

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/232318630>

Quantitative Prediction of Antitarget Interaction Profiles for Chemical Compounds

ARTICLE *in* CHEMICAL RESEARCH IN TOXICOLOGY · DECEMBER 2012

Impact Factor: 3.53 · DOI: 10.1021/tx300247r · Source: PubMed

CITATIONS

16

READS

8

4 AUTHORS, INCLUDING:



Alexey Zakharov

National Cancer Institute (USA)

10 PUBLICATIONS 127 CITATIONS

SEE PROFILE



Alexey Lagunin

Pirogov Russian National Research Medical ...

83 PUBLICATIONS 1,181 CITATIONS

SEE PROFILE



Vladimir Poroikov

Institute of Biomedical Chemistry, Moscow

163 PUBLICATIONS 1,983 CITATIONS

SEE PROFILE

Quantitative Prediction of Antitarget Interaction Profiles for Chemical Compounds

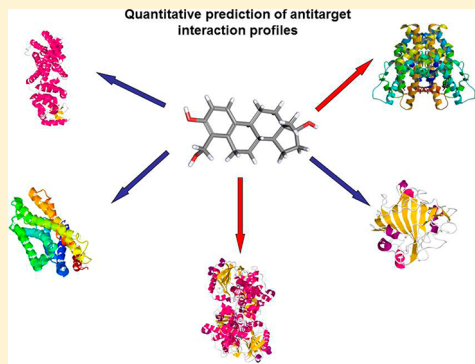
Alexey V. Zakharov,^{*,‡,‡} Alexey A. Lagunin,[‡] Dmitry A. Filimonov,[‡] and Vladimir V. Poroikov[‡]

[†]National Cancer Institute, National Institutes of Health, 376 Boyles Street, Frederick, Maryland 21702, United States

[‡]Institute of Biomedical Chemistry, Russian Academy of Medical Sciences, Pogodinskaya Str. 10, Moscow 119121, Russia

Supporting Information

ABSTRACT: The evaluation of possible interactions between chemical compounds and antitarget proteins is an important task of the research and development process. Here, we describe the development and validation of QSAR models for the prediction of antitarget end-points, created on the basis of multilevel and quantitative neighborhoods of atom descriptors and self-consistent regression. Data on 4000 chemical compounds interacting with 18 antitarget proteins (13 receptors, 2 enzymes, and 3 transporters) were used to model 32 sets of end-points (IC_{50} , K_i , and K_{act}). Each set was randomly divided into training and test sets in a ratio of 80% to 20%, respectively. The test sets were used for external validation of QSAR models created on the basis of the training sets. The coverage of prediction for all test sets exceeded 95%, and for half of the test sets, it was 100%. The accuracy of prediction for 29 of the end-points, based on the external test sets, was typically in the range of $R^2_{test} = 0.6–0.9$; three test sets had lower R^2_{test} values, specifically 0.55–0.6. The proposed approach showed a reasonable accuracy of prediction for 91% of the antitarget end-points and high coverage for all external test sets. On the basis of the created models, we have developed a freely available online service for *in silico* prediction of 32 antitarget end-points: <http://www.pharmaexpert.ru/GUSAR/antitargets.html>.



INTRODUCTION

The process of drug development is time-consuming and cost-intensive. Several years are required for lead identification, optimization, and *in vitro* and *in vivo* testing before the first clinical trials are started. Preapproval costs of a new drug exceed US \$800 million.¹ It is well known that about 90% of drug candidates fail in the first phase of clinical trials.²

Approximately 10% of new chemical entities (NCEs) show serious adverse drug reactions (ADRs) after their introduction into medical practice. More than 17 drugs were withdrawn from the market during the period from 1996 to 2006 because they had shown serious adverse drug reactions.³ For example, Amineptine (launched in 1978) is an indirect dopamine agonist, which selectively inhibits dopamine uptake and induces its release, with additional stimulation of the adrenergic system. However, microcystic, macrocystic acne, and hepatotoxicity were observed as common side effects.⁴ Also, Duract (bromfenac sodium) was launched in 1997 for the treatment of acute pain as a nonsteroidal anti-inflammatory drug (NSAID). The drug was withdrawn after postmarketing reports of severe hepatic failure that led to four deaths and eight liver transplants.⁵ Another drug, Vioxx, was launched in 1999 for the treatment of pain and inflammation as an NSAID. In 2004, a long-term study of Vioxx in patients with increased risk of colon polyps was halted because of increased cardiovascular risk. Thus, Vioxx was withdrawn from the market in 2004.⁶

Interactions with some enzymes, receptors, and channels have been identified as molecular mechanisms for certain side effects observed in the development of candidates or marketed drugs and are named antitargets.³ Several antitargets were previously considered as drug targets. Currently, it is thought that the benefit of action on these targets exceeds their side effects. For example, the main target of the well-known antidepressant and serotonin uptake inhibitor Prozac (fluoxetine) is the sodium-dependent serotonin transporter that is considered as a withdrawn target in DrugBank. Most common ADRs (e.g., hepatic toxicity, hematologic toxicity, and cardiovascular toxicity) are caused by drug action on antitargets (off-targets). α -1A adrenergic antagonists may cause orthostatic hypotension, dizziness, and fainting spells. D2 dopaminergic antagonists may cause extrapyramidal syndrome; M1 muscarinic antagonists may cause attention/memory deficits.⁷ Antitarget-mediated side effects may risk the further development of promising clinical candidates.

Therefore, to avoid potential interactions of drugs with antitargets, specific studies for their detection should ideally be conducted before the pharmaceutical is launched. For this purpose, several different computational approaches have been proposed. Three dimensional (3D) pharmacophore models rationalizing the affinity of several different chemical series have

Received: June 4, 2012

been described for α -1A, 5-HT 2A, and D2 receptors.⁸ The ligand–protein inverse docking approach was successfully used to predict the potential toxicity and side effects related to protein targets of small molecules.⁹ Different 2D and 3D QSAR techniques have also been applied for the prediction of antitarget activities.^{10–14} However, until now there was no computational approach that allowed for the prediction of the profiles of antitarget effects for chemical compounds with reasonable accuracy ($Q^2_{\text{model}} > 0.6$ and $R^2_{\text{test}} > 0.6$) and speed (~ 1 compound per second).

