Emerging Contaminants 1 (2015) 8–13

Contents lists available at ScienceDirect

Emerging Contaminants

journal homepage: http://www.keaipublishing.com/en/journals/emerging-contaminants/

Research article

A comparative QSAR study on the estrogenic activities of persistent organic pollutants by PLS and SVM

Fei Li*, Jialin Liu, Lulu Cao

Key Laboratory of Coastal Environmental Processes and Ecological Remediation, Yantai Institute of Coastal Zone Research (YIC), Chinese Academy of Sciences(CAS), Shandong Provincial Key Laboratory of Coastal Environmental Processes, YICCAS, Yantai, Shandong, 264003, PR China

ARTICLE INFO

Article history:
Received 23 March 2015
Received in revised form
12 May 2015
Accepted 13 May 2015
Available online 29 June 2015

Keywords:
Persistent organic pollutants (POPs)
Estrogen receptor (ER)
Quantitative structure activity relationship (QSAR)
Partial least square (PLS)
Support vector machine (SVM)

ABSTRACT

Quantitative structure-activity relationships (QSARs) were determined using partial least square (PLS) and support vector machine (SVM). The predicted values by the final QSAR models were in good agreement with the corresponding experimental values. Chemical estrogenic activities are related to atomic properties (atomic Sanderson electronegativities, van der Waals volumes and polarizabilities). Comparison of the results obtained from two models, the SVM method exhibited better overall performances. Besides, three PLS models were constructed for some specific families based on their chemical structures. These predictive models should be useful to rapidly identify potential estrogenic endocrine disrupting chemicals.

1. Introduction

Persistent organic pollutants (POPs) are ubiquitous and bioaccumulate in the environment, in wildlife, up through the food web and human beings [1]. Some POPs are classified as endocrine-disrupting chemicals (EDCs) that are compounds that can mimic, interfere or block the function of endogenous hormones and thereby disrupt the normal hormone homeostasis of the body. Consequently, it is necessary to screen and determine the EDCs.

It remains a labor intensive and time-costing determination considering a large number of potential EDCs. It is crucial to develop efficient and economical alternative modeling approaches for the purpose of predicting the estrogenic activities of potential EDCs. Quantitative structure activity relationships (QSARs) are promising and successful tools to provide a rapid and useful meanings for predicting the biological activity and chemical toxicity. They are considered as an important part of the priority setting process by the endocrine disruptor screening and testing advisory committee [2].

QSARs have been applied to study the mechanism of chemicals' binding for the estrogen receptors (ER) [3–6], androgen receptor and several other members of the nuclear receptor family [7]. These include PLS models, comparative molecular field analysis which considers the overall steric and electrostatic properties of the compound of interest, computer graphic and energy (electrostatic and van der Waals) based models to fit into DNA and common reactivity patterns which reflect the stereoelectronic features.

In this study, a data set consisted of experimental values which were determined by Nishihara et al. [8], including more than 500 natural, synthetic, and environmental chemicals from a broad range of structural classes. The data set was used to construct global QSAR models for the whole data set and local models for specific well-known families. Some structural descriptors were selected using Forward stepwise (FS) regression from the original DRAGON calculated descriptors and were applied to construct an optimal model based on partial least square (PLS). Another classical method, support vector machine (SVM) [9] was utilized to establish QSAR model to compare the results with those obtained by PLS. Additionally, three QSAR models for specific well-known families were examined in conjunction with knowledge of the recently reported ligand-ER crystal structures.

* Corresponding author.
E-mail address: fli@yic.ac.cn (F. Li).
Peer review under responsibility of KeAi Communications Co., Ltd.

Production and Hosting by Elsevier on behalf of KeAi

http://dx.doi.org/10.1016/j.emcon.2015.05.003

2. Materials and methods

2.1. Experiment and data set

The overall data set consisted of more than 500 organic chemicals, including natural substances, medicine, pesticides, and industrial chemicals [8]. Table 1 shows a summary of 55 positive compounds. Tested chemicals consisted of natural substances (metabolites, oxidation products, etc.), medicines, food additives, pesticides, and industrial chemicals (PCBs, PCDFs, PAHs, phenols, benzenes, phthalates and adipates, and others). The estrogenic activities to the ER, expressed as log unit of 10% relative effective concentration (logREC_{10}), are listed in Table 1.

2.2. Descriptors generation and selection

Structures of chemicals were geometry-optimized with the PM3 Hamiltonian using the software package Chemoffice 8.0 program,

Table 1
Observed and predicted logREC_{10} for the QSAR model.

