

Scientific paper

Chemometrical Exploration of Combinatorially Generated Drug-like Space of 6-fluoroquinolone Analogs: A QSAR Study

Nikola Minovski and Tom Šolmajer*

National Institute of Chemistry, Hajdrihova 19, POB 660, 1001 Ljubljana, Slovenia

* Corresponding author: E-mail: tom.solmajer@ki.si,
tel: +386 1 4760 227, fax: +386 1 4760 300

Received: 08-03-2010

This paper is dedicated to Professor Milan Randić on the occasion of his 80th birthday

Abstract

A classical virtual combinatorial chemistry approach (*CombiChem*) was applied for combinatorial generation of 5590 novel structurally-similar 6-fluoroquinolone analogs by using a virtual synthetic pathway with selected primary (43) and secondary amines (130). The obtained virtual combinatorial library was filtered using an *in-house* developed set of cheminformatics drug-likeness filters with pre-integrated Boolean options (TRUE/FALSE) for compounds reduction/selection. The retained number (304) of fluoroquinolone analogs (with TRUE outcome) defines the drug-like chemical space (*CombiData*). Quantitative structure-activity relationships (QSAR) study on these 304 virtually generated 6-fluoroquinolone analogs with unknown activity values was performed using a pre-built five-parameter multiple linear regression (MLR) model developed on a set of compounds with experimentally determined activity values ($R_{tr} = 0.8417$, $R_{tr-cv} = 0.7884$). The obtained activity values for the unknown compounds together with the model results were used to define the applicability domain (AD). The obtained AD offers a good graphical representation and establishment of structure-activity relationships (SAR) which could be used for design of new 6-fluoroquinolones with possible better activity.

Keywords: Tuberculosis, fluoroquinolones, DNA gyrase, CombiChem, QSAR, Multiple Linear Regression

1. Introduction

Tuberculosis (TB), the acutely transmissible bacterial infection, is still one of the leading threats worldwide. The causative agent of tuberculosis, *Mycobacterium tuberculosis*, is a persistent pathogen that latently infected approximately one third (two billion people) of the human population and around two million people die from tuberculosis every year worldwide (World Health Organization, 2003).¹ The tuberculosis in first instance is caused by the pathogen *M. tuberculosis*, but there are also several cases where the microorganisms such as *M. fortuitum*, *M. smegmatis* and *M. avium-intracellulare* complex (MAC) are involved in the TB development.

Although tuberculosis can be cured with chemotherapy, the whole treatment is extensively long and takes around 6–9 months. The major factors which often lead to

development of drug resistant as well as deadly multidrug resistant tuberculosis strains, are ascribed to the durability of the treatment, the toxicity of drugs and very frequently the poor patient compliance to the therapy regimen. The increasing problem of multidrug resistant tuberculosis strains is one of the challenges for design and development of new chemotherapeutics which will not only be active against resistant mycobacteria, but also shorten the length of therapy.²

The current chemotherapy of tuberculosis is based on drugs which act as cell wall biosynthesis inhibitors (isoniazide, ethambutol, ethionamide, cycloserine) as well as nucleic acid synthesis inhibitors (rifampicin, quinolones).³ The last (quinolones) are particularly interesting because of their quite invasive mechanism of bactericidal action. One of the validated molecular targets of antitubercular drugs in mycobacteria is DNA gyrase, a unique bacterial type II topoisomerase enzyme responsible for ca-

talysis of the process of introduction of negative supercoils into the double-stranded bacterial DNA molecule.⁴ This bacterial enzyme forms a functional heterodimer A_2B_2 consisting of two major subunits, GyrA and GyrB. The GyrA subunit is responsible for the process of breakage and reunion of the double-stranded DNA, where GyrB is also involved. Another bacterial enzyme is type IV topoisomerase that also forms a heterodimer and together with the topoisomerase II is involved in the process of controlling the topological state of DNA molecule.⁵ The gyrase is important for activation of the process of DNA replication and elongation, while topoisomerase IV is responsible for relaxation of the DNA strands.^{6,7} Fluoroquinolones belong to the quinolone's group of inhibitors which inhibit gyrase/topoisomerase IV. Their mechanism of action is based on the inhibition of the process of DNA synthesis in mycobacteria through a cleavage of the native mycobacterial DNA molecule in the complex formed between the DNA gyrase and type IV topoisomerase. These processes lead to topological perturbation of DNA and bacterial cell death.⁸

Quinolones belong to the class of broad-spectrum chemotherapeutics originating from the nalidixic acid which is the parent of the group. One of the important varieties of quinolone antibacterials is the 6-fluoroquinolone subset, which have an F atom attached to the central ring system at the 6 position (Figure 1).

