CODEN IEJMAT *Internet Electronic Journal of Molecular Design* **2005**, *4*, 181–193 ISSN 1538–6414
BioChem Press http://www.biochempress.com

Inter*net* Electronic Journal of
Molecular Design

February 2005, Volume 4, Number 2, Pages 181–193

Editor: Ovidiu Ivanciuc

Support Vector Regression Quantitative Structure–Activity Relationships (QSAR) for Benzodiazepine Receptor Ligands

Ovidiu Ivanciuc

Sealy Center for Structural Biology, Department of Human Biological Chemistry & Genetics, University of Texas Medical Branch, Galveston, Texas 77555–0857

Received: August 29, 2004; Revised: November 26, 2004; Accepted: December 4, 2004; Published: February 28, 2005

Citation of the article:
O. Ivanciuc, Support Vector Regression Quantitative Structure–Activity Relationships (QSAR) for Benzodiazepine Receptor Ligands, *Internet Electron. J. Mol. Des.* **2005**, *4*, 181–193, http://www.biochempress.com.

Inter*net* Electronic Journal of Molecular Design

BioChem Press
http://www.biochempress.com

Support Vector Regression Quantitative Structure–Activity Relationships (QSAR) for Benzodiazepine Receptor Ligands

Ovidiu Ivanciuc *

Sealy Center for Structural Biology, Department of Human Biological Chemistry & Genetics, University of Texas Medical Branch, Galveston, Texas 77555–0857

Received: August 29, 2004; Revised: November 26, 2004; Accepted: December 4, 2004; Published: February 28, 2005

Internet Electron. J. Mol. Des. **2005**, *4* (2), 181–193

Abstract

Support vector machines were developed by Vapnik as an effective algorithm for determining an optimal hyperplane to separate two classes of patterns. Comparative studies showed that support vector classification (SVC) usually gives better predictions than other classification methods. In a short period of time SVC found significant applications in bioinformatics and computational biology, such as cancer diagnosis, prediction of protein fold, secondary structure, protein–protein interactions, and subcellular localization. Using various loss functions, the support vector method was extended for regression (support vector regression, SVR). SVR can have significant applications in QSAR (quantitative structure–activity relationships) if it is able to predict better than other well–established QSAR models. In this study we compare QSAR models obtained with multiple linear regression (MLR) and SVR for the benzodiazepine receptor affinity using a set of 52 pyrazolo[4,3–*c*]quinolin–3–ones. Both models were developed with five structural descriptors, namely the Hammett electronic parameter $\sigma_{R'}$, the molar refractivity MR_{R8}, the Sterimol parameter $L_{R'4'}$, an indicator variable I (1/0) for 7–substituted compounds, and the Sterimol parameter B_{5R}. Extensive simulations using the dot, polynomial, radial basis function, neural, and anova kernels show that the best predictions are obtained with the neural kernel. The prediction power of the QSAR models was tested with complete cross–validation: leave–one–out, leave–5%–out, leave–10%–out, leave–20%–out, and leave–25%–out. While for the leave–one–out test SVR is better than MLR ($q^2_{LOO,MLR}$ = 0.481, $RMSE_{LOO,MLR}$ = 0.82; $q^2_{LOO,SVR}$ = 0.511, $RMSE_{LOO,SVR}$ = 0.80), in the more difficult test of leave–25%–out, MLR is better than SVR ($q^2_{L25\%O,MLR}$ = 0.470, $RMSE_{L25\%O,MLR}$ = 0.83; $q^2_{L25\%O,SVR}$ = 0.432, $RMSE_{L25\%O,SVR}$ = 0.86). The results obtained in the present study indicate that SVR applications in QSAR must be compared with other models, in order to determine if their use brings any prediction improvement. Despite many over–optimistic expectations, support vector regression can overfit the data, and SVR predictions may be worse than those obtained with linear models.

Keywords. Support vector machines; support vector regression; benzodiazepine receptor; quantitative structure–activity relationships; QSAR.

1 INTRODUCTION

Benzodiazepine receptor (BzR) ligands (either benzodiazepines or structurally unrelated chemical compounds) act as modulators of the γ–aminobutyric acid (GABA) binding to its receptor, by altering the transmembrane chloride ion conductance [1–4]. The interest for developing new

* Correspondence author; E–mail: ovidiu_ivanciuc@yahoo.com.

BzR ligands is stimulated by their ability to induce a wide spectrum of central effects, from full agonism through antagonism to inverse agonism. Full agonists have anticonvulsant, sedative, anxiolytic, amnesic effects. Full inverse agonists have proconvulsant, anxiogenic, cognition enhancement, reversal of alcohol effects. Antagonists have a selective blockade of the effects of both agonists and inverse agonists.

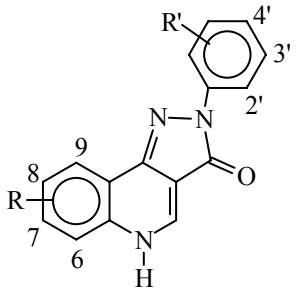

Figure 1. General formula for the pyrazolo[4,3–*c*]quinolin–3–ones **1**–**52**.

The search for BzR specific ligands resulted in the synthesis of a wide diversity of compounds, many of them structurally not related to benzodiazepine. A recent review presents quantitative structure–activity relationships (QSAR) derived from 66 sets of non–benzodiazepine BzR ligands [1]. From these data sets, we have selected a group of 52 pyrazolo[4,3–*c*]quinolin–3–ones [2] (Figure 1, Table 1) to compare multiple linear regression (MLR) with a new multivariate procedure, support vector regression (SVR).

