

Chemometric and Molecular Modeling Study of 1*H*-Indole-3-acetic Acid Derivatives with Auxin Activity*

Rudolf Kiralj and Márcia M. C. Ferreira**

Laboratório de Quimiometria Teórica e Aplicada, Instituto de Química, Universidade Estadual de Campinas, Campinas SP, 13083-970, Brazil

RECEIVED APRIL 20, 2005; REVISED AUGUST 30, 2005; ACCEPTED SEPTEMBER 1, 2005

Key words
1*H*-indole-3-acetic acid
auxin activity
chemometrics
QSAR
hierarchical cluster analysis
principal component analysis
partial least squares regression
substrate modeling at the active site
ab initio
molecular mechanics

Quantitative Structure-Activity Relationship (QSAR) study on 22 1*H*-indole-3-acetic acid derivatives with auxin activity was performed by means of Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA), Partial Least Squares Regression (PLS) and Multiple Linear Regression (MLR). Molecular geometry of the auxins was optimized at MMFF94 and *ab initio* B3LYP 6-31G** levels. Modeling of complexes of some auxin molecules with the auxin binding protein 1 (ABP1) was also carried out. Parsimonious PLS and MLR models for prediction of optimal and half-optimal auxin concentrations for *Avena L. Sativa* coleoptile elongation were obtained with 15 auxins in the training set. HCA and PCA on data for the half-optimal concentration exhibit auxin clustering with respect to substituent type and position, biological activity, and the size of the active site pockets of ABP1. Molecular graphics of ABP1 – NAA derivative complexes and of the coordination spheres around NAA (1-naphthalenic acid) hydrogen atoms in the ABP1 – NAA complex agrees well with the chemometrics/QSAR results.

INTRODUCTION

1*H*-Indole-3-acetic acid (IAA) is the first isolated and the most important plant hormone from auxin class.¹ Structure-activity relationships for auxin derivatives² have always been complex to interpret, probably because of limited number and heterogeneity of reliable activity data, lack of insight into certain IAA intrinsic properties, and missing an IAA receptor 3D structure (a structure of a receptor has just appeared recently³). The skeleton of IAA (Figure 1) consists of a heteroaromatic planar indole ring,^{4,5} and an acetic acid side chain (at position 3) which can adopt several distinct orientations with respect to the ring² (defined by torsion angles T1: C2-C3-C8-C9 and T2: C3-C8-C9-O10). IAA auxins

(IAAs) are limited to derivatives with small substituents at five substitution positions (2, 4–7). The substitutions cause systematic variation in auxin activity, lipophilicity, and other properties. Precise geometry of IAAs and various molecular descriptors contain information about aromaticity, substituent effects, conformational flexibility and other properties important for IAAs – receptor interactions. This enables the correlation of molecular properties with biological (auxin) activity of IAAs as Quantitative Structure-Activity Relationships (QSAR). This work deals with 22 IAAs (substituents: F, Cl, Br, I; alkyls Me, Et, Pr, Bu), 15 of which have experimental straight-growth promoting activities measured for *Avena Sativa L.* coleoptiles:^{2,6} the half-optimum concentration *c* and the optimum concentration *m*, relative to IAA. Be-

* Dedicated to Dr. Edward C. Kirby on the occasion of his 70th birthday.

** Author to whom correspondence should be addressed. (E-mail: marcia@iqm.unicamp.br)

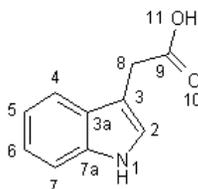


Figure 1. 1H-Indole-3-acetic acid (IAA) structure and numbering system.

sides obtaining parsimonious PLS (Partial Least Squares regression) and MLR (Multiple Linear Regression) models^{7–10} to predict biological activity and propose bioactive IAAs, the purpose of this work is to point which molecular descriptors are relevant to the auxin activity, and to interpret the QSAR results in terms of IAAs – receptor binding. For this purpose, the chemometric methods^{7–10} Hierarchical Cluster Analysis (HCA) and Principal Component Analysis (PCA), were employed.

METHODOLOGY

Optimization of molecular geometries

Molecular structure of IAA² retrieved from the Cambridge Structural Database¹¹ (CSD REFCODE: INACET03) was used for geometry optimization of IAAs at B3LYP 6-31G** level. Whilst for IAA, 7-N-IAA, Me- and halogenated IAAs the torsion angles T1 and T2 were used from the experimental structure, more complex IAAs needed conformational study prior to *ab initio* computations. For Et-derivatives, the conformation of Et group with respect to the aromatic ring was determined by MMFF94 force field¹² Monte Carlo stochastic conformational search incorporated in Titan.¹³ 5-Pr-IAA was modeled from optimized 5-Et-IAA, then the conformational search was performed, and experimental values of T1 and T2 were set before the geometry optimization. 5-Bu-IAA was modeled from 5-Pr-IAA in the same way. Molecular mechanics (MMFF94) geometry optimization was also performed on all *ab initio* geometries. This way two homogeneous sets of auxins geometry, *ab initio* and molecular mechanics geometry, were obtained.

Calculation of molecular descriptors

Steric, electronic, lipophilicity and other molecular descriptors were calculated by using Titan software, MOPAC 6.0¹⁴ (PM3 single point calculation), WebLab Viewer Pro 4.0¹⁵ and Chem3D Ultra 6.0¹⁶ on obtained structures. Topological, electrotopological, lipophilicity and other descriptors were calculated by Dragon software,¹⁷ and an online software ALOGPS 2.1¹⁸ using SMILES codes. Matlab 6.1¹⁹ was used to linearize certain descriptors.

