PARTIAL LEAST SQUARES MODEL OF MOULTING ACCELERATING COMPOUNDS WITH INSECTICIDE ACTIVITY AGAINST LEPIDOPTERAN SPECIES†

Simona Funar-Timofei[1*], Alina Bora[1], Ana Borota[1], Daniela Ionescu[2],Luminita Crisan[1]

[1]*Institute of Chemistry Timisoara of the Romanian Academy, Bul. Mihai Viteazu 24, 300223 Timisoara, Romania*
[2]*University of Medicine and Pharmacy, Faculty of Pharmacy, P-ta Eftimie Murgu 2-4, 300034 Timisoara, Romania*
e-mail: timofei@acad-icht.tm.edu.ro

Abstract

In this study the insecticidal activity of a series of 33 dibenzoylhydrazinederivatives, expressed as the pEC_{50}activity measured *in vitro*, based on an ecdysone-dependent reporter assay using cell lines derived from one lepidopteran species (the cotton leafworm *Spodoptera littoralis*), was correlated with structural descriptors using the partial least squares (PLS) approach. The data set was energy pre-optimized by molecular mechanics calculations using the MMFF94s force field. Several 0D, 1D, 2D and 3D descriptors were calculated for the minimum energy conformers. A two-components PLS model was obtained with acceptable statistical quality ($R^2X(Cum) = 0.705$, $R^2Y(cum) = 0.821$ and $Q^2(Cum) = 0.793$) for modeling the insecticidal activity. The model goodness of fit tested with the Y-randomization test indicated a stable model. Specific dibenzoylhydrazine structural features supplying information about topological distances and descriptors sensitive to any conformational change influence the insecticidal activity.

Introduction

Dibenzoylhydrazine compounds are insect growth regulators that act through the induction of an early and lethal larval molting process in vulnerable insects that belong to the species of Lepidoptera and Coleoptera [1]. These compounds are effective in the pest control because they activate the steroid receptor complex of ecdysone type at lower concentrations than the natural hormone, and because the insect cannot remove them efficiently from its body. As consequence, a constant state of ecdysteroid signaling is displayed in the insect, which avoids it to complete the molting process and for which a decrease in ecdysteroid signaling is required. Because the insect stays permanently trapped in the molting process and is unable to feed, it dies in the period of a few days from desiccation and starvation.

The importance of the unusual high affinity for the ecdysone receptor of lepidopteran insects of dibenzoylhydrazine non-steroidal ecdysone agonists has been recognized [2]. The molecular mechanism of action of ecdysteroids, was not explained until now because one of the three interaction sites of the hormone-receptor model is not present in some active compounds [3].

The objective of this paper is to determine the structural features of a series of 33 dibenzoylhydrazine ecdysone agonists [4] (Table 1), which influence the lethal larval molting process in susceptible insects that belong to the orders of one lepidopteran species, namely the cotton leafworm *Spodoptera littoralis*. The quantitative relationship between chemical features and the ecdysone agonistic activitywas determinedby means of the partial least squares (PLS) approach.

† Dedicated to the 150th anniversary of the Romanian Academy

Table 1. The smiles notation of dibenzoylhydrazine analogue structure and their experimentalinsecticidal (pEC_{50}) and predicted activity values (pEC_{50pred}) obtained using the PLS method