Support vector machines were introduced by Vapnik [5–7] as a powerful tool for pattern classification in two classes by determining an optimal hyperplane that separates the classes [8–15]. The SVM algorithm generates a separating hypersurface in the input space that optimally separates two classes of patterns. In the first step, using various kernels that perform a nonlinear mapping, the input space is transformed into a higher dimensional feature space. Then, a maximal margin hyperplane (MMH) is computed in the feature space. MMH maximizes the distance to the hyperplane of the closest patterns from the two classes. While initially support vector machines were developed for data classification, the algorithm was extended to regression by defining alternative loss functions (quadratic, Laplace, Huber, or ε–insensitive) [16–21]. SVM found interesting applications in bioinformatics and computational biology, such as for the classification of protein domain–architecture [22], brain tumor [23], cancer [24], protein class [25], cysteines [26], viral proteins [27], enzyme class [28], protein–protein interaction [29], subcellular localization [30], subcellular localization [31], membrane protein [32], and transmembrane segments [33]. The SVM applications is cheminformatics and computational chemistry are also mainly related to classification of various chemical species: odor of pyrazines [34], urine samples [35], toxicants [36], carcinogenic polycyclic aromatic hydrocarbons [37], cancer biomarkers [38], mutagenic compounds [39], drug–like substances [40], tea [41], fragrance [42], organophosphate nerve agents

[43], mechanism of toxic action [44–46].

The quantitative modeling and prediction of physical, chemical and biological properties of chemical compounds is usually made with regression methods, and several support vector regression applications in cheminformatics are reported [47–49]. While SVR seems attractive for QSAR applications, the method is relatively new and few comparative studies provide support for support vector regression, in comparison with other well–established QSAR models. In this study we compare QSAR models obtained with MLR and SVR for the benzodiazepine receptor affinity using a set of 52 pyrazolo[4,3–*c*]quinolin–3–ones [1,2].

2 MATERIALS AND METHODS

2.1 Chemical Data

In a recent study, a series of 2–aryl(heteroaryl)–2,5–dihydropyrazolo[4,3–*c*]quinolin–3–(3*H*)–ones were tested as benzodiazepine receptor ligands [2]. A Hansch–type QSAR was developed for 52 compounds from this series [1], using as descriptors the following substituent indices: the Hammett electronic parameter for the substituent R', $\sigma_{R'}$; the molar refractivity for the R_8 substituent, MR_{R8}; the Sterimol parameter L for the R'$_{4'}$ substituent, $L_{R'4'}$; an indicator variable I (1/0) for 7–substituted compounds; the Sterimol B_5 parameter for the R substituent, B_{5R}. From the large number of QSAR models for BzR ligands reported in [1] we selected this data set because it has a fairly large number of compounds, and the MLR correlation coefficient (r_{cal} = 0.798) is not too high and might be improved by non–linear regressions if the relationship between structural descriptors and the biological activity is non–linear. The compounds from Table 1 were tested for their ability to displace [^{3}H]–flunitrazepam binding from rate brain membranes. The structure of the chemical compounds, values for the five theoretical descriptors, and the biological activity (log $1/IC_{50}$) are presented in Table 1.

2.2 Support Vector Regression

All SVR models from the present paper were obtained with mySVM [50], which is freely available for download. Before computing the SVM model, the input vectors were scaled to zero mean and unit variance. The prediction power of the QSAR models (MLR and SVR) was tested with complete cross–validation: leave–one–out (LOO), leave–5%–out (L5%O), leave–10%–out (L10%O), leave–20%–out (L20%O), and leave–25%–out (L25%O). The capacity parameter C was optimized for each SVM model. The influence of the kernel type on the SVM performances was extensively explored using various kernels, namely the dot, polynomial, radial basis function, neural, and anova kernels. We present below the kernels and their parameters used in this study. The SVR performance is greatly influenced by the kernel type and parameters.

Table 1. Structure of the chemical compounds (see Figure 1), theoretical descriptors (the Hammett electronic parameter $\sigma_{R'}$, the molar refractivity MR_{R8}; the Sterimol parameter $L_{R'4'}$; an indicator variable I (1/0) for 7–substituted compounds; the Sterimol parameter B_{5R}), and binding data for the displacement of [^{3}H]–flunitrazepam, log 1/IC_{50} [1].