Chemometric analysis

All activity and molecular descriptors data were autoscaled prior to chemometric analysis. Variable selection was based on cut-off in descriptor-activity correlation coefficients (0.50), reasonable trends in descriptor-activity scatterplots, considerable contribution to regression models, and the absence of tight descriptor clustering in HCA and PCA plots. Pirouette 3.01 software²⁰ was used to carry out all chemometric calculations. Values of log *c* and log *m* for some halogenated IAAs were obtained by rescaling and normalizing the original data⁶ with respect to IAA.

QSAR studies

PLS and MLR (using all principal components, PCs, in Principal Component Regression PCR^{7–9}) models were built and validated by leave-one-out crossvalidation for 15 molecules (the training set) to predict log *c* and log *m* values. Prediction of biological activity for 7-N-IAA, 7-Et-IAA, 5-Br-IAA, 5-Cl-IAA, 6-Cl-IAA, 7-Cl-IAA and 4-Cl-IAA (the prediction set) was performed using these models.

Molecular graphics and modeling of ABP1 – IAA/NAA complexes

A monomer unit of auxin binding protein 1 (ABP1) in ABP1–NAA (1-naphthalenic acid) complex³ was employed. NAA was modified into certain NAAs. Molecular graphics on ABP1–NAA, ABP1–5-Et-NAA, ABP1–4-Et-NAA and ABP1–6-Me-NAA complexes with unoptimized geometry was carried out to visualize stereoelectronic substrate – receptor relationships. Steric properties of the NAA hydrogen atoms at the ABP1 active site hole were calculated by Matlab: the number of total and non-hydrogen atoms from ABP1 and water at van der Waals contact distance (van der Waals radii from PLATON:²¹ H – 1.20, C – 1.70, O – 1.52, N – 1.55, Zn – 2.54 Å) and inside 5.5 ± 0.2 Å cut off.²² HCA and PCA were performed on the obtained hydrogen steric descriptors. Coordination spheres with 5.5 ± 0.2 Å radius around each NAA hydrogen atom were inspected by molecular graphics. The programs Titan and WebLab Viewer were used for all molecular graphics and modeling operations.

RESULTS AND DISCUSSION

Chemometrics and QSAR for log *c*

Plant-growth activity log *c* showed to be non-linearly related to several molecular descriptors, mostly those with 3D character, calculated by using the software Dragon. Two descriptors have been selected for PLS modeling (Table I): the Randić molecular shape profile No. 8 (SP08),²³ and radial distribution function at 5.5 Å interatomic distance, weighted by molecular masses (RDF055m).²⁴ SP08

TABLE I. QSAR^(a) data for 22 IAA derivatives with auxin activity $-\log c$

Molecule	SP08	RDF055m	$-\log c_{\text{exp}}^{(b)}$	$-\log c_{\text{cal}}^{(c)}$	$-\log c_{\text{cal}}^{(d)}$	$\Delta^{(e)}$	$\Delta^{(f)}$
IAA	5	3	5.5(1)	5.5	5.5	0.0	0.0
2-Me-IAA	4	1	4.7(1)	4.7	4.7	0.0	0.0
4-Me-IAA	5	4	5.6(2)	5.6	5.6	0.0	0.0
5-Me-IAA	5	2	5.3(2)	5.4	5.4	0.1	0.1
6-Me-IAA	5	2	5.4(1)	5.4	5.4	0.0	0.0
7-Me-IAA	5	3	5.5(1)	5.5	5.5	0.0	0.0
5-Et-IAA	6	2	5.6(1)	5.5	5.5	0.1	0.1
5-Pr-IAA	7	3	5.5(1)	5.5	5.5	0.0	0.0
5-Bu-IAA	7	3	5.4(1)	5.5	5.5	0.1	0.1
4-Et-IAA	5	2	5.3(1)	5.4	5.4	0.1	0.1
6-Et-IAA	7	3	5.5(2)	5.5	5.5	0.0	0.0
4-F-IAA	5	4	5.8 ^(g)	5.6	5.6	0.2	0.2
4-Cl-IAA	5	2		5.4	5.4		
5-F-IAA	5	5	5.5 ^(g)	5.5	5.5	0.0	0.0
5-Cl-IAA	5	3		5.5	5.5		
5-Br-IAA	5	8		4.6	4.6		
6-F-IAA	5	3	5.5 ^(g)	5.5	5.5	0.0	0.0
6-Cl-IAA	6	1		5.3	5.2		
7-F-IAA	5	3	5.5 ^(g)	5.5	5.5	0.0	0.0
7-Cl-IAA	5	2		5.4	5.4		
7-N-IAA	5	2		5.4	5.4		
7-Et-IAA	6	2		5.5	5.5		

(a) Molecular descriptors: SP08 – a Randić molecular profile descriptor; RDF055m – a radial distribution function. (b) $-\log c_{\text{exp}}$ – experimental half-optimal concentration (mol/l) in $-\log$ form from Nigović *et al.*² Experimental errors are in brackets. (c) $-\log c_{\text{cal}}$ – predicted half-optimal concentration (mol/l) from PLS model. (d) $-\log c_{\text{cal}}$ – predicted half-optimal concentration (mol/l) from MLR model. (e) Δ – absolute deviation of $-\log c_{\text{cal}}$ (PLS) from $-\log c_{\text{exp}}$. (f) Δ – absolute deviation of $-\log c_{\text{cal}}$ (MLR) from $-\log c_{\text{exp}}$. (g) Literature activity data,⁶ rescaled and normalized in this work, relative to the activities for the first 11 molecules.

and RDF055m are transformed into their linear forms $L(\text{SP08})$ and $L(\text{RDF055m})$, respectively: $L(\text{SP08}) = (\text{SP08} - 6)^2$ and $L(\text{RDF055m}) = (\text{RDF055m} - 3.9)^2$. The linearization increased the correlation coefficients with activity, from 0.35 to 0.86 for SP08, and from 0.69 to 0.87 for RDF055m.