Earlier, we have shown that GUSAR software based on multilevel and quantitative neighborhoods of atoms (MNA, QNA) descriptors^{15,16} and the self-consistent regression (SCR) algorithm^{17,18} may successfully be applied for multiple QSAR tasks.¹⁵ In this work, we have applied the GUSAR program to model the interactions between drug-like organic compounds and antitargets including three types of proteins: receptors, enzymes, and transporters. A freely available online service for the quantitative prediction of antitarget interaction profiles was developed using the models derived.

MATERIALS AND METHODS

Data Sets. Data on the chemical structures and quantitative end-point values (50% inhibitory concentration, IC_{50} ; inhibition constant, K_i ; and activation constant, K_{act}) for approximately 4000 chemical compounds interacting with 18 antitarget proteins were collected from different literature sources. An example of chemical structures with end-point values is presented in Table 1. The InChI keys for all chemicals and their activities used in this work are available in the Supporting Information Table S1.

On the basis of the “Withdrawn Drug Targets” and “Withdrawn Enzymes” represented in DrugBank (<http://www.drugbank.ca/downloads>), the list of antitarget proteins includes 13 receptors, 2 enzymes, and 3 transporters. A brief description of the data sets is given below. The ranges of end-point values for the sets of compounds interacting with each receptor are presented in Table 2.

Table 2 shows that the range of the modeled values for compounds from the receptor data sets exceeds 5 logarithmic units. This is a good prerequisite for the creation of accurate, robust, and predictive QSAR models. The minimal values for all cases are less than 5 $-\log_{10}(\text{mol/L})$, which means that the created models can be used to assess both active and inactive molecules.

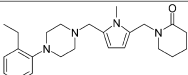
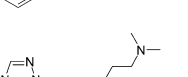
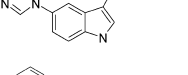
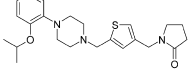
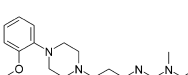
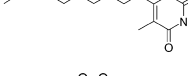
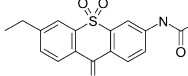
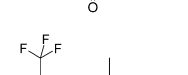
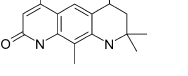
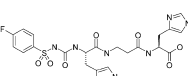
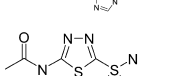
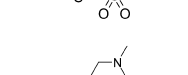
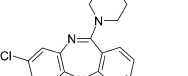
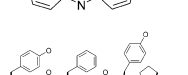
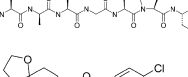
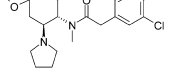
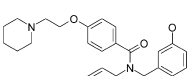
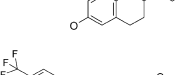
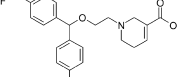
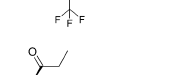
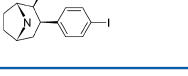
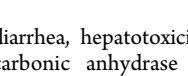
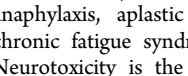
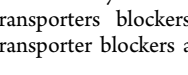


Thirteen receptors belong to the hydroxytryptamine, adrenergic, androgen, dopamine, opioid, and estrogen families. The main adverse effects of antagonists to 5-hydroxytryptamine receptors are sickness, emesis, diarrhea, sleeplessness, and anxiety.¹⁹ Adrenergic receptor antagonists may cause orthostatic hypotension, reflex tachycardia, insomnia, nasal congestion, tachycardia, and palpitation.²⁰ Antagonism to androgen receptors may lead to virilization, gynecomastia, hepatic pelioza, and hepatoma.²⁰ The main undesirable effects of dopamine receptor antagonists are palpitations, ectopic rhythm, tachycardia, retrosternal pain, hypertension, vasoconstriction, shortness of breath, and headache.⁷ Antagonism to opioid receptors may cause sickness, emesis, respiratory depression, and sedation.²¹ Adverse effects related to the action of estrogen receptor antagonists are depression, headache, obesity, sickness, hot flashes, and puffiness.⁷

The ranges of end-point values for the sets of compounds interacting with enzymes (amine oxidase and carbonic anhydrase) and transporters (GABA, dopamine and serotonin transporters) are represented in Tables 3 and 4, respectively.

Tables 3 and 4 show that the ranges of the modeled values for compounds from the enzyme and transporter data sets exceed 4 logarithmic units. The minimal values for all cases are less than 5 $-\log_{10}(\text{mol/L})$; thus, the created models can be used for the evaluation of both active and inactive molecules.

The main adverse effects of amine oxidase (MAO A) inhibitors are blood pressure lability, bradycardia, chorea, convulsions, delirium,

Table 1. Data Set Structural Information

Structure	Activity	End-point	$-\log_{10}(\text{End-point}), \text{mol/L}$
	5-hydroxytryptamine 1B receptor antagonist	K_i	7
	5-hydroxytryptamine 2A receptor antagonist	K_i	6.45
	alpha 1a adrenergic receptor antagonist	K_i	7.95
	5-hydroxytryptamine 2C receptor antagonist	IC_{50}	6.3
	5-hydroxytryptamine 2A receptor antagonist	IC_{50}	5
	alpha 1a adrenergic receptor antagonist	K_i	9.7
	alpha 1b adrenergic receptor antagonist	K_i	6.59
	alpha-2A adrenergic receptor antagonist	K_i	6.36
	alpha 1b adrenergic receptor antagonist	K_i	6.66
	alpha 1a adrenergic receptor antagonist	K_i	7.55
	amine oxidase [flavin-containing] A inhibitor	IC_{50}	8
	androgen receptor antagonist	IC_{50}	7.66
	carbonic anhydrase I activator	K_{act}	7.52
	carbonic anhydrase II activator	K_{act}	6.3
	carbonic anhydrase I inhibitor	K_i	6.05
	carbonic anhydrase II inhibitor	K_i	7.92
	d(1A) dopamine receptor antagonist	K_i	6.38
	d3 dopamine receptor antagonist	K_i	6.01
	mu-type opioid receptor antagonist	IC_{50}	8.43
	delta-type opioid receptor antagonist	K_i	6.72
	mu-type opioid receptor antagonist	K_i	6.15
	kappa-type opioid receptor antagonist	K_i	8.82
	delta-type opioid receptor antagonist	K_i	5.79
	estrogen receptor antagonist	IC_{50}	6.23
	estrogen receptor antagonist	K_i	7.43
	sodium- and chloride-dependent GABA transporter 1 antagonist	IC_{50}	6.59
	sodium-dependent dopamine transporter antagonist	IC_{50}	8.66
	sodium-dependent serotonin transporter antagonist	K_i	8.61

diarrhea, hepatotoxicity, and drowsiness. Interaction of ligands with carbonic anhydrase may be the cause of alopecia (hair loss), anaphylaxis, aplastic anemia, anxiety, bone marrow suppression, chronic fatigue syndrome, depression, and renal tubular acidosis.²⁰ Neurotoxicity is the main adverse effect of GABA and dopamine transporters blockers.²² Adverse effects caused by the serotonin transporter blockers are acute respiratory distress syndrome (ARDS),