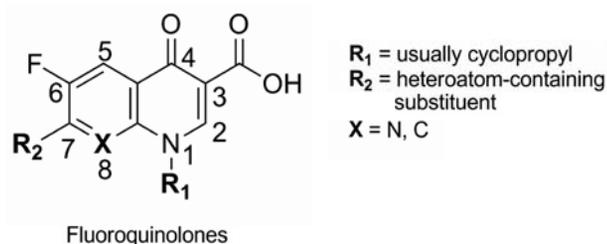


Figure 1. Generic structure of 6-fluoroquinolone chemotherapeutics.

Structure-activity relationships (SAR) studies show that main quinolone core (1,4-dihydro-4-oxo-3-pyridine-carboxylic acid moiety) is most important for activity.⁹ The substitution with F atom at position 6 is important and will result in greatly enhanced anti-mycobacterial activity. The position 1 of the main ring system, can also be substituted (lower alkyl substituents such as methyl, ethyl, and especially cyclopropyl, enhance potency and efficiency). These substitutions result in increased activity and metabolic stability of the drug through a steric bulk. Ring condensations at the positions [1,8], [5,6], [6,7] and [7,8] can also significantly increase the activity. Substitutions at position 2 of the main scaffold significantly reduce activity and potency, but on the other hand the positions 5, 6, 7 (especially), and position 8 of the fused ring core system greatly increase the anti-mycobacterial activity and potency.¹⁰

The development of new SAR rules is one of the major challenges in modern drug discovery. The total number of molecules with “drug-like” characteristics has been estimated to be approximately 10^{63} .^{11,12} The medicinal chemist's purpose is not to investigate such large pool of compounds, but to generate, isolate and identify some discrete sub-spaces of compounds that interact with the biological systems.¹³ The majority of these frequently used methods need a previous knowledge (knowledge-based methods) about the mechanism of action of the agent under investigation.¹⁴ Such a method for generating “drug-like” sub-spaces is the well known combinatorial chemistry approach. The exploration process is focused on a small number of high-quality molecules with good and well established drug-like properties.

The present study involves generation of a high-quality drug-like space of novel unknown 6-fluoroquinolone analogs. Virtual compounds were generated using the combinatorial enumeration and their selection was performed by using a pre-built statistical model and a comprehensive set of calculated molecular descriptors for prediction of the unknown activity values. The ability to navigate through the generated 6-fluoroquinolone's drug-like space as well as to investigate and define new possible SAR increases the chance for selection of new 6-fluoroquinolone analogs with possible better activity.

2. Materials and Methods

In order to investigate the practical meaning of established QSAR models for prediction of unknown activity values as well as to make a subsequent selection of one or more compounds as potentially active 6-fluoroquinolones, we used a combinatorially generated external dataset of unknown 6-fluoroquinolone analogs for which the activity values ($pMIC$, negative decade logarithm of MIC) were not known. There are several software packages which work with the combinatorial algorithm (combinatorial enumeration). For these purposes we used ChemBio-Office Ultra v11.0 (2008) software suite and its specifically integrated add-on modules for tabular data mining (*CombiChem*).¹⁵ These modules are supplied with several useful pre-integrated cheminformatics functions.

For building a high-quality combinatorial library in *in-house* conditions a viable approach is to start with an already built commercial structural database (already known drugs, other known/unknown compounds, different building-blocks, etc.). A crucial step before starting with the process of combinatorial enumeration is pre-filtering of the compounds (building-blocks) in the database using cheminformatics fragment-likeness filters such as Astex rule of 3 (Ro3).¹⁶ This procedure ensures the location of building-blocks selected (potential fragment leads) within the lead-like space. Such a database (consisting of 941 building-blocks) obtained from online internet source¹⁷ was used as a starting point for reactant selection for combinatorial enumeration.