The PLS model with only one PC (83.83 % variance, $Q^2 = 0.847$, $R^2 = 0.884$; SEV = 0.11), is better than the MLR model according to statistical parameters Q^2 and SEV ($Q^2 = 0.798$, $R^2 = 0.884$; SEV = 0.19), although there are no differences in deviations Δ from experimental values in terms of experimental errors (Table I). The activity values predicted by the PLS and MLR models are equal for all compounds except for 6-Cl-IAA. Relatively small variation in the measured activity (about 19 %) can be a reason why the models can't be distinguished in the predicted activity values. For 4F-IAA, which is considered to be highly active auxin,⁶ the predicted activity has the correct order of magnitude and Δ is twice of experimental error, being the maximum Δ among all IAAs (see Table I). The predicted activity for other halogenated IAAs shows regular tendency to be dependent on substitution type and position. 7-N-IAA and 7Et-IAA seem to be equally active as IAA. The PLS regression vector coefficients for descriptors $L(\text{SP08})$ and $L(\text{RDF055m})$ are -0.702 and 0.712 , respectively. The regression vec-

tor indicates the importance of 3D-based descriptors for this auxin activity. The PLS model exhibits complex nature of the auxin activity. Namely, classical steric, electronic, hydrogen bonding and lipophilicity descriptors showed low correlation with the activity, but descriptors of composite nature were promising in the final modeling. According to previous work on receptor – substrate interactions,²² 5.5 Å is a reasonable value for interato-

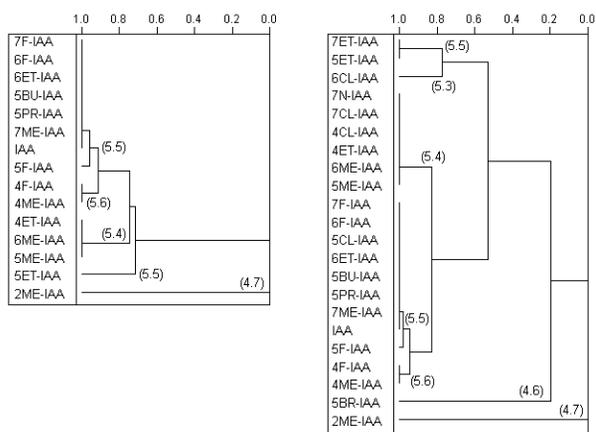


Figure 2. HCA dendrograms for training (left) and training+prediction set (right) using complete linkage method. PLS predicted activity values are in brackets next to their clusters.

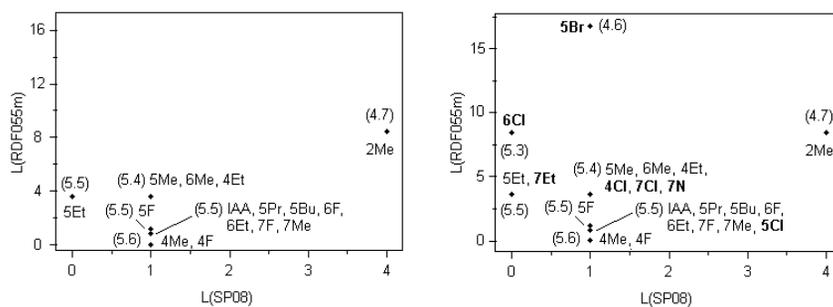


Figure 3. L(SP08) against L(RDF055m) plot for the training (left) and training+prediction set (right). PLS predicted activity values are in brackets. The samples from the prediction set are labeled bold.

mic distances between two interacting molecules, what agrees well with the selection of variable RDF055m in the final PLS model.

It is interesting to compare the clustering patterns from HCA and PCA with activity classes as predicted by the PLS regression model. HCA dendrogram for the training set (complete linkage; descriptor data matrix 15×2 ; Figure 2 left) shows certain clustering which is obvious in the dendrogram for the training + prediction set (descriptor data matrix 22×2 ; Figure 2 right). The predicted activities in the dendrograms can aid in explaining that the clustering of IAAs is based mainly on their biological activity. The compounds can be classified as less active compounds ($-\log c \leq 5.3$), moderately ($-\log c = 5.4$ or 5.5) and more active compounds ($-\log c = 5.6$).

L(SP08) – L(RDF055m) correlation plot (Figure 3 left) exhibits the clustering found in HCA dendrograms. The compounds are differentiated with respect to their increasing activities along both coordinate axes. In general, the biological activity increases when the values of both descriptors decrease. One can notice that this trend is more obvious when the samples from the prediction set are included (Figure 3 right). These predicted activities, which can be useful in rationalizing the clustering of IAAs in HCA and correlation plots, can aid in deducing the role of substitution effects. Substituent position and type, as well as possible size of the pockets at the active site hole of ABP1, are described in Table II.