Table 2. Receptor Data Sets

activity name	end-point	min ^a	max ^b
5-hydroxytryptamine 1B receptor antagonist	IC ₅₀	2.8	9.3
5-hydroxytryptamine 1B receptor antagonist	K _i	3.8	9.7
5-hydroxytryptamine 2A receptor antagonist	IC ₅₀	3.1	10.3
5-hydroxytryptamine 2A receptor antagonist	K _i	4.3	10.4
5-hydroxytryptamine 2C receptor antagonist	IC ₅₀	2.9	9.0
5-hydroxytryptamine 2C receptor antagonist	K _i	3.4	10.7
α-1A adrenergic receptor antagonist	IC ₅₀	4.0	9.9
α-1A adrenergic receptor antagonist	K _i	3.6	11.0
α-1B adrenergic receptor antagonist	K _i	4.3	10.0
α-2A adrenergic receptor antagonist	IC ₅₀	3.2	9.3
α-2A adrenergic receptor antagonist	K _i	2.4	9.9
androgen receptor antagonist	IC ₅₀	4.2	10
d(1A) dopamine receptor antagonist	IC ₅₀	3.9	9.1
d(1A) dopamine receptor antagonist	K _i	4.1	10.1
d3 dopamine receptor antagonist	K _i	4.2	10.7
delta-type opioid receptor antagonist	K _i	4.3	12.0
estrogen receptor antagonist	IC ₅₀	1.7	11.0
estrogen receptor antagonist	K _i	3.0	12.7
κ-type opioid receptor antagonist	K _i	3.1	11.4
μ-type opioid receptor antagonist	IC ₅₀	3.6	10.5
μ-type opioid receptor antagonist	K _i	3.4	12.3

^aMinimal value of $-\log_{10}(\text{end-point})$ values in the set (mol/L).

^bMaximal value of $-\log_{10}(\text{end-point})$ values in the set (mol/L).

Table 3. Enzyme Data Sets

activity name	end-point	min ^a	max ^b
amine oxidase [flavin-containing] A inhibitor	IC ₅₀	3.0	9.0
amine oxidase [flavin-containing] A inhibitor	K _i	3.5	9.5
carbonic anhydrase II activator	K _{act}	2.4	10.0
carbonic anhydrase I activator	K _{act}	1.5	10.9
carbonic anhydrase I inhibitor	K _i	0.4	9.4
carbonic anhydrase II inhibitor	K _i	4.4	9.7

^aMinimal value of $-\log_{10}(\text{end-point})$ values in the set (mol/L).

^bMaximal value of $-\log_{10}(\text{end-point})$ values in the set (mol/L).

Table 4. Transporter Data Sets

activity name	end-point	min ^a	max ^b
sodium- and chloride-dependent GABA transporter 1 antagonist	IC ₅₀	3.1	7.3
sodium-dependent dopamine transporter antagonist	IC ₅₀	3.0	10.5
sodium-dependent dopamine transporter antagonist	K _i	3.3	9.5
sodium-dependent serotonin transporter antagonist	IC ₅₀	2.4	9.8
sodium-dependent serotonin transporter antagonist	K _i	3.4	11.1

^aMinimal value of $-\log_{10}(\text{end-point})$ values in the set (mol/L).

^bMaximal value of $-\log_{10}(\text{end-point})$ values in the set (mol/L).

agitation, akathisia, constipation, diarrhea, drowsiness, emesis, glaucoma, headaches, hemorrhaging, hypomania, ischemic colitis, parkinsonism, myoclonus, QT interval prolongation, and tremor.⁷

The structural overlap was analyzed between the 32 data set end-points. The results are presented in Supporting Information, Table S2. It was found that all end-points have some structural overlap except for the sodium- and chloride-dependent GABA transporter 1 antagonists. The major cases of overlap were found between receptor end-points, belonging to the hydroxytryptamine, adrenergic, and dopamine families. More than 1000 compounds in common were found among the drug-like molecules acting on the different types of opioid receptors. In contrast to the receptor families, a small number of overlapping structures was found for the enzymes.

Each set was randomly divided into training and test sets according to the ratio 80%:20%, respectively. The training set was used to create the QSAR models, and the test set was used to assess the external predictive accuracy. QSAR models were developed using MNA and QNA descriptors^{15,16} and the self-consistent regression (SCR) algorithm.^{17,18}

QSAR Modeling on the Basis of QNA Descriptors. QSAR modeling on the basis of QNA descriptors has previously been implemented in the software program GUSAR.¹⁵ Reasonable results obtained by GUSAR modeling for different biological end-points¹⁵ prompted this study to investigate the utility of the method for modeling the interactions of chemical structures with sets of antitargets. A more detailed explanation of our approach is presented in Supporting Information. It is briefly described below.

The calculation of QNA descriptors is based on the connectivity matrix (C) and also on the standard values of ionization potential (IP) and electron affinity (EA) of atoms in a molecule.¹⁵ The main important feature of QNA descriptors is that they represent a molecule as a set of the P and Q values or, in other words, as a “constellation” in a two-dimensional QNA descriptors’ space. The P and Q values can be considered as corresponding partial atomic hardness and electro-negativity.

For any given atom *i*, the QNA descriptors are calculated as follows:

$$P_i = B_i \sum_k \left(\exp\left(-\frac{1}{2}C\right) \right)_{ik} B_k$$

$$Q_i = B_i \sum_k \left(\exp\left(-\frac{1}{2}C\right) \right)_{ik} B_k A_k$$

with $A_k = 1/2(IP_k + EA_k)$ and $B_k = (IP_k - EA_k)^{-1/2}$.

The estimation of a target property for a chemical compound is calculated as a mean value of a function of the P and Q values for the atoms in a molecule in the QNA descriptors’ space. We proposed the use of two-dimensional Chebyshev polynomials for approximating the function of P and Q values. Therefore, the independent regression variables are calculated as average values of particular two-dimensional Chebyshev polynomials of P and Q values for the atoms in a molecule.