2. 1. Combinatorial Chemistry Approach

By following the synthetic methodology of 6-fluoroquinolones (generic synthetic reaction) it was easy to define the generic combinatorial (virtual) synthetic pathway for *in silico* generation of a set of structurally-similar 6-fluoroquinolone analogs. In order to ensure that our final products will have similar motifs as already known 6-fluoroquinolones with known and confirmed activity against mycobacteria, as template for building the combinatorial synthetic pathway we used the synthetic scheme of the well known 6-fluoroquinolone ciprofloxacin.¹⁸ Since the synthetic procedure was performed in a totally virtual manner (*in silico* environment), the catalysts and other auxiliary compounds involved through the original synthetic steps were not taken into account. Thus, the virtual combinatorial synthetic pathway was reduced to two fundamental steps (Figure 2):

- (1) Amination of the starting material using primary amines and subsequent cyclization to build a common 1-monosubstituted (R_1) 6-fluoroquinolone scaffold (intermediate).
- (2) Amination of the intermediate product using secondary amines in order to produce 1,7-disubstituted (R_2, R_3) 6-fluoroquinolone moiety as a final product.

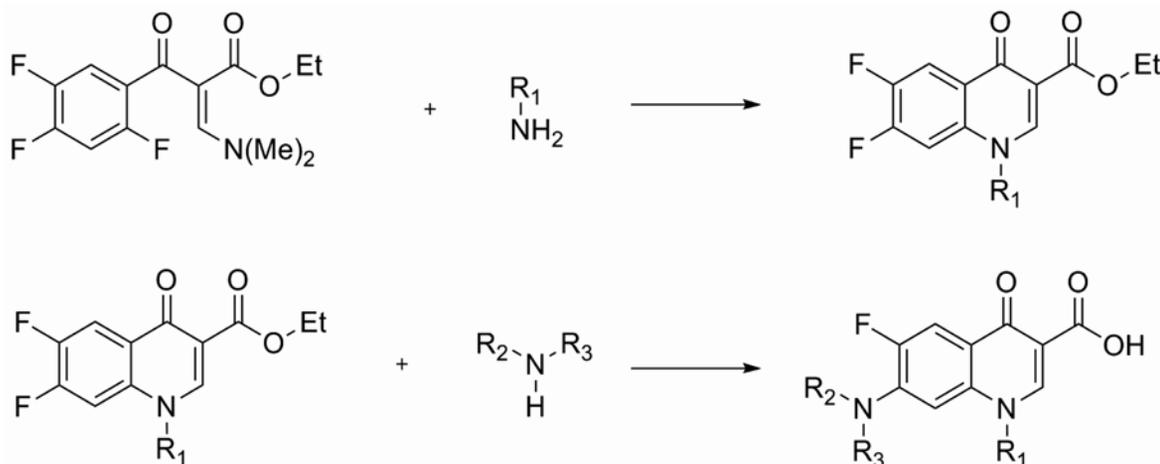


Figure 2. Generic virtual synthetic pathway for combinatorial enumeration of 6-fluoroquinolone analogs.

2. 1. 1. Reactants Selection for Combinatorial Enumeration

Substructure search (SSS) algorithm was implemented for selection of reactants subsets (primary and secondary amines) which were used for enumerating all possible combinations (products). This procedure resulted in extraction of all possible substructural fragments from the starting pool of 941 building-blocks (in which 49 primary amines and 179 secondary amines were pre-

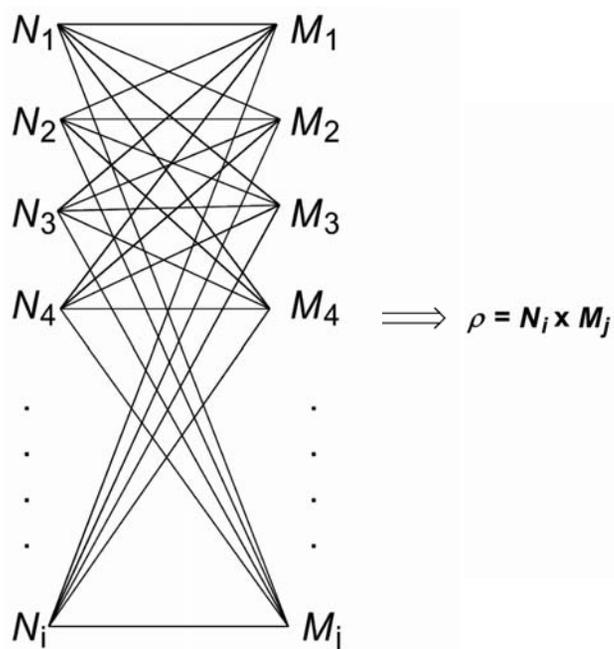


Figure 3. Graphical representation of the algorithm for combinatorial enumeration defined by ($N_1 \dots N_i$) which are the building-blocks (fragments) from the reactant 1 subset, ($M_1 \dots M_j$) are the building-blocks (fragments) from the reactant 2 subset, whereas ρ is the total number of the combinatorially-obtained products.