No	Smiles	pEC_{50}	pEC_{50pred}
1	CC(C)(C)N(NC(=O)C1=CC=CC=C1)C(=O)C1=CC=CC=C1	5.89	6.14
2	CCC1=CC=C(C=C1)C(=O)NN(C(=O)C1=CC(C)=CC(C)=C1)C(C)(C)C	8.66	7.53
3	COC1=CC=CC(C(=O)NN(C(=O)C2=CC(C)=CC(C)=C2)C(C)(C)C)=C1C	8.22	8.23
4	CC(C)(C)N(NC(=O)C1=CC=C(Cl)C=C1)C(=O)C1=CC=CC=C1	6.34	6.65
5	CC1=CC(=CC(C)=C1)C(=O)N(NC(=O)C1=C(C)C2=C(OCCC2)C=C1)C(C)(C)C	8.58	8.41
6	CC(C)(C)N(NC(=O)C1=CC=CC=C1)C(=O)C1=C(C=CC=C1)C(F)(F)F	6.1	5.99
7	CC(C)(C)N(NC(=O)C1=CC=CC=C1)C(=O)C1=C(Cl)C(Cl)=C(Cl)C=C1	5.28	5.34
8	CC(C)(C)N(NC(=O)C1=CC=CC=C1)C(=O)C1=C(Cl)C(Cl)=C(Cl)C=C1	6.3	6.99
9	CC(C)(C)N(NC(=O)C1=C(F)C=CC=C1)C(=O)C1=C(Cl)C=CC=C1	6.54	6.08
10	CC1=C(C=CC=C1)C(=O)NN(C(=O)C1=C(Cl)C=CC=C1)C(C)(C)C	6.36	6.98
11	CC(C)(C)N(NC(=O)C1=CC(F)=CC=C1)C(=O)C1=C(Cl)C=CC=C1	6.34	5.54
12	CC(C)(C)N(NC(=O)C1=CC(Cl)=CC=C1)C(=O)C1=C(Cl)C=CC=C1	5.77	5.82
13	CC(C)(C)N(NC(=O)C1=CC=C(Br)C=C1)C(=O)C1=C(Cl)C=CC=C1	6.36	6.32
14	CCCC1=CC=C(C=C1)C(=O)NN(C(=O)C1=C(Cl)C=CC=C1)C(C)(C)C	7.13	6.45
15	CC(C)C1=CC=C(C(NN(C(C)(C)C)C)C(C2=C(Cl)C=CC=C2)=O)=O)C=C1	7.76	6.81
16	COC1=CC=C(C=C1)C(=O)NN(C(=O)C1=C(Cl)C=CC=C1)C(C)(C)C	6.47	6.06
17	CC(C)(C)N(NC(=O)C1=CC=CC=C1)C(=O)C1=CC=CC=C1	8.15	7.78
18	CC1=CC(=CC(C)=C1)C(=O)N(NC(=O)C1=CC=C(C=C1)C(C)(C)C)C(C)(C)C	7.79	8.20
19	CC1=CC(=CC(C)=C1)C(=O)N(NC(=O)C1=C(C)C(C)=CC=C1)C(C)(C)C	6.96	7.24
20	CCC1=CC=C(C=C1)C(=O)NN(C(=O)C1=CC=CC=C1)C(C)(C)C	4.66	5.14
21	CCOC1=C(C=CC=C1)C(=O)N(NC(=O)C1=CC=C(OC)C=C1)C(C)(C)C	5.02	5.00
22	CCOC1=C(C=CC=C1)C(=O)N(NC(=O)C1=CC=C(Cl)C=C1)C(C)(C)C	5.16	5.27
23	CCOC1=C(C=CC=C1)C(=O)N(NC(=O)C1=CC=C(CC)C=C1)C(C)(C)C	5.76	6.02
24	CC1=CC(C(=O)N(NC(=O)C2=CC=CC=C2)C(C)(C)C)=C(Cl)C(C)=C1	6.47	6.60
25	CCC1=CC=C(C=C1)C(=O)NN(C(=O)C1=C(Cl)C(C)=CC(C)=C1)C(C)(C)C	5.95	6.13
26	CCCCC1=CC=C(C=C1)C(=O)NN(C(=O)C1=C(Cl)C(C)=CC(C)=C1)C(C)(C)C	5.69	6.36
27	COC1=CC=C(C=C1)C(=O)NN(C(=O)C1=C(Cl)C(C)=CC(C)=C1)C(C)(C)C	5.87	6.04
28	CC1=CC=C(C=C1)C(=O)NN(C(=O)C1=C(Cl)C(C)=CC(C)=C1)C(C)(C)C	5.45	6.02
29	CCCCCC1=CC=C(C=C1)C(=O)NN(C(=O)C1=C(Cl)C(C)=CC(C)=C1)C(C)(C)C	5.97	6.10
30	CC1=CC(C(=O)N(NC(=O)C2=CC=C(Cl)C=C2)C(C)(C)C)=C(Cl)C(C)=C1	7.17	7.52
31	CC1=CC(C(=O)N(NC(=O)C2=CC=C3OCCCC3=C2C)C(C)(C)C)=C(Cl)C(C)=C1	8.27	8.35
32	CC1=C2CCCOC2=CC=C1C(=O)NN(C(=O)C1=C(Cl)C=CC=C1)C(C)(C)C	6.49	6.00
33	CCCCCC1=CC=C(C=C1)C(=O)NN(C(=O)C1=C(Cl)C=CC=C1)C(C)(C)C	5.11	4.92