QNA descriptors and their polynomial transformations do not provide information on the shape and volume of a molecule. Since this type of information can be important for determining structure–activity relationships, these parameters were calculated separately and added to the variables already obtained from the Chebyshev polynomials. The topological length of the molecule is the maximal distance, calculated as the number of bonds between any two atoms (including hydrogen). The volume of a molecule is estimated as the sum of each atom’s volume.

The number of initial variables for QSAR modeling depends on the number of compounds in the training set and corresponds to the number of Chebyshev polynomials plus the number of the first, second, and third power of the values of topological length and volume of a molecule. If the number of compounds in the training set varies from 100 to 2000, then the number of initial variables equals one-half of the number of compounds in the training set.

The GUSAR algorithm uses three randomly selected parameters to generate different QSAR models based on QNA descriptors: (a) calculation of the QNA descriptors for either all atoms or for only the atoms in a molecule with two or more immediate neighbors; (b) adjustment of the connectivity matrix coefficient; and (c) adjustment of the parameters of the Chebyshev polynomials. The detailed algorithm is described in Supporting Information. The final QSAR model is the consensus of several different QNA-based models built in this way.

QSAR Modeling on the Basis of Biological Activity Profile Prediction Using MNA Descriptors. GUSAR enables QSAR models to be derived based on predicted biological activity profiles of compounds. Each compound is represented as a list of MNA descriptors, which are used as input parameters^{16–18} for predicting biological activity profiles in the PASS (prediction of activity spectra

Table 5. Characteristics of QSAR Models for the Antitarget Sets

activity name	end-point	number of compounds training set/test set	number of models	R ² training set	Q ² training set	average R ² of test set during L10%Out	R ² test set	coverage,%
5-hydroxytryptamine 1B receptor antagonist	IC ₅₀	297/74	8	0.83	0.79	0.63 ± 0.02	0.67	100.0
5-hydroxytryptamine 1B receptor antagonist	K _i	266/66	7	0.73	0.66	0.55 ± 0.06	0.72	100.0
5-hydroxytryptamine 2A receptor antagonist	IC ₅₀	555/143	13	0.83	0.78	0.74 ± 0.02	0.71	98.6
5-hydroxytryptamine 2A receptor antagonist	K _i	1010/252	13	0.72	0.65	0.58 ± 0.05	0.59	99.6
5-hydroxytryptamine 2C receptor antagonist	IC ₅₀	128/32	18	0.77	0.73	0.66 ± 0.03	0.58	100.0
5-hydroxytryptamine 2C receptor antagonist	K _i	487/121	14	0.74	0.66	0.53 ± 0.02	0.62	99.2
α-1a adrenergic receptor antagonist	IC ₅₀	438/111	16	0.79	0.73	0.58 ± 0.02	0.72	98.2
α-1a adrenergic receptor antagonist	K _i	1366/344	5	0.83	0.79	0.75 ± 0.02	0.80	97.0
α-1b adrenergic receptor antagonist	K _i	410/102	17	0.73	0.66	0.54 ± 0.03	0.63	100.0
α-2A adrenergic receptor antagonist	IC ₅₀	109/27	16	0.88	0.84	0.77 ± 0.04	0.75	100.0
α-2A adrenergic receptor antagonist	K _i	525/131	17	0.84	0.79	0.68 ± 0.03	0.77	99.2
amine oxidase [flavin-containing] A inhibitor	IC ₅₀	286/71	9	0.80	0.75	0.58 ± 0.02	0.72	100.0
amine oxidase [flavin-containing] A inhibitor	K _i	60/15	5	0.73	0.62	0.57 ± 0.06	0.64	100.0
androgen receptor antagonist	IC ₅₀	116/29	8	0.79	0.73	0.68 ± 0.06	0.67	100.0
carbonic anhydrase II activator	K _{act}	104/26	20	0.92	0.90	0.76 ± 0.04	0.91	100.0
carbonic anhydrase I activator	K _{act}	108/27	12	0.98	0.97	0.92 ± 0.01	0.93	100.0
carbonic anhydrase I inhibitor	K _i	935/234	11	0.91	0.86	0.73 ± 0.02	0.86	98.3
carbonic anhydrase II inhibitor	K _i	866/217	7	0.87	0.79	0.74 ± 0.04	0.76	98.6
d(1A) dopamine receptor antagonist	IC ₅₀	126/31	11	0.76	0.72	0.68 ± 0.03	0.80	100.0
d(1A) dopamine receptor antagonist	K _i	291/73	10	0.72	0.66	0.53 ± 0.02	0.57	100.0
d3 dopamine receptor antagonist	K _i	822/206	9	0.73	0.66	0.59 ± 0.02	0.62	98.0
delta-type opioid receptor antagonist	K _i	1044/261	16	0.75	0.70	0.60 ± 0.04	0.65	98.5
estrogen receptor antagonist	IC ₅₀	402/100	4	0.66	0.61	0.66 ± 0.05	0.70	97.0
estrogen receptor antagonist	K _i	255/68	13	0.76	0.71	0.66 ± 0.04	0.70	100.0
κ-type opioid receptor antagonist	K _i	884/221	7	0.74	0.67	0.59 ± 0.04	0.65	100.0
μ-type opioid receptor antagonist	IC ₅₀	545/136	7	0.67	0.61	0.60 ± 0.03	0.70	97.8
μ-type opioid receptor antagonist	K _i	1354/338	4	0.69	0.62	0.60 ± 0.02	0.60	96.7
sodium- and chloride-dependent GABA transporter 1 antagonist	IC ₅₀	75/19	10	0.9	0.86	0.80 ± 0.03	0.89	100.0
sodium-dependent dopamine transporter antagonist	IC ₅₀	920/230	5	0.7	0.65	0.65 ± 0.04	0.67	98.3
sodium-dependent dopamine transporter antagonist	K _i	655/164	7	0.77	0.69	0.59 ± 0.04	0.64	100.0
sodium-dependent serotonin transporter antagonist	IC ₅₀	796/199	7	0.8	0.75	0.67 ± 0.02	0.69	97.5
sodium-dependent serotonin transporter antagonist	K _i	823/206	2	0.72	0.65	0.62 ± 0.01	0.61	95.6

for substances) software program. The PASS algorithm is based on a Bayesian approach and is used to calculate this profile. A detailed description of the PASS algorithm is presented in Supporting Information.