No	R	R'	$\sigma_{R'}$	MR_{R8}	$L_{R'4'}$	I	B_{5R}	log 1/IC_{50}
1	H	H	0	0.103	2.06	0	1	9.35
2	H	4–Cl	0.23	0.103	3.52	0	1	9.00
3	H	4–OCH_3	–0.27	0.103	3.98	0	1	9.17
4	6–F	H	0	0.103	2.06	0	1.35	8.16
5	6–CF_3	H	0	0.103	2.06	0	3.07	5.73
6	6–OCH_3	H	0	0.103	2.06	0	1	5.66
7	8–F	H	0	0.092	2.06	0	1	9.54
8	8–F	3–NH_2	–0.16	0.092	2.06	0	1	9.26
9	8–F	4–OCH_3	–0.27	0.092	3.98	0	1	9.48
10	8–F	4–OH	–0.37	0.092	2.74	0	1	9.34
11	8–Cl	H	0	0.600	2.06	0	1	9.37
12	8–OCH_3	H	0	0.787	2.06	0	1	9.17
13	8–OC_2H_5	H	0	1.247	2.06	0	1	8.85
14	8–C_4H_9	H	0	1.959	2.06	0	1	9.00
15	8–C_4H_9	4–COOH	0.45	1.959	3.91	0	1	5.93
16	8–cyC_6H_{11}	H	0	2.669	3.91	0	1	8.35
17	8–cyC_6H_{11}	4–COOH	0.45	2.669	3.91	0	1	5.55
18	8–$OCH_2C_6H_5$	H	0	3.219	2.06	0	1	7.75
19	8–OCF_3	H	0	0.786	2.06	0	1	9.15
20	8–OCF_3	2–F	0.06	0.786	2.06	0	1	9.40
21	8–OCF_3	2–Cl	0.23	0.786	2.06	0	1	8.60
22	8–OCF_3	2–CH_3	–0.17	0.786	2.06	0	1	8.47
23	8–OCF_3	3–Br	0.39	0.786	3.82	0	1	7.46
24	8–OCF_3	3–CH_3	–0.07	0.786	2.87	0	1	8.20
25	8–OCF_3	3–Cl	0.37	0.786	3.52	0	1	7.62
26	8–OCF_3	3–NO_2	0.71	0.786	3.44	0	1	7.20
27	8–OCF_3	3–NH_2	–0.16	0.786	2.78	0	1	9.62
28	8–OCF_3	4–Br	0.23	0.786	3.82	0	1	7.82
29	8–OCF_3	4–CH_3	–0.17	0.786	2.87	0	1	8.79
30	8–OCF_3	4–Cl	0.23	0.786	3.52	0	1	7.90
31	8–OCF_3	4–F	0.06	0.786	2.65	0	1	9.00
32	8–OCF_3	4–NO_2	0.78	0.786	3.44	0	1	7.40
33	8–OCF_3	4–OCH_3	–0.27	0.786	3.98	0	1	9.22
34	8–OCF_3	4–OH	–0.37	0.786	2.74	0	1	9.63
35	9–OH	H	0	0.103	2.06	0	1	9.62
36	9–OCH_3	H	0	0.103	2.06	0	1	8.84
37	6,8–F	H	0	0.092	2.06	0	1.35	7.87
38	6,8–F	3–F	0.34	0.092	2.06	0	1.35	8.02
39	6,8–F	4–Br	0.23	0.092	3.82	0	1.35	6.79
40	6,8–F	4–OCH_3	–0.27	0.092	3.98	0	1.35	8.12
41	6,8–F	2–pyridyl–2′–yl	0.17	0.092	2.06	0	1.35	7.82
42	6,8–F	2–pyrimidyl–2′–yl	0.53	0.092	2.06	0	1.35	6.47
43	7,9–Cl	H	0	0.103	2.06	1	1	8.43
44	6,7,8–F	H	0	0.092	2.06	1	1.35	7.70
45	6,7,8–F	4–CH_3	–0.17	0.092	2.87	1	1.35	7.15
46	6,7,8–F	4–Cl	0.23	0.092	3.52	1	1.35	7.13
47	6,7,8–F	4–F	0.06	0.092	2.65	1	1.35	7.68
48	6,7,8–F	4–OCH_3	–0.27	0.092	3.98	1	1.35	8.14
49	7,8,9–OCH_3	H	0	0.787	2.06	1	1	8.90
50	7,8,9–OCH_3	4–COOH	0.45	0.787	3.91	1	1	5.52
51	7,8,9–OCH_3	2–pyridyl–2′–yl	0.17	0.787	2.06	1	1	8.50
52	7,8,9–OCH_3	2–pyrimidyl–2–yl	0.53	0.787	2.06	1	1	7.24

The dot kernel. The inner product of x and y defines the dot kernel:

$$K(x, y) = x \cdot y \tag{1}$$

The polynomial kernel. The polynomial of degree d (values 2, 3, 4, and 5) in the variables x and y defines the polynomial kernel:

$$K(x, y) = (x \cdot y + 1)^d \tag{2}$$

The radial kernel. The following exponential function in the variables x and y defines the radial basis function kernel, with the shape controlled by the parameter γ (values 0.25, 0.5, 1.0, 1.5, and 2.0):

$$K(x, y) = \exp(-\gamma \parallel x - y \parallel^2) \tag{3}$$

The neural kernel. The hyperbolic tangent function in the variables x and y defines the neural kernel, with the shape controlled by the parameters a (values 0.5, 1.0, and 2.0) and b (values 0, 1, and 2):

$$K(x, y) = \tanh(ax \cdot y + b) \tag{4}$$

The anova kernel. The sum of exponential functions in x and y defines the anova kernel, with the shape controlled by the parameters γ (values 0.25, 0.5, 1.0, 1.5, and 2.0) and d (values 1, 2, and 3):

$$K(x, y) = \left(\sum_i \exp(-\gamma(x_i - y_i)) \right)^d \tag{5}$$

3 RESULTS AND DISCUSSION

3.1 Multiple Linear Regression QSAR

The calibration and cross–validation results for the MLR QSAR model are presented in Eq. (6). The values in parenthesis represent the confidence interval for the MLR parameters at the 95% level. For each cross–validation experiment we present the correlation coefficient r, q^2, and root mean square error RMSE.

$$\begin{gathered}
\log 1/IC_{50} = 11.538(\pm 2.869) - 2.320(\pm 0.577)\, \sigma_{R'} - 0.294(\pm 0.073)\, MR_{R8} - 0.326(\pm 0.081)\, L_{R'4'} \\
- 0.560(\pm 0.139)\, I - 1.795(\pm 0.446)\, B_{5R} \\
n = 52 \quad r_{cal} = 0.798 \quad RMSE_{cal} = 0.69 \quad s_{cal} = 0.73 \quad F_{cal} = 16.18 \\
r_{LOO} = 0.721 \quad q^2_{LOO} = 0.481 \quad RMSE_{LOO} = 0.82 \\
r_{L5\%O} = 0.716 \quad q^2_{L5\%O} = 0.458 \quad RMSE_{L5\%O} = 0.84 \\
r_{L10\%O} = 0.711 \quad q^2_{L10\%O} = 0.448 \quad RMSE_{L10\%O} = 0.85 \\
r_{L20\%O} = 0.733 \quad q^2_{L20\%O} = 0.502 \quad RMSE_{L20\%O} = 0.81 \\
r_{L25\%O} = 0.712 \quad q^2_{L25\%O} = 0.470 \quad RMSE_{L25\%O} = 0.83
\end{gathered} \tag{6}$$

The partial correlation coefficients are: $r(\sigma_{R'}) = -0.587$, $r(MR_{R8}) = -0.146$, $r(L_{R'4'}) = -0.262$, $r(I) = -0.221$, $r(B_{5R}) = -0.403$. These values show that the Hammett electronic substituent is the most

important descriptor, followed by B_5 for *R*. All five descriptors have significant inverse correlation with log $1/IC_{50}$, and the intercorrellation between descriptors is small, as one can see from the following matrix of intercorrellation coefficients:

	$\sigma_{R'}$	MR_{R8}	$L_{R'4'}$	*I*	B_{5R}
$\sigma_{R'}$	1	0.20	0.13	0.05	–0.04
MR_{R8}	0.20	1	0.16	–0.18	–0.30
$L_{R'4'}$	0.13	0.16	1	–0.02	–0.11
I	0.05	–0.18	–0.02	1	0.09
B_{5R}	–0.04	–0.30	–0.11	0.09	1

The prediction statistics will be compared with those obtained from SVR. RMSE increases from 0.69 in calibration to 0.82 for LOO, to 0.84 in L5%O, and to 0.85 in L10%O. For L20%O RMSE decreases to 0.81 and then increases to 0.83 for L25%O.

