3D-QSAR on 1-(Substituted Phenoxyacetoxy)alkylphosphonates and Phosphinates Using CoMFA and CoMSIA

Junbo He, Guoqiang Dai, Hao Peng*, and Hongwu He*

Key Laboratory of Pesticide and Chemical Biology, Ministry of Education, and College of Chemistry, Central China Normal University, Wuhan, China

Email: jb112he@yahoo.com.cn

Supplemental Materials

Materials and Methods

Data set for analysis.

The herbicidal activity of compounds **1-49** (**Table S 1**) against dicotyledonous stem was tested according to the modified method described previously. Forty-three compounds served as a training set and six additional compounds (footnoted molecules in **Table S 1**) were used as a test set to evaluate the predictive ability of the model obtained in this experiment. The compounds in the test set have been randomly selected and not been included in the training set during study. The IC_{50} (concentration causing 50% inhibitory effect) values for the training set were converted to pIC_{50} ($-\log IC_{50}$) values and used as dependent variables in the CoMFA and CoMSIA QSAR.

Table S 1. Structure and biological activities of molecules used in training and testing set.

Compd.	R ¹	R ²	R ³	R ⁴	X	Y	pIC ₅₀
1	CH ₃ O	CH ₃ O	CH ₃	H	2-Cl	4-Cl	5.12
2	CH ₃ O	CH ₃ O	H	H	2-Cl	4-Cl	5.11
3	CH ₃ O	CH ₃ O	Et	H	2-Cl	4-Cl	4.85
4	CH ₃ O	CH ₃ O	CCl ₃	H	2-Cl	4-Cl	5.31
5	CH ₃ O	CH ₃ O	Ph	H	2-Cl	4-Cl	5.05
6	CH ₃ O	CH ₃ O	2-ClPh	H	2-Cl	4-Cl	5.10
7	CH ₃ O	CH ₃ O	3-BrPh	H	2-Cl	4-Cl	5.26
8 [#]	CH ₃ O	CH ₃ O	Pr	H	2-Cl	4-Cl	5.03
9	CH ₃ O	CH ₃ O	H	CH ₃	2-Cl	4-Cl	5.17
10	CH ₃ O	CH ₃ O	CH ₃	CH ₃	2-Cl	4-Cl	5.12
11	CH ₃ O	CH ₃ O	Et	CH ₃	2-Cl	4-Cl	5.03
12	CH ₃ O	CH ₃ O	Pr	CH ₃	2-Cl	4-Cl	4.86
13	CH ₃ O	CH ₃ O	2-ClPh	CH ₃	2-Cl	4-Cl	5.03
14	CH ₃ O	CH ₃ O	3-NO ₂ Ph	CH ₃	2-Cl	4-Cl	4.99
15	CH ₃ O	CH ₃ O	4-NO ₂ Ph	CH ₃	2-Cl	4-Cl	5.53
16 [#]	CH ₃ O	CH ₃ O	Ph	CH ₃	2-Cl	4-Cl	4.52
17	CH ₃ O	CH ₃ O	3-NO ₂ Ph	H	2-Cl	4-Cl	5.30
18	CH ₃ O	CH ₃ O	4-NO ₂ Ph	H	2-Cl	4-Cl	5.65
19	CH ₃ O	CH ₃ O	Ph	H	2-H	4-NO ₂	3.72

20	CH ₃ O	CH ₃ O	Pr	H	2-H	4-NO ₂	2.97
21	C ₂ H ₅ O	C ₂ H ₅ O	H	H	2-Cl	4-Cl	4.87
22	C ₂ H ₅ O	C ₂ H ₅ O	CH ₃	H	2-Cl	4-Cl	5.31
23	C ₂ H ₅ O	C ₂ H ₅ O	Et	H	2-Cl	4-Cl	4.59
24 [#]	C ₂ H ₅ O	C ₂ H ₅ O	Pr	H	2-Cl	4-Cl	5.58
25	C ₂ H ₅ O	C ₂ H ₅ O	4-CH ₃ Ph	H	2-Cl	4-Cl	4.82
26	C ₂ H ₅ O	C ₂ H ₅ O	2-ClPh	H	2-Cl	4-Cl	4.68
27	C ₂ H ₅ O	C ₂ H ₅ O	H	H	2-Cl	H	4.53
28	C ₂ H ₅ O	C ₂ H ₅ O	CH ₃	H	2-Cl	H	4.85
29	C ₂ H ₅ O	C ₂ H ₅ O	CCl ₃	H	2-Cl	H	4.89
30	C ₂ H ₅ O	C ₂ H ₅ O	Pr	H	2-Cl	H	4.66
31	C ₂ H ₅ O	C ₂ H ₅ O	Ph	H	2-Cl	H	4.71
32 [#]	C ₂ H ₅ O	C ₂ H ₅ O	Et	H	2-Cl	H	4.52
33	C ₂ H ₅ O	C ₂ H ₅ O	CH ₃	CH ₃	2-Cl	4-Cl	5.39
34	C ₂ H ₅ O	C ₂ H ₅ O	Ph	CH ₃	2-Cl	4-Cl	4.88
35	OH	OH	H	H	2-Cl	4-Cl	5.78
36	OH	OH	CH ₃	H	2-Cl	4-Cl	5.77
37	OH	OH	Et	H	2-Cl	4-Cl	5.46
38	OH	OH	Pr	H	2-Cl	4-Cl	5.91
39	OH	OH	Bu	H	2-Cl	4-Cl	5.94
40	OH	OH	Ph	H	2-Cl	4-Cl	6.27
40 [#]	CH ₃	CH ₃ O	Ph	H	2-Cl	4-Cl	5.06
42	CH ₃	CH ₃ O	4-CH ₃ Ph	H	2-Cl	4-Cl	5.12
43	CH ₃	CH ₃ O	2-ClPh	H	2-Cl	4-Cl	5.39
44	CH ₃	CH ₃ O	4-ClPh	H	2-Cl	4-Cl	5.21
45	CH ₃	CH ₃ O	4-CH ₃ OPh	H	2-Cl	4-Cl	5.47
46	CH ₃	CH ₃ O	3-NO ₂ Ph	H	2-Cl	4-Cl	5.81
47	CH ₃	CH ₃ O	Pr	H	2-Cl	4-Cl	5.38