sent) contained in the database. Each entry in the subsets was visually inspected for presence of non-individualistic forms (salt forms, dual and triple forms, non-electroneutral forms). These entries were eliminated from the virtual synthetic pathway through a simple filtering procedure employing a Boolean operational algorithm (Y/N (yes/no)). The retained substructural fragments (43 primary amines and 130 secondary amines, separately) with (Y) outcome were subsequently used for combinatorial enumeration.

2. 1. 2. Combinatorial Enumeration

The process of combinatorial enumeration is a straightforward process of statistical non-repetitive permutation where each fragment of a building-blocks subset (*reactant 1*) interact with each fragment of another building-blocks subset (*reactant 2*) within the previously defined core (main molecular scaffold) as presented in Figure 3.

For a system configured of two subsets of reactants (Figure 2), the total number of final possible products (6-fluoroquinolone analogs) can be calculated as a multiplication product between each pair of the fragments (*reactant 1* and *reactant 2*). Mathematically this process can be described with the following simple equation (Eq.1):

$$\rho = N_i \times M_j \quad (1)$$

where N_i is the total number of fragments in the *reactant 1* subset, M_j is the total number of fragments in the *reactant 2* subset, whereas ρ is the total number of the obtained products.¹⁹ Thus, for $N_i = 43$ (*reactant 1*, primary amines) and $M_j = 130$ (*reactant 2*, secondary amines), the total number of all possible combinations (6-fluoroquinolone structural analogs) will be $\rho = 5590$. The obtained combinatorial library of 5590 structural analogs was subsequently used for assessing the quality and for defining the drug-like chemical space of novel 6-fluoroquinolones.

2. 2. Defining the Drug-like Chemical Space

In order to be bioavailable, a drug molecule must be transported through the biological membranes to reach the systemic circulation. According to this postulate, molecular properties that correlate with the poor membrane permeability (in the absence of active transport) can be effectively used as a drug-likeness filter for elimination of undesirable molecules. Lipinski et al.²⁰ stated that a drug molecule overpassing any two of the following rules is likely to be poorly absorbed:

- (1) molecular weight (mass) less than 500 Da.
- (2) number of hydrogen bond donors (OH/NH groups) equal or less than 5.
- (3) Number of hydrogen bond acceptors (O/N) less than 10.
- (4) Calculated logP less than 5.0 (by ClogP).

Another important molecular parameter that strongly correlates with the membrane permeability is polar surface area (PSA) defined as a sum of van der Waals surface areas of the polar atoms in the molecule (N and O). Veber et al.,²¹ assessed the influence of PSA of the molecule over the oral bioavailability in rats and found that a drug molecule will have a poor oral bioavailability if the following criteria passed:

- (1) PSA equal or less than 140 \AA^2 (or 12 or fewer H-bond donors and acceptors).
- (2) number of rotatable bonds is equal or less than 10.

According to these definitions, we developed three different algorithms (Lipinski, Veber and CombiVL (Lipinski-Veber combination)) which were implemented as drug-likeness filtering tools. These drug-likeness filters were constructed to work as a Boolean operator (TRUE/FALSE). We assessed our combinatorial library of 5590 structures employing all three filtering algorithms and we found that CombiVL pre-selectional algorithm offers the best outcome (304 structures out of 5590 were signed with TRUE as most promising compounds for further analysis). The best results (structures marked with TRUE) were isolated and subsequently used as an external dataset (*CombiData*) for prediction of their unknown activity values ($pMIC_{pred-combi}$) as well as the selection of new possible active principles.^{22,23} The chemical structures of the combinatorially obtained 6-fluoroquinolones (*CombiData*, 304 compounds) as well as the selected *reactant 1* and *reactant 2* substructural fragments are available as Supplementary material (Table S1).

2. 3. Applicability Domain

According to OECD QSAR Validation Principles, a QSAR model is usable in the boundaries of its applicability domain.²⁴ The applicability domain of a (Q)SAR is defined as physico-chemical, structural, or biological space, knowledge or information on which the training set of the model has been developed, and for which it is applicable to make predictions for new compounds.²⁵ It should be described in terms of molecular descriptors of the model which are the most relevant parameters. The activity predictions can be made only within the domain's boundaries. Therefore, the applicability domain can be defined as a theoretical region in the space represented by the model's descriptors and the response (predicted activity values) in which a (Q)SAR gives reliable outcome.