Material and methods

Definition of target property and molecular structures

A series of 33 dibenzoylhydrazine analogues (Table 1) was used, having the insecticidal activity pEC_{50} measured *in vitro*, based on an ecdysone-dependent reporter assay using cell lines derived from the lepidopteran speciesthe cotton leafworm *Spodoptera littoralis*,as dependent variable.

These insecticides were energy pre-optimized by molecular mechanics calculations using the MMFF94s force field included in the OMEGA (version 2.5.1.4, OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com) software [5, 6]. Structural 0D, 1D, 2D and 3D descriptors were calculated for the minimum energy structures using the DRAGON (Dragon Professional 5.5 (2007), Talete S.R.L., Milano, Italy) and InstantJchem (which was used for structure database management, search and prediction) (InstantJchem 6.0.0, 2013, ChemAxon (http://www.chemaxon.com) software

The Partial Least Squares (PLS) method

Projections to latent structures (PLS) represent a regression technique for modeling the relationship between projections of dependent factors and independent responses. In this approach data analysis features link a block (or a column) of response variables to a block of explanatory variables [7]. The relationship between the dependent and independent variables is described as a latent variable approach [8]. The PLS approach leads to stable, correct and highly predictive models even for correlated descriptors [9].PLS calculations were performed by the SIMCA package (SIMCA P+ 12.0.0.0, May 20 2008, Umetrics, Sweeden, http://www.umetrics.com/). The QSAR matrix (including the dependent and independent matrices) was analyzed in a first step by the principal component analysis (PCA) [10],and subsequently by the partial least squares(PLS)approaches. The squared correlation regression coefficient R^2, and the squared cross-validated correlation coefficient, Q^2, are the most important statistical parameters that provide a measure of the quality and validity for the final PLS model, while the Variables Importance in the Projection (VIP) values and the sign of the variables' coefficients are more relevant in explaining the activity mechanism. The significant principal components were selected by 7 cross-validation groups.

The Y-randomization test is a widely used technique that displays the robustness of a QSAR model, being a measure of model overfit. The dependent variable (biological activity) is randomly shuffled and a QSAR model is built using the same descriptor matrix. The obtained PLS models (after 999 randomizations) must have the minimal R^2 and Q^2 values[11].

Results and discussion

A statistical analysis of the dibenzoylhydrazine analogues was performed using the calculated variables.A PCA model was built for the whole X matrix (including N=33 compounds and X = 1462 descriptors). From the total of 7 significant principal components resulted from this analysis, the first three components already explained 61.7% of the information content of the descriptor matrix. PLS calculations were, also, performed using the same program forthe same dataset.

Table 2. The coefficients in descending order of VIP values for the two principal components of model M2.

No	Variable ID*	CoefCS[2]	VIP[2]	No	Variable ID*	CoefCS[2]	VIP[2]
1	BEHe2	0.07	0.98	8	Mor32p	0.20	1.22
2	BELm1	0.10	1.01	9	Mor32v	0.19	1.21
3	F09[C-C]	0.05	0.94	10	R4u	0.14	1.07
4	G3m	0.17	0.88	11	R5u	0.12	0.95
5	Infective-80	0.16	0.94	12	RDF085v	0.08	0.84
6	Mor02m	0.11	0.80	13	VEA1	0.11	0.97
7	Mor29e	0.17	1.01	14	VEA2	0.20	1.08

*BEHe2-highest eigenvalue n. 2 of Burden matrix / weighted by atomic Sanderson electronegativities, BELm1-lowest eigenvalue n. 1 of Burden matrix / weighted by atomic masses, F09[C-C]-frequency of C-C at topological distance 9, G3m-3st component symmetry directional WHIM index / weighted by atomic masses, Infective-80 - Ghose-Viswanadhan-Wendoloski antiinfective-like index at 80%, Mor02m-3D-MoRSE - signal 02 / weighted by atomic masses, Mor29e-3D-MoRSE - signal 29 / weighted by atomic Sanderson electronegativities, Mor32p-3D-MoRSE - signal 32 / weighted by atomic polarizabilities, Mor32v-3D-MoRSE - signal 32 / weighted by atomic van der Waals volumes, R4u-R autocorrelation of lag 4 / unweighted,R autocorrelation of lag 5 / unweighted, R5u-Radial Distribution Function - 8.5 / weighted by atomic van der Waals volumes, VEA1-eigenvector coefficient sum from adjacency matrix, VEA2-average eigenvector coefficient sum from adjacency matrix