GUSAR incorporates PASS version 10.1, which predicts 4130 types of biological activity with a mean prediction accuracy of about 95%. The list of predictable biological activities includes 501 pharmacotherapeutic effects (e.g., antihypertensive, hepatoprotectant, nootropic, etc.), 3295 mechanisms of action (e.g., 5 hydroxytryptamine antagonist, acetylcholine M1 receptor agonist, cyclooxygenase inhibitor, etc.), 57 adverse and toxic effects (e.g., carcinogenic, mutagenic, hematotoxic, etc.), 199 metabolic terms (e.g., CYP1A inducer, CYP1A1 inhibitor, CYP3A4 substrate, etc.), 49 transporter proteins (e.g., P-glycoprotein 3 inhibitor, nucleoside transporters inhibitors), and 29 activities related to gene expression (e.g., TH expression enhancer, TNF expression inhibitor, VEGF expression inhibitor). The results of a PASS prediction are given as a list of

biological activities, for which the difference between the probability that a compound is active (Pa) and that it is inactive (Pi) is calculated.

To obtain different QSAR models, the Pa–Pi values for the activities, randomly selected from the total list of predicted biological activities, were input as independent variables for the regression analysis. Similar to the QSAR analysis with QNA descriptors, topological length and volume of molecules were added as variables to the biological activity profile; the number of initial variables for creating regression models was also selected depending on the number of compounds in the training set.

Self-Consistent Regression. GUSAR uses self-consistent regression (SCR) for building (Q)SAR models. SCR is based on the regularized least-squares method described in refs 15 and 17. Unlike stepwise regression and other methods of combinatorial search, the initial SCR model includes all regressors. The basic purpose of the SCR method is to remove the variables, which poorly describe the appropriate target value.^{15,17} The final number of variables in the QSAR equation, selected after the self-consistent regression procedure,

is significantly less compared to the initial number of variables. Nevertheless, the final model contains a set of variables that correctly represents the existing relationship.

Nearest Neighbor Correction. It is well known that the use of both global and local models for noncongeneric sets improves the quality of QSAR models.²³ We used the experimental data on the three nearest neighbors (NN) to each input compound to correct the prediction values obtained from the regression model. The correction value is estimated by taking an average of three values from the training set that are the most similar to the compound under prediction. The similarity of each pair of compounds is estimated as Pearson's coefficient calculated in the space of the independent variables obtained after SCR. The mean experimental value obtained for the three nearest neighbor compounds from the training set is averaged with the predicted value of the test compound.

Applicability Domain. The average similarity to the three nearest neighbor compounds in the training set was also used for the assessment of applicability domain (AD) of the model. If the average similarity exceeds the threshold, then the chemical compound under prediction is considered to be in the AD of the model and vice versa. The higher the value selected for the threshold, the closer in similarity the compounds must be to fall in the AD of the model. In this study, a threshold for the AD equal to 0.7 was used.

Consensus Modeling. The final predicted value for each end-point is estimated by including a weighted average of the predicted values from the set of QSAR models (for predictions that are within their respective applicability domains). The value obtained from each model is weighted by the similarity value calculated for the estimation of its applicability domain. This algorithm combines the results of QSAR modeling on the basis of QNA descriptors and on the basis of PASS-predicted biological activity profiles.

Interpretation of Results. Typically, the affinity of pharmaceutical agent to the drug target should exceed the affinity to off-targets for at least one to 2 orders of magnitude. The medium affinity of current small molecule drugs to drug targets is about 16 nM, ranging from 16 mM to 1.6 pM.²⁴ Therefore, GUSAR prediction of interaction with antitarget(s) should be carefully considered in each individual case taking into account the predicted/measured affinity of the analyzed compound to the drug target. Particular attention should be paid to the compounds, for which predicted affinity of interaction with three or more antitargets exceeded 1 μ M.

RESULTS AND DISCUSSION

QSAR Modeling and Validation of Antitarget End-Points. The initial data for each antitarget end-point was randomly divided into the training and external test sets in a ratio of 80% and 20%, respectively. External test sets were used for the assessment of predictivity in the obtained QSAR models. The number of compounds in the training and external

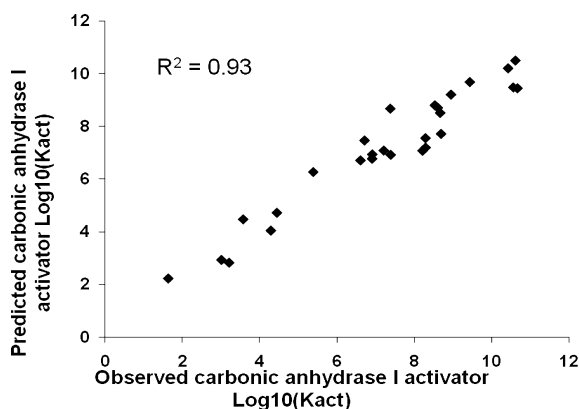


Figure 1. Carbonic anhydrase I activator test set, observed versus predicted $\log_{10}(K_{act})$ values.

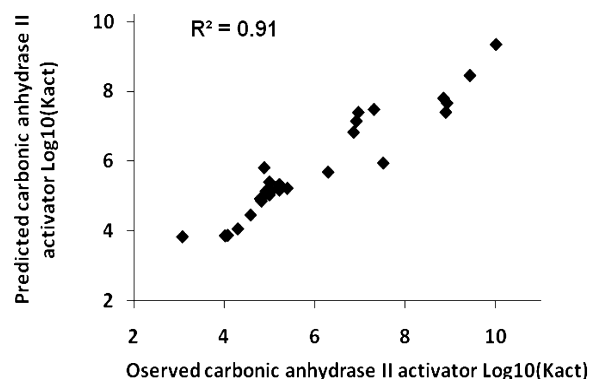


Figure 2. Carbonic anhydrase II activator test set, observed versus predicted $\log_{10}(K_{act})$ values.

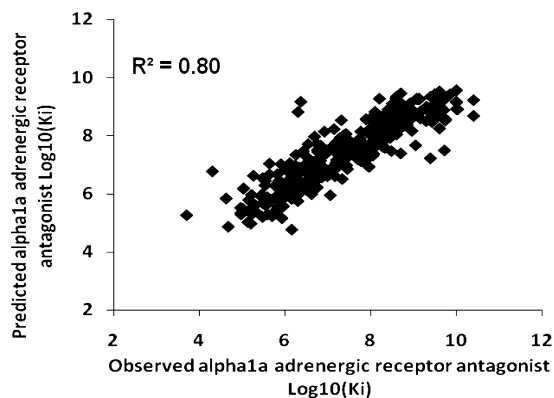


Figure 3. α -1a adrenergic receptor antagonist test set, observed versus predicted $\log_{10}(K_i)$ values.