The drug-like chemical space (applicability domain) of the investigated 6-fluoroquinolones (pre-built five-parameter MLR model) was calculated using the leverage approach.²⁶ The plot defined by standardized residuals as a function of the leverage (Williams plot) was employed to visualize the drug-like chemical space. The leverage is defined as a compound's distance from the centroid of X . Mathematically, the leverage (h_i) of a given compound in the multidimensional descriptor space, can be calculated as (Eq.2):

$$h_i = x_i^T (X^T X)^{-1} x_i \quad (2)$$

where x_i is the descriptor vector of the compound under investigation, and the X is the descriptor matrix rendered from the descriptor values of the training set.²⁷ According to Eriksson et. al.,²⁸ the cut-off leverage value (h^*) is defined as:

$$h^* = \frac{3(p+1)}{n} \quad (3)$$

where n is the number of compounds in the training set, whereas p is the number of descriptors used for modeling. In our model the cut-off value for $h^* = 0.391$ (Eq.3). Eriksson et. al., proposed that the prediction for compounds with ($h_i > h^*$) can be considered as unreliable, and vice versa. The value 3 for standardized residuals in the Williams plot is frequently used as a limit (cut-off value) for accepting predictions (3.0 standard deviation units, $\pm 3.0\sigma$). The compounds that lie in this region cover 99% of the normally distributed data.²⁴

2. 4. Prediction of Activity Values

The prediction of activity values ($pMIC$) for the dataset of 304 unknown 6-fluoroquinolone analogs (*CombiData*) was performed using a pre-built five-parameter multiple linear regression (MLR) model. For modeling and testing the predictive performances we employed a dataset of 65 fluoroquinolone structures (*Assay3*) active against *M. tuberculosis* as well as a large set of approximately 600 molecular descriptors.²⁹ The selection of statistically significant molecular descriptors for activity was performed using Heuristic algorithm and intercorrelation matrix.³⁰ This procedure resulted in a model with five molecular descriptors most important for activity and simultaneous elimination of the molecular descriptors for which intercorrelation coefficients were more than or equal than 0.40, i.e. $R^2(P_i, P_m) \leq 0.40$. An upper limit of 0.40 was proposed for R^2 to eliminate the chance correlation.³¹

The dataset was previously divided using random dividing approach into training set (*Assay3*, 46 structures) and the external validation set (*Assay3*, 19 structures). The model was built on the training set ($R_{tr} = 0.8417$, $R_{tr-cv} = 0.7884$) using MLR method and subsequently was validated for its predictive power using the external validation set ($R_{val} = 0.7993$). This model was subsequently used for further analysis and graphical representation of 304 virtual compounds from the drug-like chemical space.

3. Results and Discussion

Combinatorial chemistry approach (*CombiChem*) as a methodology for building combinatorial explosion i.e. defining the chemical space of structurally similar analogs as well as its exploration employing different cheminformatics tools and algorithms is well documented.^{32–36} In order to build a combinatorial library of small structural analogs for medicinal chemistry purposes (drug-like chemical space), first the generic synthetic methodology for the entity of interest must be taken into account, and second one or more subsets of building-blocks (structural fragments) with acceptable quality must be utilized. The aim of our study was to generate such a combinatorial library of new structurally-similar 6-fluoroquinolone ana-

logs in a virtual environment using mathematical-chemistry algorithms for combinatorial enumeration as well as to explore the structural diversity (drug-like chemical space) in order to establish SAR rules which could subsequently be used for design of new 6-fluoroquinolones with possible enhanced activity.