Chemometrics and QSAR for log *m*

There are four molecular descriptors in Table III that have been selected for QSAR study of log optimal auxin concentration log *m*: torsion angle T2 (C3-C8-C9-O10) from auxin MMFF94 optimized geometries, the second-order polarizability β calculated by MOPAC, a Geary autocorrelation – lag6 weighted by atomic Sanderson electronegativities (GATS6e), and the first component size directional WHIM index weighted by the same set of electronegativities (L1e).²⁵ GATS6e and L1e were obtained by using Dragon software. Descriptor L1e, due to its non-linear relationship with the auxin activity, was transformed into $L(L1e) = (L1e - 9.1)^2$ for modeling. It is in-

TABLE II. Substitution effects on predicted $-\log c$ relative to IAA $-\log c = 5.5$

Substitution	Higher activity (5.6)	IAA like activity (5.4–5.5)	Lower activity (≤ 5.3)	Pocket size
2	–	–	Me	small
4	F, Me	Et, Cl	–	medium
5	–	F, Cl, Me, Et, Pr, Bu	Br	large
6	–	F, Me, Et	–	medium
7	–	F, Cl, N, Me, Et	–	medium or large

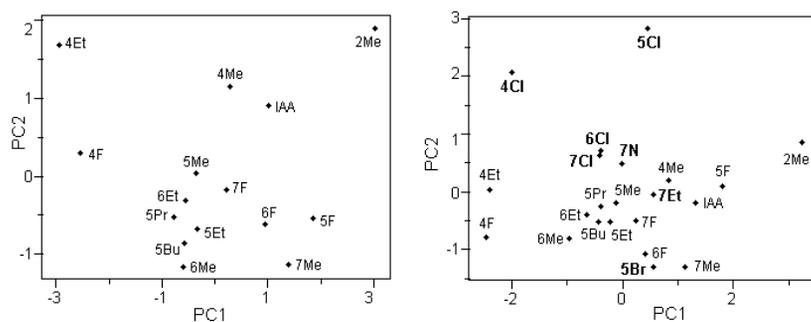


Figure 4. PC1-PC2 scores plots from PCA analysis on training set (left) and training+prediction set (right). The samples from the prediction set are labeled bold.

TABLE III. QSAR^(a) data for 22 IAA derivatives with auxin activity $-\log m$

Molecule	T2 / °	β / a.u.	GATS6e	L1e	$-\log m_{\text{exp}}^{(b)}$	$-\log m_{\text{cal}}^{(c)}$	$-\log m_{\text{cal}}^{(d)}$	$\Delta^{(e)}$	$\Delta^{(f)}$
IAA	71.32	18.844	1	6	4.1(0)	4.1	4.1	0.0	0.0
2-Me-IAA	72.62	51.591	2	5	3.7(1)	3.7	3.7	0.0	0.0
4-Me-IAA	52.09	-7.036	1	6	4.3(1)	4.2	4.2	0.1	0.1
5-Me-AA	18.79	-3.235	1	7	4.2(1)	4.3	4.3	0.1	0.1
6-Me-IAA	-19.94	26.251	1	9	4.2(1)	4.3	4.3	0.1	0.1
7-Me-IAA	84.79	52.507	1	8	4.2(2)	4.2	4.2	0.0	0.0
5-Et-IAA	23.47	3.679	1	8	4.3(1)	4.3	4.3	0.0	0.0
5-Pr-IAA	22.79	-20.436	1	10	4.4(1)	4.4	4.4	0.0	0.0
5-Bu-IAA	18.14	-3.015	1	9	4.4(1)	4.4	4.4	0.0	0.0
4-Et-IAA	-49.13	-48.207	0	6	4.3(1)	4.4	4.4	0.1	0.1
6-Et-IAA	-13.91	14.705	1	11	4.3(2)	4.2	4.2	0.1	0.1
4-F-IAA	-42.27	-16.065	0	7	4.7 ^(g)	4.5	4.5	0.2	0.2
4-Cl-IAA	-37.73	-168.631	1	7		4.7	4.7		
5-F-IAA	45.77	40.317	2	7	4.1 ^(g)	4.0	4.0	0.1	0.1
5-Cl-IAA	31.92	-168.631	2	7		4.5	4.5		
5-Br-IAA	31.02	81.260	1	7		4.1	4.1		
6-F-IAA	27.42	64.402	1	7	4.1 ^(g)	4.1	4.1	0.0	0.0
6-Cl-IAA	24.81	-73.952	1	7		4.5	4.5		
7-F-IAA	32.37	19.414	1	7	4.1 ^(g)	4.2	4.2	0.1	0.1
7-Cl-IAA	21.62	-67.491	1	7		4.4	4.4		
7-N-IAA	9.53	-22.361	1	6		4.2	4.2		
7-Et-IAA	81.11	-43.395	1	8		4.5	4.5		

^(a) Molecular descriptors: T2 – torsion angle T2 (C3-C8-C9-O10) obtained from MMFF94 geometry optimization; β – the second-order molecular polarizability in atomic units; GATS6e – a Geary autocorrelation – lag6 weighted by atomic Sanderson electronegativities; L1e – the first component size directional WHIM index weighted by the same set of electronegativities. ^(b) $-\log m_{\text{exp}}$ – experimental optimal concentration (mol/l) in $-\log$ form from Nigović *et al.*² Experimental errors are in brackets. ^(c) $-\log m_{\text{cal}}$ – predicted half-optimal concentration from PLS model. ^(d) $-\log m_{\text{cal}}$ – predicted half-optimal concentration from MLR model. ^(e) Δ – absolute deviation of $-\log m_{\text{cal}}$ (PLS) from $-\log m_{\text{exp}}$. ^(f) Δ – absolute deviation of $-\log m_{\text{cal}}$ (MLR) from $-\log m_{\text{exp}}$. ^(g) Literature activity data,⁶ rescaled and normalized in this work, relative to the activities for the first 11 molecules.