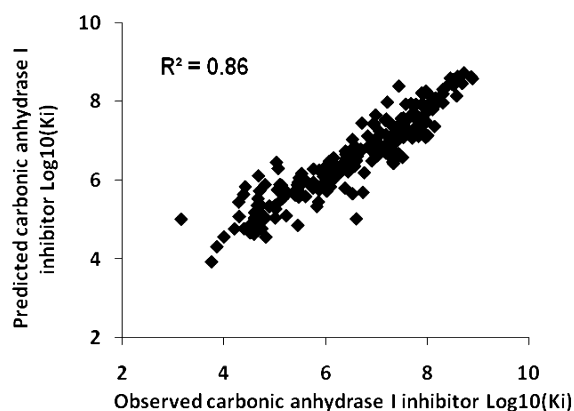
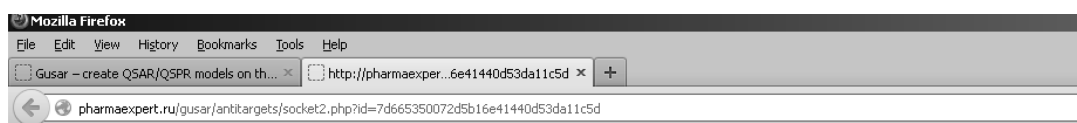


Figure 4. Carbonic anhydrase I inhibitor test set, observed versus predicted $\log_{10}(K_i)$ values.

test sets is shown in Table 5. The four antitarget end-points (K_i) include more than 1000 chemical compounds in the training set and more than 250 compounds in the external test set: 5-hydroxytryptamine 2A receptor antagonists, α -1a adrenergic receptor antagonists, δ -type opioid receptor antagonists, and μ -type opioid receptor antagonists. Thirty of the 32 end-points ($\sim 94\%$) include more than 100 compounds in the training set. Therefore, only two training sets contain a small number of compounds: "sodium- and chloride-dependent GABA transporter 1 antagonist" with IC_{50} values and "amine oxidase [flavin-containing] A inhibitor" with K_i values (75 and 60 compounds, respectively).



Quantitative prediction of antitarget interaction profiles for chemical compounds

Activity	Prediction Value, -Log10(Value), Mole	Applicability Domain
5-hydroxytryptamine 1B receptor antagonist IC50	6,515	Out of AD
5-hydroxytryptamine 1B receptor antagonist Ki	6,143	In AD
5-hydroxytryptamine 2A receptor antagonist IC50	7,076	In AD
5-hydroxytryptamine 2A receptor antagonist Ki	7,394	In AD
5-hydroxytryptamine 2C receptor antagonist IC50	6,134	In AD
5-hydroxytryptamine 2C receptor antagonist Ki	8,060	In AD
alpha1a adrenergic receptor antagonist IC50	7,525	In AD
alpha1a adrenergic receptor antagonist Ki	7,671	Out of AD
alpha1b adrenergic receptor antagonist Ki	6,344	In AD
Alpha-2A adrenergic receptor antagonist IC50	6,619	In AD
Alpha-2A adrenergic receptor antagonist Ki	5,971	In AD
amine oxidase [flavin-containing] A inhibitor IC50	5,138	In AD
amine oxidase [flavin-containing] A inhibitor Ki	5,471	In AD
androgen receptor antagonist IC50	6,365	In AD
carbonic anhydrase I activator Kact	8,097	Out of AD
carbonic anhydrase I inhibitor Ki	7,436	In AD
Carbonic anhydrase 2 activator Kact	5,594	In AD
carbonic anhydrase II inhibitor Ki	7,851	In AD
D(1A) dopamine receptor antagonist IC50	6,390	In AD
D(1A) dopamine receptor antagonist Ki	5,953	In AD
D3 dopamine receptor antagonist Ki	6,881	Out of AD
delta-type opioid receptor antagonist Ki	6,774	Out of AD
estrogen receptor antagonist IC50	5,439	Out of AD
estrogen receptor antagonist Ki	3,955	Out of AD
kappa-type opioid receptor antagonist Ki	7,122	Out of AD
mu-type opioid receptor antagonist IC50	5,522	Out of AD
mu-type opioid receptor antagonist Ki	6,436	Out of AD
sodium- and chloride-dependent GABA transporter 1 antagonist IC50	5,645	In AD
sodium-dependent dopamine transporter antagonist IC50	6,305	Out of AD
sodium-dependent dopamine transporter antagonist Ki	6,003	In AD
sodium-dependent serotonin transporter antagonist IC50	5,423	Out of AD
sodium-dependent serotonin transporter antagonist Ki	8,176	Out of AD

Significant level of interaction (6 -Log10(Mole) and higher) is predicted for the follow number of antitarget(s): 11

in AD - compound falls in the applicability domain of the model
out of AD - compound is out of the applicability domain of the model

<http://pharmaexpert.ru/GUSAR/antitargets.html>

Figure 5. Results of the antitarget interaction profile prediction for Temafloxacin using our online service.

For each training set, 40 models based on MNA descriptors and 40 models based on QNA descriptors were created. A leave-10%-out cross-validation procedure was performed 20 times for each model, and the results were used to select the most predictive models. The average R^2 value and the standard deviation of R^2 values calculated for the test sets during the L-10%-out procedure are presented in Table 5. From the full set of 80 models, we selected only those models that satisfied the following conditions: a value of Q^2 exceeding 0.6 and an R^2 value from the leave-10%-out cross validation procedure exceeding 0.5. Thus, over 10 models were selected for 16 of the end-points (50%) and only the sodium-dependent serotonin transporter antagonist activity was represented by two models. The selected models were used for consensus predictions on the external test set obtained for each antitarget

end-point, taking into account the applicability domain of these models.