No	Compounds	Observed logREC_{10}	Predicted logREC_{10}	
			PLS	*SVM*
A. natural products and related				
1	17α-Estradiol	3.13	3.46	2.91
2	Apigenin	6.52	6.00	6.30
3	Coumestrol	6.52	5.88	6.31
4	Daidzein	5.00	5.34	4.84
5	Dihydrogenistein	5.00	4.08	4.75
6	Equol	6.52	4.90	5.15
7	Estrone	2.00	3.80	3.37
8	Genistein	4.52	5.08	4.74
B. medicines, food additives, and related				
9	17α-Ethynylestradiol	1.82	2.71	3.50
10	β-Estradiol-17-acetate	5.22	3.06	4.10
11	Diethylstilbesterol (DES)	1.82	2.50	2.04
12	Ethyl 4-hydroxybenzoate	7.52	6.84	7.30
13	Methyl 4-hydroxybenzoate	8.13	7.04	7.91
14	n-Butyl 4-hydroxybenzoate	6.00	5.66	6.31
15	n-Propyl 4-hydroxybenzoate	6.52	6.09	6.43
C. PCBs, PCDFs, PAHs, and related				
16	2-Hydroxy benzo[a]pyrene	7.22	6.34	6.48
17	2-Hydroxy fluorene	7.52	7.48	7.59
18	3,8-Dihydroxy-2-chlorodibenzofuran	5.43	6.22	5.88
19	3-Hydroxy benzo[a]pyrene	6.52	7.22	6.43
20	4-Hydroxy-2′,4′,6'-trichlorobiphenyl	5.13	6.19	6.08
21	4-Hydroxy-2′,4′,6'-trichlorobiphenyl	6.30	6.19	6.08
22	8-Hydroxy-2,3,4-trichlorodibenzofuran	6.52	7.04	6.74
23	8-Hydroxy-2-monochlorodibenzofuran	6.52	7.22	6.67
24	8-Hydroxy-3,4,6-trichlorodibenzofuran	6.52	7.13	6.30
25	8-Hydroxy-3,4-dichlorodibenzofuran	6.22	7.17	6.40
26	8-Hydroxy-3-monochlorodibenzofuran	6.37	7.24	6.54
D. Phenols				
27	2,2-Bis(4-hydroxy-3-methylphenyl)propane	6.00	5.55	5.90
28	2,2-Bis(4-hydroxy-phenyl)butane	6.00	5.68	5.78
29	2,4-Dichlorophenol	7.13	7.18	6.92
30	3,4-Dichlorophenol	6.82	7.43	7.30
31	4,4′-Dihydroxybenzophenone	8.00	7.21	7.29
32	4,4′-Dihydroxybiphenyl	6.22	6.41	6.37
33	4,4′-Thiobiphenyl	6.00	5.99	5.78
34	4-Bromophenol	7.43	7.45	7.21
35	4-Chloro-3,5-xylenol	7.52	7.66	7.74
36	4-Chloro-3-methylphenol	7.22	7.36	7.25
37	4-Chlorophenol	7.82	7.54	7.61
38	4-Ethylphenol	7.00	7.33	7.32
39	4-Hydroxyacetophenone	7.82	7.60	7.92
40	4-Hydroxybiphenyl	7.00	7.46	7.48
41	4-Methylphenol (p-cresol)	8.00	7.52	7.60
42	4-*n*-Butylphenol	6.52	7.16	6.74
43	4-*n*-Hexylphenol	6.52	6.85	6.31
44	4-*n*-Pentylphenol	6.00	6.93	6.64
45	4-*n*-Propylphenol	7.37	7.11	6.88
46	4-*sec*-Butylphenol	6.52	7.16	6.98
47	4-*tert*-Butylphenol	7.00	6.21	6.18
48	4-*tert*-Octylphenol	4.82	5.81	5.66
49	4-*tert*-Pentylphenol	5.52	5.90	5.74
50	Bis(4-hydroxyphenyl)methane	6.82	6.98	7.10
51	Bisphenol A	6.00	5.71	6.02
E. Benzenes and heterocyclics				
52	*cis*-1,2-Diphenylcyclobutane	8.00	7.57	7.78
F. Phthalates and adipates				
53	Benzylbutyl phthalate (BBP)	8.22	7.13	7.83
54	Di-*iso*-propyl phthalate	8.82	7.84	8.61
55	Di-*n*-propyl phthalate	8.52	7.66	8.31