Theoretically, we expect RMSE to increase and q^2 to decrease for the sequence of cross–validation experiments: LOO, L5%O, L10%O, L20%O, and L25%O. The unexpected RMSE decrease for the cross–validation experiments that have the largest perturbation of the MLR model (L20%O and L25%O) can be a result of a random improvement in the prediction for a particular set of calibration data.

3.2 Support Vector Regression QSAR

Support vector regression gives calibration results that greatly improve those obtained with MLR (Table 2). The highest SVR calibration correlation coefficient (0.896 *vs*. 0.798 for MLR) is obtained with the anova kernel in experiments 29 and 31–34. Overall, high calibration *r* is obtained with the anova, radial basis and polynomial (degrees 4 and 5) kernels. As expected, the dot kernel has statistics close to those of the MLR model. The neural kernel has low calibration statistics, with *r* between 0.699 and 0.055.

As Eq. (6) indicates, the MLR model is stable in the LOO cross–validation test (r_{cal} = 0.798 and r_{LOO} = 0.721), with a small decrease for *r*, as expected. The comparison between calibration and LOO statistics for SVR shows dramatic changes for the polynomial, radial, and anova kernels (Table 2). The LOO statistics for these kernels indicate that the corresponding SVR models are not able to make reliable predictions. To have a perception of the inability of these SVR models in prediction, we consider, for each of these kernels, the experiment with maximum r_{cal}: polynomial kernel, experiment 5, r_{cal} = 0.895 and r_{LOO} = – 0.294; radial kernel, experiment 10, r_{cal} = 0.894 and r_{LOO} = 0.452; anova kernel, experiment 31, r_{cal} = 0.896 and r_{LOO} = 0.176. Compared with MLR LOO results, all these three kernels have very bad prediction statistics (r_{LOO}, q^2_{LOO}, and $RMSE_{LOO}$). These results are important, because they indicate that the SVR models can easily overfit the data, and a significant effort should be invested in evaluating the prediction ability of different kernels. The LOO results for the dot kernel (r_{LOO} = 0.668) are slightly worse than those obtained with MLR,

but they significantly outperform the polynomial, radial, and anova kernels. The LOO prediction results for the neural kernel are also worse than the MLR predictions, with the exception of those obtained in the experiment 11 (r_{LOO} = 0.740, q^2_{LOO} = 0.511, and $RMSE_{LOO}$ = 0.80). These results are slightly better than those from MLR LOO (r_{LOO} = 0.721, q^2_{LOO} = 0.481, and $RMSE_{LOO}$ = 0.82), but we have to consider that this SVR model is the only one, from the set of 34 experiments, that outperforms the MLR QSAR.

Table 2. Kernel type and corresponding parameters for each experiment (Exp), calibration statistics (r_{cal} and $RMSE_{cal}$) and leave–one–out statistics (r_{LOO}, q^2_{LOO}, and $RMSE_{LOO}$). Five kernels were tested: dot D; polynomial P (parameter: degree *d*); radial basis function R (parameter: γ); neural N (parameters: *a* and *b*); anova A (parameters: γ and *d*).

Exp	Kernel			r_{cal}	$RMSE_{cal}$	r_{LOO}	q^2_{LOO}	$RMSE_{LOO}$
1	D			0.793	0.72	0.668	0.275	0.97
2	P	2		0.837	0.64	–0.267	<–100	>10
3	P	3		0.876	0.58	–0.287	<–100	>10
4	P	4		0.894	0.52	–0.295	<–100	>10
5	P	5		0.895	0.52	–0.294	<–100	>10
6	R	0.25		0.885	0.54	0.651	0.348	0.92
7	R	0.5		0.889	0.53	0.609	0.302	0.95
8	R	1.0		0.893	0.52	0.527	0.221	1.01
9	R	1.5		0.894	0.52	0.497	0.214	1.01
10	R	2.0		0.894	0.52	0.452	0.174	1.04
11	N	0.5	0.0	0.690	0.85	0.740	0.511	0.80
12	N	1.0	0.0	0.596	0.94	0.618	0.355	0.92
13	N	2.0	0.0	0.654	0.90	0.636	0.379	0.90
14	N	0.5	1.0	0.612	0.93	0.581	0.333	0.93
15	N	1.0	1.0	0.676	0.86	0.678	0.413	0.87
16	N	2.0	1.0	0.684	0.88	0.498	0.197	1.02
17	N	0.5	2.0	0.236	1.11	0.453	0.197	1.02
18	N	1.0	2.0	0.055	1.76	0.292	–0.062	1.17
19	N	2.0	2.0	0.699	0.91	0.645	0.358	0.91
20	A	0.25	1	0.822	0.66	0.656	0.389	0.89
21	A	0.5	1	0.841	0.63	0.639	0.332	0.93
22	A	1.0	1	0.861	0.59	0.699	0.428	0.86
23	A	1.5	1	0.863	0.59	0.652	0.326	0.94
24	A	2.0	1	0.868	0.58	0.635	0.286	0.96
25	A	0.25	2	0.886	0.54	0.258	–2.017	1.98
26	A	0.5	2	0.894	0.52	0.238	–1.728	1.88
27	A	1.0	2	0.895	0.52	0.298	–0.672	1.47
28	A	1.5	2	0.895	0.52	0.385	–0.285	1.29
29	A	2.0	2	0.896	0.52	0.444	–0.098	1.19
30	A	0.25	3	0.895	0.52	0.168	–3.567	2.44
31	A	0.5	3	0.896	0.52	0.176	–1.466	1.79
32	A	1.0	3	0.896	0.52	0.371	–0.127	1.21
33	A	1.5	3	0.896	0.52	0.422	–0.010	1.15
34	A	2.0	3	0.896	0.52	0.449	0.024	1.13