interesting to note that even the torsion angle T2 from MMFF94 optimized ABP1 – auxin geometries²⁶ shows moderate correlation with $-\log m$ (correlation coefficient -0.62). PLS model with three PCs (92.38 % variance, $Q^2 = 0.539$, $R^2 = 0.799$; SEV = 0.15) and MLR ($Q^2 = 0.531$, $R^2 = 0.799$; SEV = 0.15), predicted the activities of IAAs equally well, with Δ not greater than experimental errors (Table III). For 4F-IAA, which is considered to be a highly active auxin,⁶ deviation Δ of calculated from experimental activity reached maximum value 0.2. 7-N-IAA and 7Et-IAA are predicted to be more active than IAA. In fact, experimental findings²⁷ confirm the auxin activity for 7-N-IAA. Among halogenated IAAs, 4F-IAA were underpredicted by $\Delta = 0.2$, and chlorinated IAAs were probable overpredicted. Perhaps the lack of reliable activity data for all halogenated IAAs is what might improve the regression models and produce different predicted activities for halogenated samples. The PLS regression vector coefficients for descriptors T2, β , GATS6e and L(L1e) are 0.016, -0.374 , -0.404 , and -0.460 , respectively. The regression vector shows the significance of de-

scriptors based on 2D (GATS6e) and 3D (T2, L1e) molecular geometry as well as electronic features (β , partially GATS6e and L1e).

PCA analysis on the training set (PC1: 59.81 %; PC2: 24.17 %; PC3: 8.76 %) and also on the full data set (PC1: 43.22 %, PC2: 25.22 %; PC3: 23.17 %) exhibits rather complex nature of IAAs clustering (Figure 4). As can be observed from Table III, halogenated and alkylated IAAs cannot be distinguished according to their activities; they are rather mixed in the PCA scores (Figure 4) and HCA dendrograms (not shown). Discrimination of samples with respect to the predicted activities in PCA is rather visible along PC1 and PC2. Such behavior of the studied IAAs reminds on observed trends with respect to $-\log c$ activity (previous section).

Molecular graphics and modeling

The only known structure of a substrate with ABP1 receptor is that of 1-naphthalenic acid3 (NAA, Figure 5 left). This auxin is structurally very much similar to IAA, what one can observe when superimposing the two structures

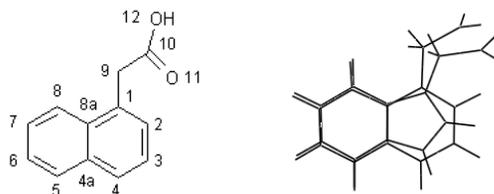


Figure 5. Left: 1-Naphthalenic acid (NAA) structure and numbering system. Right: Superimposed MMFF94 geometries of IAA and NAA anions.

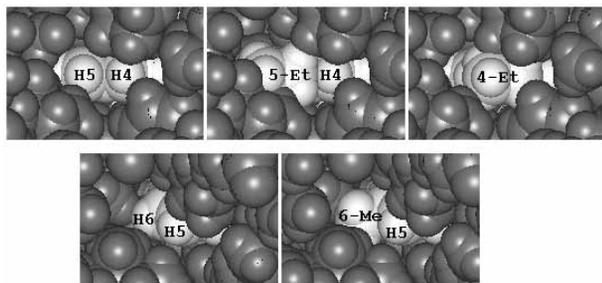


Figure 6. Molecular graphics and modeling of ABP1-NAA complexes. Upper row: ABP1-NAA, ABP1-5-Et-NAA and ABP1-4-Et-NAA complexes, showing the pocket around substitution positions 4 and 5. Lower row: ABP1-NAA and ABP1-6-Me-IAA complexes. Water molecule No. 3176 was removed to show the substitution position 6.

(Figure 5 right). This is the reason why ABP1-NAA relationships can be used to predict ABP1-IAA binding properties. Benzene hydrogen atoms in IAA and in NAA can be almost completely superimposed (Figure 5 right). IAA hydrogen atom at N1 overlaps with H4 in NAA, while H2 (IAA) is placed along the direction in between H2 and H3 of NAA. Methylene hydrogens, due to slight difference in orientation of the acetic groups in both acids, can be useful in describing the whole active site hole. Molecular graphics on the crystal structure of ABP1-NAA complex (Figures 6, 7) shows that H5 and partially H4 of NAA lie in a large pocket visible from the outside of the protein. Accommodation of Et at these two positions in the pocket is quite reasonable. H6 is po-

sitioned in a smaller pocket, and hence, Me substituent seems to be the most appropriate alkyl group for this pocket. When studying the coordination sphere around each NAA hydrogen atom at 5.5 Å cut-off (Figure 7), one can notice that there are pockets of different size and even shape, as characterized in Table IV. The pockets around the equivalent IAA hydrogen atoms were identified in the same way as from predicted activities $-\log c$ (see Tables I and II, and also Figures 2, 3). Table IV contains six steric parameters for NAA hydrogen atoms. HCA on these data produces four clusters (Figure 8 left): two of them are formed by hydrogens placed in very small or small pockets (clusters B and C), and the others are mixtures of hydrogens in medium and large pockets (clusters A and D). In other words, 9H-anti and the hydrogen atoms at the imidazole ring with its substituents are placed in smaller ABP1 pockets; the hydrogens of