Using the approved synthetic pathway of the well known 6-fluoroquinolone chemotherapeutic ciprofloxacin (Figure 2) as a template for virtual synthesis for generation of all possible analog's combinations and a lead-likeness pre-filtered set of building-blocks, initially we generated a large dataset of 5590 compounds. The quality of the combinatorial dataset was investigated through analysis of the druggability properties: MW (molecular weight (mass)), Hansch-Leo's calculated partition coefficient for n -octanol/water bi-phase system (ClogP), number of hydrogen-bond donors ($nHBD$), number of hydrogen-bond acceptors ($nHBA$), polar surface area (PSA, calculated as TPSA (topological polar surface area)) and number of rotatable bonds (nRB). The druggability was assessed by implementing histogram-type of analysis for graphical representation i.e. properties distribution comparison³⁷ between the dataset used for modeling (*Assay3*, 46 structures, Figure 4a)²⁹ and the combinatorial one (5590 structures, Figure 4b). Normal Gaussian distribution was observed for MW,

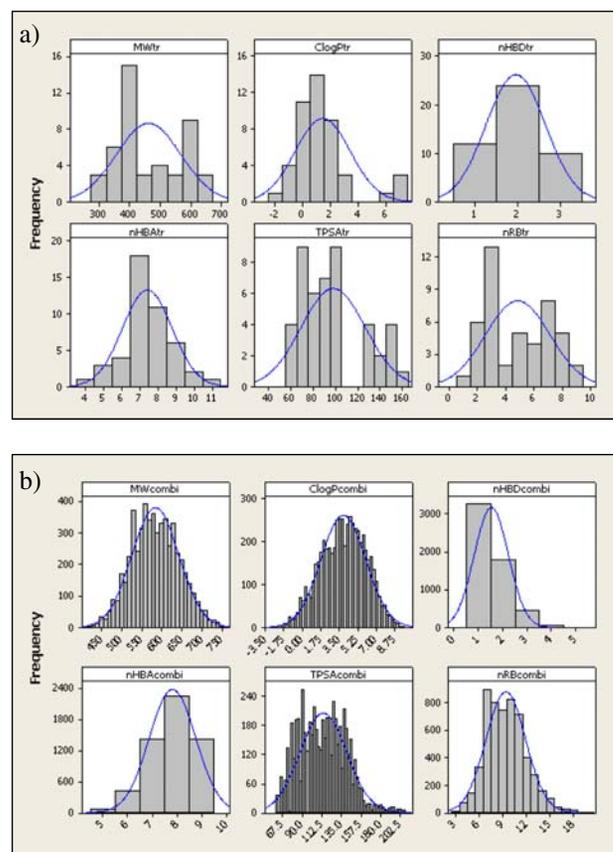


Figure 4. Histogram-type of analysis for assessing the properties distribution. **a)** Training set (*Assay3*), 46 compounds. **b)** Combinatorial set, 5590 compounds.

ClogP and nRB , whereas for other properties ($nHBD$, $nHBA$ and $TPSA$) an asymmetric distribution was observed in the two datasets. The peak analysis (point where the distribution of the property of interest reaches 50%) in Figure 4a (training set of 6-fluoroquinolone compounds with approved activity) shows the optimal molecular parameters for drug-likeness: $MW_{tr} = 463$, $ClogP_{tr} = 1.40$, $nHBD_{tr} = 2$, $nHBA_{tr} = 7$, $TPSA_{tr} = 98$, $nRB_{tr} = 5$. The corresponding mid-50% values (the intervals between the 25% and 75% of the distribution) were approximately: $MW_{tr} = 406$, $ClogP_{tr} = (-0.09)–2.91$, $nHBD_{tr} = 1–3$, $nHBA_{tr} = 6–9$, $TPSA_{tr} = 73–121$, $nRB_{tr} = 3–6$. These results apparently show the attendance of Lipinski's "rule of five" as well as Veber's rules for oral bioavailability as measures for drug-likeness.^{20,21} As a comparison, for combinatorially generated dataset the situation for property distribution (peak analysis) was as follows: $MW_{combi} = 585$, $ClogP_{combi} = 3.80$, $nHBD_{combi} = 2$, $nHBA_{combi} = 8$, $TPSA_{combi} = 113$, $nRB_{combi} = 9$, and the mid-50% values were approximately in the interval: $MW_{combi} = 516–652$, $ClogP_{combi} = 1.27–6.35$, $nHBD_{combi} = 1–2$, $nHBA_{combi} = 7–9$, $TPSA_{combi} = 87–140$, $nRB_{combi} = 6–12$ (Figure 4b).

Two out of six parameters (MW_{combi} and $ClogP_{combi}$) show increased values, while the other parameters ($nHBD_{combi}$, $nHBA_{combi}$, $TPSA_{combi}$, and nRB_{combi}) were again in the drug-likeness boundaries. The increased values for MW_{combi} and $ClogP_{combi}$ significantly point to the increased molecular complexity. Hann et al.,³⁸ demonstrated that the probability of a good ligand-receptor interaction, quickly decreases with increased molecular complexity. According to this observation, some sort of drug-likeness pre-filtering must be implemented in order to define the drug-like chemical sub-space. The distribution analysis results for each property (mean and standard deviation values) are presented in Table 1.