Table 5 shows the results of the consensus predictions on the external test sets for each antitarget end-point. Sixteen test sets (50%) were predicted with 100% coverage, and the remaining test sets were predicted with coverage exceeding 95%. QSAR models obtained for 29 of the antitarget end-points (91%) showed good statistical criteria for the external test sets. The accuracy of prediction for these end-points in external tests was in the range of 0.6–0.9. Thus, two activities (6%) were predicted with an accuracy higher than 0.9, the K_{act} of carbonic anhydrase I activator and carbonic anhydrase II activator. Plots of observed versus predicted values for these activities are shown in Figures 1 and 2. Six activities (19%) were predicted with an accuracy higher than 0.8 and 16 (50%) with an

Table 6. Prediction Results for Withdrawn and Marketed Drugs

drug name	state	number of predicted antitargets
amineptine	withdrawn	13
Duract	withdrawn	8
Vioxx	withdrawn	7
astemizole	withdrawn	17
cerivastatin	withdrawn	8
chlormezanone	withdrawn	10
fenfluramine	withdrawn	11
flosequinan	withdrawn	11
glafenine	withdrawn	14
grepafloxacin	withdrawn	12
mibefradil	withdrawn	16
rofecoxib	withdrawn	7
troglitazone	withdrawn	14
ximelagatran	withdrawn	14
aspirin	marketed	2
ibuprofen	marketed	2
Valtrex	marketed	3
microzide	marketed	3
neurontin	marketed	3
enoxaparin	marketed	2
Lyrica	marketed	2

accuracy higher than 0.7. Plots of observed versus predicted values for K_i of α -1a adrenergic receptor antagonist and carbonic anhydrase II inhibitor are shown in Figures 3 and 4. Only three antitarget end-points, K_i of 5-hydroxytryptamine 2A receptor antagonist, IC_{50} of 5-hydroxytryptamine 2C receptor antagonist, and K_i of d(1A) dopamine receptor antagonist, showed not so high but reasonable values for accuracy exceeding 0.5.

Some points in Figures 1–4 can be considered to be outliers; nevertheless, the number of outliers (~10–15) is negligible in comparison to the total number of points. In general, Figures 1–4 show that there are no visible artifacts in the predicted values.

These results show that MNA and QNA descriptors with self-consistent regression can be successfully used for the development of accurate and predictive QSAR models of antitarget effects.

Online Service for Quantitative Prediction of Antitarget Interaction Profiles for Chemical Compounds. On the basis of our created QSAR models, we have developed a freely available online service for the simultaneous prediction of 32 antitarget end-points available at: <http://www.pharmaexpert.ru/GUSAR/Antitargets/>. It includes an online chemical editor (ChemAxon Marvin Sketch)²⁵ for drawing the studied structure. This service provides a reasonable computational speed (about 2 compounds per second for the simultaneous prediction of 32 antitarget end-points). The assessment of the prediction results for the 32 antitarget end points may be done in the same way as an assessment for *in vitro* experimental assays. We consider that 1 μ M (6 in the units used on the Web site: $-\log_{10}(\text{value})$, mol/L) is a minimal cutoff value for any end-point. This means that a compound predicted to have an IC_{50} or K_i of less than 1 μ M is active against that undesirable target (antitarget). The significance of these predicted values depends on the therapeutic dose of the drug and the magnitude of the interaction between the drug and its target. However, the probability of adverse drug

reactions may increase with the number of interactions with antitargets. For statistical reasons, any compound interacting with three or more antitargets may be considered potentially risky for further drug development and should be filtered out at the early stage of studies.

An example of prediction results for the antibiotic agent Temafloxacin (Omniflox) obtained with this service is represented in Figure 5. Temafloxacin was marketed in the United States from February to June 1992. During the three months of its use, the FDA received 50 reports of multiple side effects including three cases of death.⁶ It was shown that Temafloxacin may produce hemolytic anemia and other blood cell abnormalities, kidney dysfunction, and liver dysfunction.²⁶ Hemolytic anemia and blood cell abnormalities may be related to the inhibition of carbonic anhydrase. The cause of kidney dysfunction may be the interaction of Temafloxacin with the dopamine receptor and inhibition of the carbonic anhydrase. Liver dysfunction may be induced by action on the androgen receptor.

Figure 5 shows that for these modeled antitarget end-points (carbonic anhydrase, dopamine, and androgen receptors), Temafloxacin falls in the applicability domain and is predicted to be active (K_i and IC_{50} exceed 6 units in $-\log_{10}(\text{mol/L})$). Thus, the GUSAR predictions correspond to the multiple side effects that were found during Temafloxacin usage in medical practice.

The Web service also provides the total number of targets for which the input compound has been predicted to be active. This can be useful for the selection and prioritization of compounds during the drug discovery process. A particular compound can be considered as a potential source of adverse drug reactions if interactions with three or more antitargets are predicted and exceed the cutoff value (1 μ M). Compounds for which antitargets are not predicted can be selected for further development as potential drugs. In addition, the service can help medical chemists determine on which targets (molecular mechanism of toxicity) a particular compound should be tested experimentally to avoid ADR.

We have applied the Web service to 14 known drugs that have been withdrawn from the market to estimate the number of antitargets with which they are predicted to interact. In addition to the withdrawn drugs, seven currently marketed drugs were also analyzed to find out the difference in the number of antitargets. These prediction results are presented in Table 6. The results show that more than seven antitargets were predicted for each withdrawn drug, while less than three antitargets were predicted for each existing drug. Thus, our service can successfully be applied to the selection and prioritization of compounds during the drug discovery process.

CONCLUSIONS

We developed QSAR models for 32 antitarget end-points based on MNA and QNA descriptors and self-consistent regression. These models showed good accuracy of prediction for 91% of the activities and high coverage of the external test sets for all end-points. Thus, the QSAR models could successfully be used for filtering out chemical compounds with a high probability of antitarget activity during the R&D process.

Our freely available online service for quantitative prediction of antitarget interaction profiles of chemical compounds is useful for researchers to increase the efficacy of finding drug-like leads with desirable pharmacological effects but without the side effects and toxicity caused by interactions with antitargets.

■ ASSOCIATED CONTENT

● Supporting Information

Detailed description of the QNA descriptors and the PASS algorithm, the chemical structures used in this work, and an overlap analysis of the data sets. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Tel: +1-301-846-5993. E-mail: alexey.zakharov@nih.gov.

Funding

This work is partially supported by RFBR grant number 12-07-00597 (to D.A.F. and A.A.L.) and RFBR/NIH grant number 12-04-91445/RUB1-31081-MO-12 (to V.V.P.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the two anonymous reviewers for their many helpful suggestions to improve the manuscript.