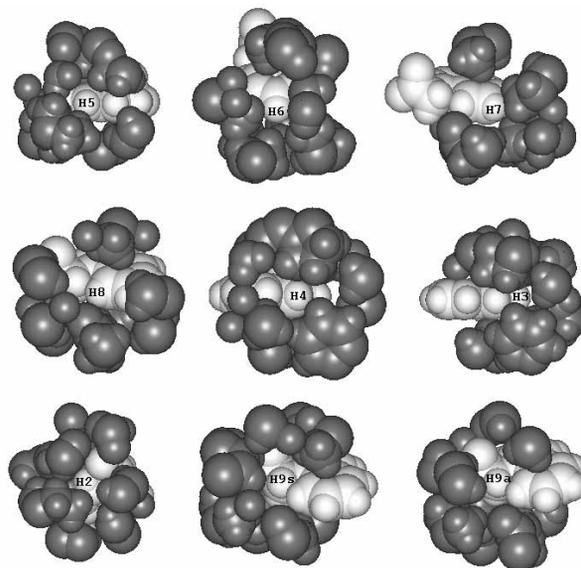


Figure 7. Molecular graphics of coordination sphere (dark) around each NAA hydrogen atom (light). H9a and H9s stand for H9-anti and H9-syn, respectively.

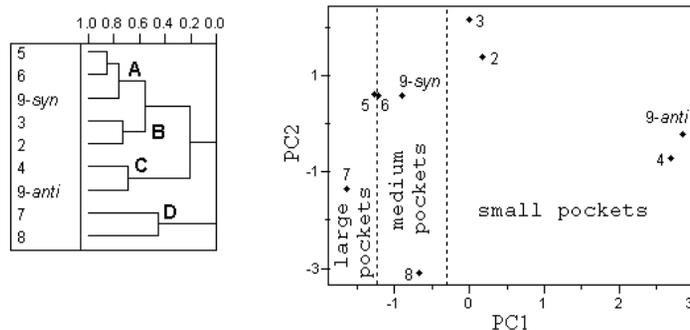


Figure 8. Left: HCA on nine NAA hydrogen atoms steric parameters using complete linkage method. Right: PC1-PC2 scores plot from PCA on the same data set, showing the discrimination of the hydrogens with respect the ABP1 counterpart pockets.

TABLE IV. Steric properties^(a) of NAA hydrogen atoms and the observed size of the ABP1 counterpart pockets

NAA ^(b)	IAA eq. ^(c)	Count1	Count2	Min. <i>d</i>	Count3	Non-H	Count4	Pocket size ^(d)
9 <i>anti</i>	8 <i>anti</i>	2	1	2.90	27	0.54	21	very small
9 <i>syn</i>	8 <i>syn</i>	0	0	3.43	23	0.46	17	medium
2	[2] ^(e)	0	0	3.35	26	0.48	20	small
3	[2] ^(e)	0	0	3.70	27	0.44	22	small
4	1	3	1	2.91	27	0.46	11	small
8	4	1	1	2.68	17	0.38	15	medium
7	5	2	0	3.31	19	0.35	16	large
6	6	0	0	3.46	24	0.39	18	medium
5	7	0	0	3.54	22	0.42	18	large

^(a) Steric properties of the ABP1 active size hole around each NAA hydrogen atom inside 5.5 ± 0.2 Å cut-off: Count1 – the number of van der Waals contacts; Count2 – the number of van der Waals contacts including only non-hydrogen atoms; Min. *d* – the minimum distance observed for the van der Waals contacts; Count3 – the number of non-hydrogen ABP1/water atoms inside the cut-off; Non-H – the number fraction of non-hydrogen atoms count; Count4 – the number of ABP1 carbon atoms inside the cut-off. ^(b) NAA hydrogen atoms, according to the numbering system from Figure 5; *syn* and *anti* atoms at C9 are named whether they are oriented up or down with respect to the aromatic plane in orientation as in Figure 5. ^(c) IAA hydrogen atoms according to the numbering system from Figure 1, equivalent to the analogous NAA hydrogen atoms; *syn* and *anti* atoms at C8 with respect to the aromatic ring plane. ^(d) The pocket size as observed from molecular graphics. ^(e) This IAA hydrogen is in the mid-way between the two NAA hydrogen atoms (see Figure 5 right).

the benzene ring and its substituents are situated in larger ABP1 pockets. PCA analysis on the data from Table IV (2 PCs contain 87.56 % of the original data) discriminate small from medium and large pockets along PC1 (Figure 8 right).

Substitution positions on IAA can be considered as having pocket counterparts as follows: 2 – very small to small (vss); 4 and 6 – medium (md); 5 and 7 – large (l). Experimental and predicted $-\log c$ growth-promoting activities of alkylated² and fluorinated⁶ IAAs agree with the pocket size. The order of increasing activities 7-F – IAA, 5-F – IAA < IAA < 6-F – IAA < 4-F – IAA is preserved when substitution position and pocket sizes are arranged as 7(l), 5(l), 0(for IAA) < 6(md) < 4 (md). Small fluorine atom prefers small pockets. The 4(md) pocket with two hydrogens and an oxygen from ABP1 inside van der Waals contact distance (see Table IV) is certainly smaller than the 6(md) pocket without van der Waals contacts. Another good correlation exists between the activities 4-Et-IAA < 6-Et-IAA, IAA < 5-Et-IAA and the corresponding sequence of pockets 4(md) < 6(md), 0 < 5(l). Et substituent requires large pockets. Obviously, 4(md) is, due to the existence of van der Waals contacts (Table IV, Count1 and Count2 parameters), smaller than the 6(md) pocket. Activity – hydrogen parameters correlation is less notable for Me derivatives. While activity increases in order 2-Me-IAA < 5-Me-IAA < 6-Me-IAA < 7-Me-IAA, IAA < 4-Me-IAA, the pockets can be arranged as 2(vss) < 5(l) < 6(md) < 7(l), 0 < 4(md) knowing that 4(md) < 6(md) and 5(l) < 7(l) is due to van der Waals contacts (Count1 and Count2 parameters in Table IV). As noticed before, Me group, like F atoms, prefers smaller pockets. Figure 6 shows that Me is more appropriated for position 6 (6 in NAA also) and Et for 7 (5 in NAA). 5-Alkyl derivatives² can be arranged in order of

their $-\log c$ values as Me-IAA < Bu-IAA < Pr-IAA, IAA < Et-IAA. As 5 is characterized as position with large pocket in ABP1, which is practically of the same size as that for position 7, it is clear that Et is the best 7-substituent.