Table 1. The property distribution analysis results for the training set and the combinatorial set (N_{tr} , number of training set objects; N_{combi} , number of combinatorially generated compounds; Prop, Property; StDev, Standard Deviation).

	N_{tr}	Prop	Mean	StDev
Training set	46	MW_{tr}	462.90	105.90
		$ClogP_{tr}$	1.40	2.00
		$nHBD_{tr}$	1.96	0.70
		$nHBA_{tr}$	7.44	1.38
		$TPSA_{tr}$	97.61	28.92
		nRB_{tr}	4.85	2.29
	N_{combi}	Prop	Mean	StDev
Combinatorial set	5590	MW_{combi}	584.800	58.660
		$ClogP_{combi}$	3.842	2.137
		$nHBD_{combi}$	1.524	0.704
		$nHBA_{combi}$	7.820	0.934
		$TPSA_{combi}$	113.400	26.980
		nRB_{combi}	9.376	2.524

The drug-like chemical sub-space was assessed through implementation of three different algorithms (Lipinski, Veber, CombiVL (Lipinski-Veber combined filter)). The best outcome was observed using the third combined filter (CombiVL) which resulted in 304 compounds out of 5590 combinatorially-generated structures. Such an algorithm ensures that each of the retained structures have optimal drug-like properties according to drug-likeness rules ($MW < 500$, $ClogP < 5.0$, $nHBD \leq 5$, $nHBA < 10$, $TPSA \leq 140$, $nRB \leq 10$). These filtered 6-fluoroquinolone compounds define the drug-like chemical space (*CombiData*) which was used for prediction of unknown activity values of the compounds ($pMIC_{pred-combi}$).

The activity of the unknown 304 filtered structures was predicted using a five-parameter MLR model pre-built on a dataset for which the biological activity values was known (*Assay 3*, 65 compounds).²⁹ These results indicated that activity ($pMIC$) can be correlated with the five most important molecular descriptors: MSD, D/Dr03, MATS7p, GATS8e and EEig05d. The MSD and D/Dr03 parameters belong to the class of topological and constitutional parameters and clearly describe the importance of the shape i.e. main quinolone core (1,4-dihydro-4-oxo-3-pyridinecarboxylic acid moiety) for activity. On the other hand, MATS7p, GATS8e and EEig05d, are molecular descriptors from the class of topological and electro-topological parameters. These molecular parameters describe the possibility for good accommodation of the drug into the binding pocket (GyrA subunit) as well as possible π - π stacking interactions between the aromatic quinolone scaffold and planar aromatic systems into the bacterial DNA molecule. Such an interaction will procure a topological stress of the DNA, inhibition of the replication/transcription processes and cell-death.³⁹ EEig05d alone is also a pure electrostatic parameter and it is of significant importance for the anti-mycobacterial activity. This molecular descriptor, clearly indicates that *in vitro/in vivo* anti-mycobacterial activity against *M. tuberculosis*, is potentially dependent on the charge indices for the O atom (sp^2) of the carboxyl and carbonyl group in the main core (Figure 1). This molecular parameter describes the possibility for establishing hydrogen bonding interactions between these substituents and the amino acid residues within the GyrA binding pocket. Another electrostatic parameter in the model is the GATS8e. This parameter suggest that the position 6 of the main 6-fluoroquinolone core is of significant importance for good accommodation of the molecule into the active binding site and suggest the possibility of establishing an electrostatic interaction between the F atom in position 6 and the target (possible inter-molecular electrostatic interactions with the amino acid residues of the GyrA subunit active site).³⁹ These electrostatic interactions may result in enhanced binding of the 6-fluoroquinolone analog to the complex.⁴⁰ The parameters (molecular descriptors) used for modeling/prediction with their detailed description are presented in Table 2, whe-

Table 2. Specification of molecular descriptors used for modeling/prediction (C, Constitutional; T, Topological; E, Electrostatic).