In Tables 3 and 4 we present the SVR cross–validation results obtained for the L5%O, L10%O, L20%O, and L25%O tests. The cross–validation sets of compounds are identical for MLR and SVR, which is convenient for the comparison of the statistical indices. For the L5%O test, the evaluation is made with the MLR results: $r_{L5\%O}$ = 0.716, $q^2_{L5\%O}$ = 0.458, and $RMSE_{L5\%O}$ = 0.84. The results from Table 3 indicate that, by far, the predictions of the polynomial kernel are the worst. An uneven performance is identified for the anova kernel, with reasonable predictions for

experiments 20–24, and then with large errors for experiments 25–34. The best prediction with an anova kernel is obtained in experiment 22, with $r_{L5\%O}$ = 0.689, $q^2_{L5\%O}$ = 0.398, and $RMSE_{L5\%O}$ = 0.88. The results of the radial kernel show a decrease quality in going from experiment 6 to experiment 10, with the best prediction for experiment 6: $r_{L5\%O}$ = 0.665, $q^2_{L5\%O}$ = 0.368, and $RMSE_{L5\%O}$ = 0.91. These results are slightly lower than those of the best anova kernel, but also slightly better than those of the dot kernel: $r_{L5\%O}$ = 0.667, $q^2_{L5\%O}$ = 0.261, and $RMSE_{L5\%O}$ = 0.98. The predictions of the neural kernel span also a large range, with the best results obtained in experiment 11: $r_{L5\%O}$ = 0.696, $q^2_{L5\%O}$ = 0.453, and $RMSE_{L5\%O}$ = 0.84. This is the best SVR prediction, and it is very close to the MLR results for the L5%O test. Despite the complexity of the SVR algorithm, its predictive power in the L5%O test is not even equal to the predictions obtained with the classical MLR model. The prediction quality varies widely with the kernel type and parameters, which makes very difficult and time consuming the identification of a best SVR model.

Table 3. Support vector regression statistics for leave–5%–out ($r_{L5\%O}$, $q^2_{L5\%O}$, and $RMSE_{L5\%O}$) and leave–10%–out ($r_{L10\%O}$, $q^2_{L10\%O}$, and $RMSE_{L10\%O}$) cross–validation tests

Exp	Kernel	$r_{L5\%O}$	$q^2_{L5\%O}$	$RMSE_{L5\%O}$	$r_{L10\%O}$	$q^2_{L10\%O}$	$RMSE_{L10\%O}$
1	D	0.667	0.261	0.98	0.672	0.273	0.97
2	P	–0.270	<–100	>10	–0.265	<–100	>10
3	P	–0.290	<–100	>10	–0.291	<–100	>10
4	P	–0.297	<–100	>10	–0.300	<–100	>10
5	P	–0.301	<–100	>10	–0.300	<–100	>10
6	R	0.665	0.368	0.91	0.676	0.370	0.91
7	R	0.591	0.226	1.00	0.636	0.324	0.94
8	R	0.538	0.221	1.01	0.590	0.310	0.95
9	R	0.519	0.236	1.00	0.535	0.258	0.98
10	R	0.483	0.208	1.01	0.481	0.205	1.02
11	N	0.696	0.453	0.84	0.729	0.498	0.81
12	N	0.665	0.416	0.87	0.659	0.411	0.87
13	N	0.651	0.396	0.89	0.653	0.394	0.89
14	N	0.391	0.070	1.10	0.446	0.120	1.07
15	N	0.641	0.389	0.89	0.145	–21.114	5.36
16	N	0.562	0.297	0.96	0.596	0.348	0.92
17	N	0.236	–0.365	1.33	0.187	–0.405	1.35
18	N	0.358	0.106	1.08	0.505	0.243	0.99
19	N	0.634	0.345	0.92	0.649	0.376	0.90
20	A	0.660	0.397	0.89	0.652	0.377	0.90
21	A	0.635	0.324	0.94	0.644	0.331	0.93
22	A	0.689	0.398	0.88	0.690	0.412	0.87
23	A	0.645	0.299	0.95	0.670	0.374	0.90
24	A	0.641	0.293	0.96	0.653	0.339	0.93
25	A	0.305	–1.289	1.73	0.446	–0.921	1.58
26	A	0.238	–1.375	1.76	0.367	–0.873	1.56
27	A	0.304	–0.606	1.44	0.306	–0.465	1.38
28	A	0.401	–0.241	1.27	0.385	–0.195	1.25
29	A	0.455	–0.071	1.18	0.438	–0.060	1.17
30	A	0.125	–2.998	2.28	0.238	–1.934	1.95
31	A	0.181	–1.282	1.72	0.221	–0.983	1.61
32	A	0.406	–0.060	1.17	0.375	–0.094	1.19
33	A	0.447	0.060	1.11	0.356	–0.097	1.19
34	A	0.434	–0.062	1.18	0.394	–0.073	1.18

Table 4. Support vector regression statistics for leave–20%–out ($r_{L20\%O}$, $q^2_{L20\%O}$, and $RMSE_{L20\%O}$) and leave–25%–out ($r_{L25\%O}$, $q^2_{L25\%O}$, and $RMSE_{L25\%O}$) cross–validation tests