The statistical results of the PLS model: $R^2_Y(CUM) = 0.837$ and $Q^2(CUM) = 0.613$ obtained for three principal components demonstrated the model overfit ($R^2_{X(CUM)}$ and $R^2_{Y(CUM)}$ are the cumulative sum of squares of all the X and Y values). This inconvenience was overstepped by excluding the noise variables from this model (e.g. coefficient values insignificantly different from 0). Thus, a robust model, M2 (N= 33 and X= 14) with two latent variables, which explaines70.5% of the information content of the descriptor matrix, $R^2_Y(CUM) = 0.821$ and $Q^2(CUM) = 0.793$ was obtained.
All selected variables in M2(Table 2) had VIP values greater than 1 and were considered to be the most relevant for the model.For the test set the Y-randomization procedure was applied using the SIMCA-P+ 12.0 software (for the final PLS model). It gave the following intercept (PLS) values of the regression lines obtained by the correlation between the calculated R^2, respectively Q^2 values of the original Y-variable and the shuffled Y-variable, respectively: 0.134 for the R^2_Y line and -0.249 for the Q^2_Y line. The slope values close to zero indicate stable models.

Conclusion

The final modelof dibenzoylhydrazine non-steroidal ecdysone agonists obtained using the PLS method have good statistical parameters.The most important molecular descriptors for the insecticidal activity are related to the geometric representation of molecules, providing information on interatomic and topological distances, structural fragments, descriptors sensitive to any conformational change, antiinfective drug-like index having a qualifying range that covers approximately 80% of the drugs studied.

Acknowledgements

This project was financially supported by Project 1.1 of the Institute of Chemistry of the Romanian Academy. Access to the Chemaxon Ltd. and OpenEye Ltd. software are greatly acknowledged by the authors.

References

[1] L. Swevers, T. Soin, H. Mosallanejad, K. Iatrou, G. Smagghe, Insect Biochem. 38 (2008) 825.
[2] T. Soin, M. Iga, L. Swevers, P. Rougé, C.R. Janssen, G. Smagghe, Insect Biochem. 39 (2009) 523.
[3] N. Ferro, J.E. Tacoronte, T. Reinard, P. Bultinck, L.A. Montero, J. Mol. Struct.-THEOCHEM 758 (2006) 263.
[4] T. Soin, E. De Geyter, H. Mosallanejad, M. Iga, D. Martín, S. Ozaki, S. Kitsuda, T. Harada, H. Miyagawa, D. Stefanou, G. Kotzia, R. Efrose, V. Labropoulou, D. Geelen, K. Iatrou, Y. Nakagawa, C.R. Janssen, G. Smagghe, L. Swevers, Pest. Manag. Sci.66 (2010) 526.
[5]. P.C.D. Hawkins, G.A. Skillman, G.L. Warren, B.A. Ellingson, M.T. Stahl, J. Chem. Inf. Model. 50(2010) 572.
[6]. P.C.D. Hawkins, A. Nicholls, J. Chem. Inf. Model. 52 (2012) 2919.
[7] H. Wold, in: S. Kotz and N.L. Johnson (Eds.), Encyclopedia of Statistical Sciences, (Vol. 6), Wiley, New York, 1985, pp. 581.
[8] L. Eriksson, J. Gottfries, E. Johansson, S. Wold, Chemom. Intel. Lab. Syst.73 (2004) 73.
[9] A. Höskuldsson, J. Chemometrics 2(3) (1998) 211.
[10] M. Daszykowski, K. Kaczmarek, V. Heyden, B. Walczak, Chemom. Intell .Lab. Syst. 85 (2007) 203.
[11] P.P. Roy, S. Paul, I. Mitra, K. Roy, Molecules 14(2009) 1660.