Experimental and predicted $-\log m$ activities do not have well-defined correlation with ABP1 pocket characteristics as was the case with $-\log c$. This means that $-\log m$ cannot be interpreted in terms of ABP1 – auxin binding in terms of selected molecular descriptors. The bell-shaped dose-response curves for auxin growth-promoting activity² clearly show that, due to variations in the curves for auxins under study, there is a moderate correlation between the positions of the peaks (optimal concentration *m*) and the position of the inflexion points determined at the half value of coleoptile elongation (half-optimal concentration *c*), with correlation coefficient being equal to 0.771. That is why mechanistic interpretation of $-\log m$ should not include necessarily the same effects related to the growth-promoting activity $-\log c$.

Torsion angles T1 (C2–C3–C8–C9) and T2 (C3–C8–C9–O10) define the relative orientation of the IAA side chain with respect to the indole ring, and the orientation of the carboxy group, respectively. In the crystal structure of IAA and its derivatives² the side chain is perpendicular to (T1 around -90° , anticlinal conformation) or parallel with the indole ring (T1 around 0° , synperiplanar conformation). *Ab initio* geometries of 22 IAAs in this work, yield strong correlation between T1 and T2 (Figure 9). Steric hindrance of the acetic group by 2-Me is the reason why the conformation around C3–C8 bond is almost anticlinal. Other substituents shift conformation into the synclinal region. Substituents at position 4 interact in the same way with H8-*syn* and H8-*anti*, what

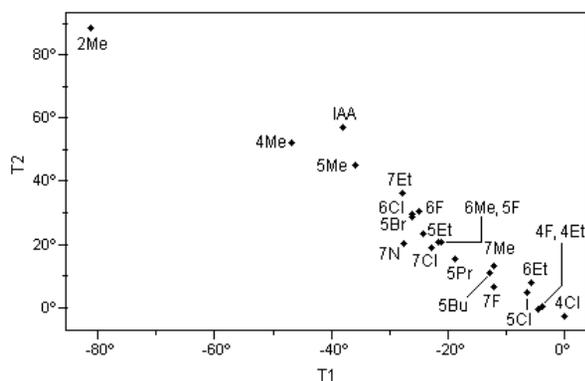


Figure 9. Torsion angles T1 against T2 in *ab initio* structures of 22 IAAs.

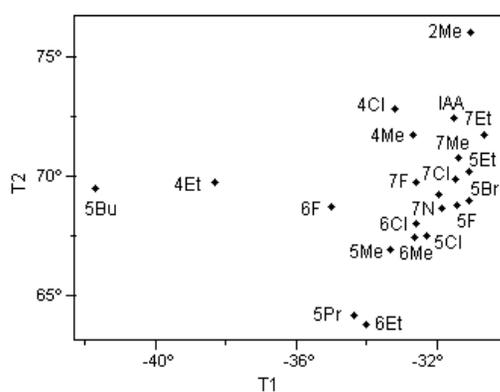


Figure 10. Torsion angles T1 against T2 in MMFF94 structures of 22 IAAs complexed with ABP1.²⁶

turns T1 close to 0°. MMFF94 structures of all 22 IAAs in free and complexed state (an ABP1 protomer complexed to an auxin at the active site was optimized without constraints in a recent work²⁶) show some conformational change due to complex formation. T1 and T2 in complexed state vary only $\approx 10^\circ$ (Figure 10). In free state, auxins with more bulky substituents (Pr, Bu) and substituents sterically hindered by the IAA side chain (4Et, 6Et), substituents placed in small ABP1 pockets (2Me and 6Et) or too small substituents (6F, 5Me), are the farthest samples from the region with the highest concentration of samples in the T1-T2 plot (Figure 9). An auxin molecule is bound to hexacoordinated Zn²⁺ ion in ABP1 *via* two Zn – O chemical bonds.³ Thus ABP1 – auxin binding mode may be directly related to biological activity. Torsion angles in Figure 10 might be qualitatively correlated with $-\log c$ activities like has been demonstrated for pocket sizes. For example, $-T1$ for 5-alkyls increases in order Bu > Pr > Me > 0 > Et, and the corresponding experimental $-\log c$ values increase in order Me < Bu < Pr, 0 < Et. The behavior of other series of auxins is not so obvious. However, correlation between T1, T2 and $-\log c$ is worth mentioning as a possible starting point for QSAR and molecular modeling studies of auxins.

CONCLUSIONS

This work has demonstrated a good complementarity of different methods applied to the study on auxin plant-growth promoting activity: computational methods, chemometrics, QSAR analysis, molecular graphics and molecular modeling, and methods of structural chemistry. The aim of this work was to describe ABP1 – auxin interaction physically, including some aspects of ABP1-auxin binding, and to propose new active compounds – (7N-IAA, 7-Et-IAA) with their most representative molecular properties based on geometry, topology and electronic structure. It is important to emphasize that the predicted half-optimal concentration activity gave the same description of the ABP1 pockets at the active site as molecular graphics and modeling of ABP1–NAA derivative complexes. The auxin side-chain geometry in free and complexed state is a good indicator of ABP1 – auxin binding. The stereoelectronic properties of the ABP1 pockets as well as of substituents on NAA or IAA are the major factors defining ABP1 – auxin interaction.