Exp	Kernel	$r_{L20\%O}$	$q^2_{L20\%O}$	$RMSE_{L20\%O}$	$r_{L25\%O}$	$q^2_{L25\%O}$	$RMSE_{L25\%O}$
1	D	0.674	0.228	1.00	0.667	0.189	1.03
2	P	–0.277	<–100	>10	0.332	<–100	>10
3	P	–0.288	<–100	>10	–0.297	<–100	>10
4	P	–0.313	<–100	>10	0.289	<–100	>10
5	P	–0.302	<–100	>10	0.298	<–100	>10
6	R	0.633	0.317	0.94	0.680	0.344	0.92
7	R	0.568	0.188	1.03	0.638	0.282	0.97
8	R	0.558	0.261	0.98	0.604	0.352	0.92
9	R	0.529	0.249	0.99	0.516	0.259	0.98
10	R	0.487	0.211	1.01	0.443	0.189	1.03
11	N	0.695	0.450	0.85	0.666	0.431	0.86
12	N	0.670	0.418	0.87	0.691	0.432	0.86
13	N	0.644	0.386	0.89	0.644	0.396	0.89
14	N	0.340	–0.019	1.15	0.548	0.292	0.96
15	N	0.686	0.439	0.85	0.640	0.392	0.89
16	N	0.598	0.355	0.92	0.497	0.210	1.01
17	N	–0.243	–123.006	12.70	0.408	0.107	1.08
18	N	0.429	0.166	1.04	0.335	0.050	1.11
19	N	0.634	0.363	0.91	0.636	0.368	0.91
20	A	0.678	0.413	0.87	0.660	0.277	0.97
21	A	0.678	0.390	0.89	0.641	0.116	1.07
22	A	0.698	0.386	0.89	0.653	0.220	1.01
23	A	0.646	0.195	1.02	0.656	0.340	0.93
24	A	0.632	0.155	1.05	0.675	0.376	0.90
25	A	0.415	–0.543	1.42	0.291	–2.497	2.13
26	A	0.401	–0.509	1.40	0.321	–1.072	1.64
27	A	0.444	–0.068	1.18	0.421	–0.311	1.31
28	A	0.466	0.023	1.13	0.477	–0.032	1.16
29	A	0.484	0.076	1.10	0.505	0.067	1.10
30	A	0.415	–0.619	1.45	0.176	–3.121	2.31
31	A	0.440	–0.149	1.22	0.297	–0.983	1.61
32	A	0.302	–0.418	1.36	0.365	–0.555	1.42
33	A	0.305	–0.368	1.33	0.419	–0.163	1.23
34	A	0.315	–0.303	1.30	0.501	0.148	1.05

The trend identified in the L5%O test is also apparent from the prediction results for the L10%O, L20%O, and L25%O tests (Tables 3 and 4). Consistently, the worst predictions are obtained with the polynomial kernel, while the dot kernel is more robust, with RMSE between 0.97 and 1.03. The anova kernel gives predictions with a large range of variation. Typically, experiments 20–24 have acceptable statistics, while the predictions from experiments 25–34 are bad. The SVR models with radial kernel show a constant decrease of prediction quality from experiment 6 to experiment 10. Similarly with the LOO and L5%O, experiment 11 (representing an SVR with neural kernel) has the best prediction statistics among all 34 SVR models for L10%O and L20%O, while for L25%O the best results are obtained in the experiment 12, also with a neural kernel. Overall, the predictions obtained with the neural kernel fluctuate, and some unexpected bad predictions are obtained (L10%O, experiment 15; L20%O, experiment 17).

It is interesting to compare the best SVR model with the corresponding MLR prediction. For L10%O, SVR has slightly better prediction statistics: MLR, $r_{L10\%O} = 0.711$, $q^2_{L10\%O} = 0.448$, and

$RMSE_{L10\%O}$ = 0.85; SVR, neural kernel, experiment 11, $r_{L10\%O}$ = 0.729, $q^2_{L10\%O}$ = 0.498, and $RMSE_{L10\%O}$ = 0.81. For L20%O, MLR has better prediction statistics: MLR, $r_{L20\%O}$ = 0.733, $q^2_{L20\%O}$ = 0.502, and $RMSE_{L20\%O}$ = 0.81; SVR, neural kernel, experiment 11, $r_{L20\%O}$ = 0.695, $q^2_{L20\%O}$ = 0.450, and $RMSE_{L20\%O}$ = 0.85. The same situation is found for the L25%O test: MLR, $r_{L25\%O}$ = 0.712, $q^2_{L25\%O}$ = 0.470, and $RMSE_{L25\%O}$ = 0.83; SVR, neural kernel, experiment 12, $r_{L25\%O}$ = 0.691, $q^2_{L25\%O}$ = 0.432, and $RMSE_{L25\%O}$ = 0.86.

The five prediction tests show that the SVR QSAR is not able to outperform the simple MLR model for the QSAR example considered in this study. SVR models with the neural kernel have slightly better predictions than MLR for LOO and L10%O, while MLR is better for L5%O, L20%O, and L25%O. Our results indicate that for more difficult prediction tests MLR is more reliable than SVR.

4 CONCLUSIONS

Considering the good performances of the support vector machines in classification, it is expected that support vector regression would be a reliable model for QSAR applications. While support vector classification was tested for a large number of bioinformatics and cheminformatics problems, SVR is relatively new and extensive comparative studies are necessary in order to evaluate this new QSAR model. In this study we compared MLR and SVR QSAR models for the benzodiazepine receptor affinity of 52 2–aryl(heteroaryl)–2,5–dihydropyrazolo[4,3–*c*]quinolin–3–(3*H*)–ones [1,2]. Both models were developed with five structural descriptors, namely the Hammett electronic parameter $\sigma_{R'}$, the molar refractivity MR_{R8}, the Sterimol parameter $L_{R'4'}$, an indicator variable *I* (1/0) for 7–substituted compounds, and the Sterimol parameter B_{5R}.

The SVR prediction power depends on the kernel type and the parameters that control the kernel shape. For the moment there are no clear rules on selecting the most predictive kernel, and as a result we explored a group of 34 SVR experiments obtained with five kernels, namely the dot, polynomial, radial basis function, neural, and anova kernels. The QSAR models were tested with complete cross–validation: leave–one–out, leave–5%–out, leave–10%–out, leave–20%–out, and leave–25%–out.