48	CH ₃	CH ₃ O	Et	H	2-Cl	4-Cl	5.48
49 [#]	CH ₃	CH ₃ O	CH ₃	H	2-Cl	4-Cl	5.51

[#] The compounds in the test set for model validation.

Molecular modeling and molecular alignment.

All molecular modeling studies, CoMFA and CoMSIA, were performed using SYBYL 7.1 running on a Silicon Graphics Fuel workstation. Initially, the most potency compound **40** (since it having the highest IC₅₀) was sketched. Minimization of this compound was performed with Gaussian 03 and then in SYBYL using Tripos forces field, the MMFF94 charges with a distance-dependant dielectric and Powell conjugate gradient algorithm with convergence criterion of 0.005 kcal/(mol·Å). Then, the lowest energy conformation of compound **40** was used as a template to build the other molecular structures. Each structure was energy-minimized using conjugate gradient minimization algorithm with the Tripos force field until a gradient convergence of 0.005 kcal/(mol·Å). MMFF94 charges were calculated for all compounds.

Generally, a geometric similarity should exist between the modeled structure and that of the bioactive conformation for 3D-QSAR. A database of the energy-minimized structure of **1-49** was aligned using the “align database” method in SYBYL with the molecular **40** was chosen as the template of the

alignment (**Figure S 1**), asterisks in the structural formula as shown in the top of **Table S 1** indicate the atoms (P and O) used for the alignment.

Figure S 1. Stereoview of all aligned compounds in the training and test set.

CoMFA and CoMSIA Calculations.

The steric and electrostatic potential fields for CoMFA were calculated at each lattice intersection of a regularly spaced grid of 2.0 Å in x, y and z direction. The grid box dimensions were determined automatically in such a way that the region boundaries were extended beyond 4 Å in each direction from the coordinates of each molecule. A sp^3 carbon atom with a distance dependent dielectric constant of 1.0 Å and +1 charge served as probe atom to calculate steric and electrostatic fields. The steric and electrostatic contributions were truncated to +30 kcal/mol and electrostatic contribution were ignored at the lattice intersections with maximal steric interactions. The CoMFA fields generated automatically were scaled by the CoMFA-STD method in SYBYL.

CoMSIA similarity indices descriptors were derived according to Klebe et al.¹⁸ with same lattice box as was used for the CoMFA calculations, with a grid spacing of 2 Å employing a C^{+1} probe atom with a radius of 1.0 Å as implemented in SYBYL. CoMSIA similarity indices (A_F) for a molecule j with atoms i at a grid point q are calculated as follows,

$$A_{F,K}^q(j) = \sum_i \omega_{probe,k} \omega_{ik} e^{-ar} i q^2$$

where $w_{probe,k}$ is the probe atom with radius 1 Å, charge +1, hydrophobicity +1, hydrogen bond donating +1, hydrogen bond accepting +1; w_{ik} is the actual value of the physicochemical property k of atom i ; r_{iq} is the mutual distance between the probe atom at grid point q and atom i of the test molecules; a is the attenuation factor, with a default value of 0.3. A Gaussian type distance dependence was considered between the grid point q and each atom i of the molecule, where r represents the distance. A lattice of 2 Å grid spacing was generated automatically.

Partial Least-Square (PLS) Calculations and Validations.

Partial least-square (PLS) methodology was used for all 3D-QSAR analysis. The CoMFA and CoMSIA descriptors were used as independent variables, and pIC_{50} values were used as dependent variables in partial least-squares regression analyses to derive 3D-QSAR models using the standard implementation in the SYBYL package. Cross-validation is used to check the predictivity of the derived model. Results of the analyses correspond to a regression equation with thousands of coefficients. The cross-validation analysis was performed using leave-one-out (LOO) method in which one compounds is removed from the dataset and its activity is predicted using the model derived from the rest of the dataset. The cross-validated q^2 that resulted in optimum number of components and the lowest standard error of prediction were taken. The optimum number of components (N) used to derive the non cross-validated model was defined as the number of components with the highest R^2 and the lowest standard error of prediction (SEP). To obtain the statistical confidence limits on the analyses, bootstrapping was carried out with 100 groups.