Acknowledgement. – The authors thank FAPESP for financial support.

REFERENCES

1. R. Arteca, *Plant Growth Substances: Principles and Applications*, Chapman&Hall, New York, 1996.
2. B. Nigović, S. Antolić, B. Kojić-Prodić, R. Kiralj, V. Magnus, and B. Salopek-Sondi, *Acta Crystallogr., Sect. B* **56** (2000) 94–111.
3. E. J. Woo, J. Marshall, J. Baully, J. G. Chen, M. Venis, R. M. Napier, and R. W. Pickersgill, *EMBO J.* **21** (2002) 2877–2885.
4. D. T. Davies, *Aromatic Heterocyclic Chemistry*, Oxford University Press, Oxford, 1995, Chapter 7.
5. T. L. Gilchrist, *Heterocyclic Chemistry*, 3rd ed., Addison-Wesley-Longman Publishing Co., Harlow, UK, 1997.
6. S. Antolić, B. Kojić-Prodić, S. Tomić, B. Nigović, V. Magnus, and J. D. Cohen, *Acta Crystallogr., Sect. B* **52** (1996) 651–661.
7. K. Beebe, R. Pell, and M. B. Seasholtz, *Chemometrics: a practical guide*, John Wiley & Sons, New York, 1998.
8. M. M. C. Ferreira and R. Kiralj, *Métodos Quimiométricos em QSAR*, in: C. A. Montanari (Ed.), *Proceedings of the First Brazilian Symposium On Medicinal Chemistry*, in press.
9. M. M. C. Ferreira, *J. Braz. Chem. Soc.* **13** (2002) 742–753.
10. R. Kiralj and M. M. C. Ferreira, *J. Mol. Graphics Modell.* **21** (2003) 435–448.
11. F. H. Allen, *Acta Crystallogr., Sect. B* **58** (2002) 380–388.
12. T. A. Halgren, *J. Comput. Chem.* **17** (1996) 490–519.
13. Titan, Version 1, Wavefunction, Inc., Irvine, CA, USA, 2000.
14. V. S. Lobanov, MOPAC, version 6.0 for Microsoft Windows, University of Florida, USA, 1996.
15. WebLab ViewerPro, Version 4.0, Accelrys, Inc., Burlington, MA, USA, 2000.
16. Chem3D Ultra, Version 6.0, CambridgeSoft.Com, Cambridge, MA, USA, 2000.

17. R. Todeschini, V. Consonni, A. Mauri, and M. Pavan, Dragon, Web Version 3.0, Milano, Italy, 2003.
18. ALOGPS, Version 2.1., Virtual Computational Chemistry Laboratory, 2002. URL: <http://146.107.217.178/lab/alogps/>
19. Matlab, Version 6.1.0.450 release 12.1, MathWorks, Inc., Natick, MA, 2001.
20. Pirouette, Version 3.01, Infometrix, Inc., Woodinville, WA, USA, 2001.
21. A. L. Spek, PLATON – A Multipurpose Crystallographic Tool, Version 31000, Utrecht University, Utrecht, The Netherlands, 2000.
22. R. Kiralj and M. M. C. Ferreira, *J. Mol. Graphics Modell.* **21** (2003) 435–448.
23. M. Randić, *J. Chem. Inf. Comput. Sci.* **35** (1995) 372–382.
24. M. C. Hemmer, V. Steinhauer, and J. Gasteiger, *Vib. Spectrosc.* **19** (1999) 151–164.
25. P. Gramatica, M. Corradi, and V. Consonni, *Chemosphere* **41** (2000) 763–777.
26. R. Kiralj and M. M. C. Ferreira, *Chemom. Intell. Lab. Syst.*, submitted.
27. S. Antolić, B. Kojić-Prodić and V. Magnus, *Acta Crystallogr., Sect. C* **56** (2000) 1026–1027.

SAŽETAK

Kemometričko proučavanje i molekularno modeliranje derivata 1*H*-indol-3-octene kiseline s auksinskom aktivnošću

Rudolf Kiralj i Márcia M. C. Ferreira

Proučavanje kvantitativnih relacija između strukture i auksinske aktivnosti za 22 derivata 1*H*-indol-3-octene kiseline sprovedeno je pomoću analize glavnih komponenata (PCA), hijerarhijske grozdaste analize (HCA), regresije parcijalnih najmanjih kvadrata (PLS) i višestruke linearne regresije (MLR). Molekularna geometrija auksina optimizirana je na razini MMFF94 i *ab initio* B3LYP 6-31G** teorije. Izvršeno je i modeliranje kompleksa nekih auksinskih molekula i auksinskoga proteina 1 (ABP1). Zadovoljavajući modeli PLS i MLR za predviđanje optimalne i polu-optimalne auksinske koncentracije za produljenje koleoptila *Avena L. Sativa* dobiveni su pomoću test-skupova s 15 auksina. Analize podataka HCA i PCA za polu-optimalnu koncentraciju klasificirale su auksine s obzirom na vrstu i mjesto supstitucije, biološku aktivnost i veličinu džepova u aktivnom mjestu proteina ABP1. Molekularna grafika kompleksa ABP1–NAA derivati kao i koordinacijskih sfera oko vodikovih atoma u NAA (1-naftalenska kiselina) dobro se slaže s kemometričkim i QSAR rezultatima.