The results obtained for the set of 52 benzodiazepine receptor ligands show that SVR QSAR models have lower prediction statistics than the MLR model, as measured in *k*–fold cross–validation tests, especially for 5–fold (L20%O) and 4–fold (L25%O) cross–validation. While the calibration SVR models obtained with the polynomial, radial, and anova kernels are better than the MLR QSAR, the cross–validation statistics for these kernels are much lower than the corresponding MLR cross–validation statistics. As expected, the dot kernel gives prediction statistics slightly lower than the MLR results. The neural kernel has low calibration statistics (r_{cal}= 0.690 and

$RMSE_{cal}$ = 0.85 for experiment 11) compared with MLR (r_{cal}= 0.798 and $RMSE_{cal}$ = 0.69) and SVR QSAR with the polynomial, radial, and anova kernels (best calibration statistics obtained with an anova kernel in experiments 29 and 31–34, with r_{cal}= 0.896 and $RMSE_{cal}$ = 0.52). However, the neural kernel (experiment 11 for all cross–validation tests, and experiment 12 for L25%O) has the best prediction statistics in the group of 34 SVR models computed for each cross–validation test.

The extensive cross–validation tests performed in this study do not reveal any advantage of SVR over the classical MLR model. SVR models with the neural kernel have slightly better predictions than MLR for LOO and L10%O, while MLR is better for L5%O, L20%O, and L25%O. The five prediction tests show that SVR is not able to outperform the simple MLR model for the QSAR example considered in this study. The prediction quality of the SVR model varies widely with the kernel type and parameters, which makes very difficult and time consuming the identification of a best QSAR equation. The results obtained in the present study indicate that SVR applications in QSAR must be compared with other models, in order to determine if their use brings any prediction improvement. Despite many over–optimistic expectations, support vector regression can overfit the data, and SVR predictions may be worse than those obtained with linear models.

5 REFERENCES

[1] D. Hadjipavlou–Litina, R. Garg, and C. Hansch, Comparative quantitative structure–activity relationship studies (QSAR) on non–benzodiazepine compounds binding to benzodiazepine receptor (BzR), *Chem. Rev.* **2004**, *104*, 3751–3793.
[2] L. Savini, P. Massarelli, C. Nencini, C. Pellerano, G. Biggio, A. Maciocco, G. Tuligi, A. Carrieri, N. Cinone, and A. Carotti, High affinity central benzodiazepine receptor ligands: Synthesis and structure–activity relationship studies of a new series of pyrazolo[4,3–*c*]quinolin–3–ones, *Bioorg. Med. Chem.* **1998**, *6*, 389–399.
[3] L. Savini, P. Massarelli, C. Nencini, C. Pellerano, G. Biggio, A. Maciocco, G. Tuligi, A. Carrieri, N. Cinone, and A. Carotti, High affinity central benzodiazepine receptor ligands: synthesis and structure–activity relationship studies of a new series of pyrazolo[4,3–*c*]quinolin–3–ones, *Bioorg. Med. Chem.* **2001**, *9*, 431–444.
[4] A. Carotti, C. Altornare, L. Savini, L. Chlasserini, C. Pellerano, M. P. Mascia, E. Maciocco, F. Busonero, M. Mameli, G. Biggio, and E. Sanna, High affinity central benzodiazepine receptor ligands. Part 3: Insights into the pharmacophore and pattern recognition study of intrinsic activities of pyrazolo[4,3–*c*]quinolin–3–ones, *Bioorg. Med. Chem.* **2003**, *11*, 5259–5272.
[5] V. Vapnik, *Estimation of Dependencies Based on Empirical Data*, Nauka, Moscow, 1979.
[6] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, 1995.
[7] V. Vapnik, *Statistical Learning Theory*, Wiley–Interscience, New York, 1998.
[8] C. J. C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition, *Data Mining Knowledge Discov.* **1998**, *2*, 121–167.
[9] B. Schölkopf, C. J. C. Burges, and A. J. Smola, *Advances in Kernel Methods: Support Vector Learning*, MIT Press, Cambridge, MA, 1999.
[10] N. Cristianini and J. Shawe–Taylor, *An Introduction to Support Vector Machines*, Cambridge University Press, Cambridge, 2000.
[11] B. Scholkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, New support vector algorithms, *Neural Computation* **2000**, *12*, 1207–1245.
[12] R. Collobert and S. Bengio, SVMTorch: Support vector machines for large–scale regression problems, *J. Mach. Learn. Res.* **2001**, *1*, 143–160.
[13] O. L. Mangasaian and D. R. Musicant, Lagrangian support vector machines, *J. Mach. Learn. Res.* **2001**, *1*, 161–177.
[14] B. Scholkopf and A. J. Smola, A short introduction to learning with kernels, *Advanced Lectures on Machine Learning, Vol. 2600*, pp. 41–64, 2002.