■ ABBREVIATIONS

GUSAR, general unrestricted structure–activity relationships; PASS, prediction of activity spectra for substances; SCR, self-consistent regression; QSAR, quantitative structure–activity relationships; MNA, multilevel neighborhoods of atoms descriptors; QNA, quantitative neighborhoods of atoms descriptors

■ REFERENCES

- (1) European Federation of Pharmaceutical Industries and Associations. (2010) in *The Pharmaceutical Industry in Figures*, EFPIA, Brussels, Belgium.
- (2) Czerepak, E., and Ryser, S. (2008) Drug approvals and failures: implications for alliances. *Nature Rev. Drug Discovery* 7, 197–198.
- (3) Schuster, D., Laggner, C., and Langer, T. (2008) Why Drugs Fail: A Study on Side Effects in New Chemical Entities, in *Antitargets* (Vaz, R. J., and Klabunde, T., Eds.) pp 3–22, Wiley-VCH, Weinheim, Germany.
- (4) Larrey, D., Berson, A., Habersetzer, F., Tinel, M., Castot, A., Babany, G., Letteron, P., Freneaux, E., Loeper, J., Dansette, P., and Pessyre, D. (1989) Genetic predisposition to drug hepatotoxicity: role in hepatitis caused by amineptine, a tricyclic antidepressant. *Hepatology* 10, 168–173.
- (5) FDA Duract Voluntarily Withdrawn, <http://www.fda.gov/downloads/Safety/MedWatch/SafetyInformation/UCM189811.zip> (accessed Dec 14, 2010).
- (6) FDA (2004) FDA Issues Public Health Advisory on Vioxx as Its Manufacturer Voluntarily Withdraws the Product, in *FDA News*, FDA, Silver Spring, MD.
- (7) Dukes, M. N. G., and Aronson, J. K. (2006) In *Meyler's Side Effects of Drugs*, 14th ed., An Encyclopedia of Adverse Reactions and Interactions, Elsevier, Atlanta, GA.
- (8) Klabunde, T., and Evers, A. (2005) GPCR anti-target modeling: pharmacophore models for biogenic amine binding GPCRs to avoid GPCR-mediated side effects. *ChemBioChem* 6, 876–889.
- (9) Chen, Y. Z., and Ung, C. Y. (2001) Prediction of potential toxicity and side effect protein targets of a small molecule by a ligand–protein inverse docking approach. *J. Mol. Graphics Modell.* 20, 199–218.
- (10) Fanelli, F., Menziani, M., Carotti, A., Benedetti, P. A., and Benedetti, De. (1994) Theoretical quantitative structure–activity relationship analysis on three dimensional models of ligand–m1 muscarinic receptor complexes. *Bioorg. Med. Chem.* 2, 195–211.
- (11) Lopez-Rodriguez, M., Rosado, M., Benhamu, B., Morcillo, M., Fernandez, E., and Schaper, K. (1997) Synthesis and structure–activity relationships of a new model of arylpiperazines. 2. Three dimensional quantitative structure–activity relationships of new hydantoin–phenyl–piperazine derivatives with affinity for 5-HT1A and alpha1 receptors. A comparison of CoMFA models. *J. Med. Chem.* 40, 1648–1656.
- (12) Vistoli, G., Pedretti, A., Villa, L., and Tesla, B. (2005) Range and sensitivity as descriptors of molecular property spaces in dynamic QSAR analyses. *J. Med. Chem.* 48, 4947–4952.
- (13) Hegyemegi-Barakonyi, B., Eros, D., Szántai-Kis, C., Breza, N., Bánhegyi, P., Szabó, G., Várkonyi, E., Peták, I., Orfi, L., and Kéri, G. (2009) Tyrosine kinase inhibitors - small molecular weight compounds inhibiting EGFR. *Curr. Opin. Mol. Ther.* 11 (3), 308–321.
- (14) Matthews, E., and Frid, A. (2010) Prediction of drug-related cardiac adverse effects in humans. A: Creation of a database of effects and identification of factors affecting their occurrence. *Regul. Toxicol. Pharmacol.* 56 (3), 247–275.
- (15) Filimonov, D. A., Zakharov, A. V., Lagunin, A. A., and Poroikov, V. V. (2009) QNA based 'Star Track' QSAR approach. *SAR and QSAR Environ. Res.* 20 (7–8), 679–709.
- (16) Lagunin, A. A., Zakharov, A. V., Filimonov, D. A., and Poroikov, V. V. (2011) QSAR modelling of rat acute toxicity on the basis of PASS prediction. *Mol. Inf.* 30 (2–3), 241–250.
- (17) Lagunin, A. A., Zakharov, A. V., Filimonov, D. A., and Poroikov, V. V. (2007) A new approach to QSAR modelling of acute toxicity. *SAR QSAR Environ. Res.* 18 (3–4), 285–298.
- (18) Sady, A., Lagunin, A., Filimonov, D., and Poroikov, V. (2003) Prediction of biological activity spectra via Internet. *SAR QSAR Environ. Res.* 14, 339–347.
- (19) Maurer, I., and Volz, H. (2001) Cell-mediated side effects of psychopharmacological treatment. *Arzneimittelforschung* 51 (10), 785–792.
- (20) Rang, H. P., Dale, M. M., and Ritter, J. M. (1999) In *Pharmacology*, 4th ed., Churchill Livingstone, Philadelphia, PA.
- (21) Gutstein, H.B., and Akil, H. (2006) In *Goodman & Gilman's The Pharmacological Basis of Therapeutics* (Brunton, L. L., Lazo, J. S., and Parker, K. L., Eds.) 11th ed., pp 547–590, The McGraw-Hill Companies, New York.
- (22) Maus, M., Glowinski, J., and Premont, J. (2002) GABA is toxic for mouse striatal neurones through a transporter-mediated process: GABA toxicity in striatal neurones. *J. Neurochem.* 82 (4), 763–773.
- (23) Lei, B., Xi, L., Li, J., Liu, H., and Yao, X. (2009) Global, local and novel consensus quantitative structure–activity relationship studies of 4-(phenylaminomethylene) isoquinoline-1,3 (2H, 4H)-diones as potent inhibitors of the cyclin-dependent kinase 4. *Anal. Chim. Acta* 644 (1–2), 17–24.
- (24) Overington, J. P., Al-Lazikani, B., and Hopkins, A. L. (2006) How many drug targets are there? *Nat. Rev. Drug Discovery* 5, 993–996.
- (25) ChemAxon Ltd, MarvinSketch 5.4.1.1, <http://www.chemaxon.com/marvin/help/index.html> (accessed May 10, 2011).
- (26) FDA. (1992) In *Omniflox Recalled - Antibiotic*, FDA Consumer, U.S. Government Printing Office, Washington, DC.