[15] V. D. Sanchez, Advanced support vector machines and kernel methods, *Neurocomputing* **2003**, *55*, 5–20.
[16] A. J. Smola and B. Scholkopf, On a kernel–based method for pattern recognition, regression, approximation, and operator inversion, *Algorithmica* **1998**, *22*, 211–231.
[17] O. L. Mangasarian and D. R. Musicant, Robust linear and support vector regression, *IEEE Trans. Pattern Analysis Mach. Intell.* **2000**, *22*, 950–955.
[18] O. L. Mangasarian and D. R. Musicant, Large scale kernel regression via linear programming, *Mach. Learn.* **2002**, *46*, 255–269.
[19] J. B. Gao, S. R. Gunn, and C. J. Harris, SVM regression through variational methods and its sequential implementation, *Neurocomputing* **2003**, *55*, 151–167.
[20] J. B. Gao, S. R. Gunn, and C. J. Harris, Mean field method for the support vector machine regression, *Neurocomputing* **2003**, *50*, 391–405.
[21] A. J. Smola and B. Scholkopf, A tutorial on support vector regression, *Statistics Comput.* **2004**, *14*, 199–222.
[22] K. Vlahovicek, U. Kajan, V. Agoston, and S. Pongor, The SBASE domain sequence resource, release 12: prediction of protein domain–architecture using support vector machines, *Nucl. Acids Res.* **2005**, *33*, D223–D225.
[23] L. Lukas, A. Devos, J. A. K. Suykens, L. Vanhamme, F. A. Howe, C. Majos, A. Moreno–Torres, M. Van Der Graaf, A. R. Tate, C. Arus, and S. Van Huffel, Brain tumor classification based on long echo proton MRS signals, *Artif. Intell. Medicine* **2004**, *31*, 73–89.
[24] A. Bertoni, R. Folgieri, and G. Valentini, Bio–molecular cancer prediction with random subspace ensembles of support vector machines, *Neurocomputing* **2005**, *63*, 535–539.
[25] C. S. Leslie, E. Eskin, A. Cohen, J. Weston, and W. S. Noble, Mismatch string kernels for discriminative protein classification, *Bioinformatics* **2004**, *20*, 467–476.
[26] A. Passerini and P. Frasconi, Learning to discriminate between ligand–bound and disulfide–bound cysteines, *Protein Eng. Des. Selection* **2004**, *17*, 367–373.
[27] L. Y. Han, C. Z. Cai, Z. L. Ji, and Y. Z. Chen, Prediction of functional class of novel viral proteins by a statistical learning method irrespective of sequence similarity, *Virology* **2005**, *331*, 136–143.
[28] P. D. Dobson and A. J. Doig, Predicting enzyme class from protein structure without alignments, *J. Mol. Biol.* **2005**, *345*, 187–199.
[29] A. Koike and T. Takagi, Prediction of protein–protein interaction sites using support vector machines, *Protein Eng. Des. Selection* **2004**, *17*, 165–173.
[30] M. S. Scott, D. Y. Thomas, and M. T. Hallett, Predicting subcellular localization via protein motif co–occurrence, *Genome Res.* **2004**, *14*, 1957–1966.
[31] C. S. Yu, C. J. Lin, and J. K. Hwang, Predicting subcellular localization of proteins for Gram–negative bacteria by support vector machines based on n–peptide compositions, *Protein Sci.* **2004**, *13*, 1402–1406.
[32] Y. D. Cai, X. J. Liu, X. Xu, and K. C. Chou, Support vector machines for predicting membrane protein types by incorporating quasi–sequence–order effect, *Internet Electron. J. Mol. Des.* **2002**, *1*, 219–226, http://www.biochempress.com.
[33] Z. Yuan, J. S. Mattick, and R. D. Teasdale, SVMtm: Support vector machines to predict transmembrane segments, *J. Comput. Chem.* **2004**, *25*, 632–636.
[34] O. Ivanciuc, Structure–odor relationships for pyrazines with support vector machines, *Internet Electron. J. Mol. Des.* **2002**, *1*, 269–284, http://www.biochempress.com.
[35] S. Zomer, C. Guillo, R. G. Brereton, and M. Hanna–Brown, Toxicological classification of urine samples using pattern recognition techniques and capillary electrophoresis, *Anal. Bioanal. Chem.* **2004**, *378*, 2008–2020.
[36] G. Steiner, L. Suter, F. Boess, R. Gasser, M. C. de Vera, S. Albertini, and S. Ruepp, Discriminating different classes of toxicants by transcript profiling, *Environm. Health Persp.* **2004**, *112*, 1236–1248.
[37] O. Ivanciuc, Support vector machine classification of the carcinogenic activity of polycyclic aromatic hydrocarbons, *Internet Electron. J. Mol. Des.* **2002**, *1*, 203–218, http://www.biochempress.com.
[38] O. Ivanciuc, Support vector machines for cancer diagnosis from the blood concentration of Zn, Ba, Mg, Ca, Cu, and Se, *Internet Electron. J. Mol. Des.* **2002**, *1*, 418–427, http://www.biochempress.com.
[39] C. Helma, T. Cramer, S. Kramer, and L. De Raedt, Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1402–1411.
[40] C. Merkwirth, H. A. Mauser, T. Schulz–Gasch, O. Roche, M. Stahl, and T. Lengauer, Ensemble methods for classification in cheminformatics, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1971–1978.
[41] O. Ivanciuc, Support vector machines classification of black and green teas based on their metal content, *Internet Electron. J. Mol. Des.* **2003**, *2*, 348–357, http://www.biochempress.com.
[42] A. Kovatcheva, A. Golbraikh, S. Oloff, Y. D. Xiao, W. F. Zheng, P. Wolschann, G. Buchbauer, and A. Tropsha, Combinatorial QSAR of ambergris fragrance compounds, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 582–595.
[43] O. Sadik, W. H. Land, A. K. Wanekaya, M. Uematsu, M. J. Embrechts, L. Wong, D. Leibensperger, and A. Volykin, Detection and classification of organophosphate nerve agent simulants using support vector machines

with multiarray sensors, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 499–507.
[44] O. Ivanciuc, Support vector machine identification of the aquatic toxicity mechanism of organic compounds, *Internet Electron. J. Mol. Des.* **2002**, *1*, 157–172, http://www.biochempress.com.
[45] O. Ivanciuc, Aquatic toxicity prediction for polar and nonpolar narcotic pollutants with support vector machines, *Internet Electron. J. Mol. Des.* **2003**, *2*, 195–208, http://www.biochempress.com.
[46] O. Ivanciuc, Support vector machines prediction of the mechanism of toxic action from hydrophobicity and experimental toxicity against *Pimephales promelas* and *Tetrahymena pyriformis*, *Internet Electron. J. Mol. Des.* **2004**, *3*, 802–821, http://www.biochempress.com.
[47] R. Kumar, A. Kulkarni, V. K. Jayaraman, and B. D. Kulkarni, Structure–activity relationships using locally linear embedding assisted by support vector and lazy learning regressors, *Internet Electron. J. Mol. Des.* **2004**, *3*, 118–133, http://www.biochempress.com.
[48] U. Thissen, M. Pepers, B. Ustun, W. J. Melssen, and L. M. C. Buydens, Comparing support vector machines to PLS for spectral regression applications, *Chemom. Intell. Lab. Syst.* **2004**, *73*, 169–179.
[49] I. S. Han, C. H. Han, and C. B. Chung, Melt index modeling with support vector machines, partial least squares, and artificial neural networks, *J. Appl. Polymer Sci.* **2005**, *95*, 967–974.
[50] S. Rüping, mySVM, University of Dortmund, http://www–ai.cs.uni–dortmund.de/SOFTWARE/MYSVM/.