ID	Descriptor	Source	Definition	Class
1	MSD	DRAGON	Mean square distance index (Balaban)	T
2	D/Dr03	DRAGON	Distance/detour ring index of order 3	C/T
3	MATS7p	DRAGON	Moran autocorrelation-lag7/weighted by atomic polarizabilities	T/E
4	GATS8e	DRAGON	Geary autocorrelation-lag8/weighted by atomic Sanderson electronegativities	T/E
5	EEig05d	DRAGON	Eigenvalue 05 from edge adj. matrix weighted by dipole moments	T/E

Table 3. The experimental vs. predicted activity values for training set compounds ($pMIC_{exp-tr}$, $pMIC_{pred-tr}$) together with the numerical values of the molecular descriptors involved in the modeling procedure.

ID	$pMIC_{exp-tr}$	$pMIC_{pred-tr}$	MSD	D/Dr03	MATS7p	GATS8e	EEig05d
1	-0.477	0.360	0.204	0.000	0.239	0.519	3.287
2	1.939	0.797	0.203	0.000	0.321	0.562	2.909
3	0.523	0.369	0.222	0.000	0.272	0.663	2.843
4	-0.544	-0.109	0.200	0.000	0.430	0.627	3.228
5	1.000	1.227	0.197	35.550	0.386	0.452	3.320
6	-0.146	0.130	0.197	0.000	0.386	0.452	3.306
7	1.301	0.558	0.200	0.000	0.235	0.459	3.285
8	0.301	0.719	0.205	0.000	-0.111	1.388	3.312
9	-0.204	-0.528	0.205	0.000	0.380	2.388	2.688
10	-0.301	0.263	0.222	0.000	0.303	0.646	2.853
11	0.602	1.185	0.211	36.263	0.250	0.653	3.257
12	2.000	2.289	0.204	34.391	-0.004	0.573	3.237
13	1.509	0.393	0.197	0.000	0.118	1.257	3.285
14	0.903	1.510	0.221	36.985	0.205	0.781	2.941
15	2.398	2.369	0.203	37.719	0.005	0.484	3.291
16	1.222	1.650	0.197	38.954	0.278	0.647	3.255
17	2.699	2.787	0.197	83.212	0.312	0.675	3.323
18	1.921	1.458	0.201	37.496	0.386	0.356	3.200
19	1.796	2.063	0.203	37.719	0.085	0.516	3.301
20	0.222	0.408	0.197	0.000	0.299	0.430	3.319
21	2.799	1.731	0.223	54.141	0.104	0.580	3.341
22	-0.818	-0.157	0.228	0.000	0.081	0.667	3.350
23	-0.716	0.200	0.223	0.000	0.019	0.681	3.350
24	-0.810	0.187	0.222	0.000	-0.029	0.682	3.458
25	-0.799	-0.216	0.225	0.000	0.141	0.621	3.350
26	-0.797	-0.460	0.222	0.000	0.115	0.681	3.553
27	1.523	1.870	0.196	39.096	0.213	0.613	3.286
28	0.108	0.876	0.235	51.151	0.160	0.805	3.340
29	-0.796	-1.369	0.243	0.000	0.228	0.961	3.349
30	-0.496	0.397	0.232	55.659	0.334	0.871	3.393
31	1.204	1.791	0.195	42.157	0.047	1.293	3.393
32	2.527	1.458	0.223	56.156	0.063	0.641	3.566
33	2.572	2.202	0.218	63.619	-0.040	0.661	3.554
34	2.583	1.669	0.218	63.619	0.099	0.660	3.597
35	2.907	2.266	0.218	63.619	-0.044	0.638	3.537
36	2.857	2.354	0.218	63.619	-0.035	0.672	3.459
37	1.966	1.682	0.218	63.619	0.115	0.579	3.597
38	2.222	1.437	0.199	40.052	0.298	0.708	3.291
39	3.000	2.110	0.203	37.719	0.085	0.516	3.276
40	0.409	0.782	0.196	43.134	0.489	1.027	3.275
41	1.903	2.677	0.202	65.911	0.055	0.654	3.459
42	-0.100	-0.999	0.226	0.000	0.328	0.747	3.370
43	0.482	-0.411	0.228	0.000	0.128	0.752	3.367
44	0.796	1.454	0.225	67.219	0.111	0.733	3.596
45	0.495	-0.296	0.215	0.000	0.225	0.668	3.403
46	1.398	1.951	0.207	71.526	0.185	0.748	3.586

reas the experimental and predicted activity values ($pMIC_{exp-tr}$, $pMIC_{pred-tr}$) together with the numerical values of the descriptors used for modeling are showed in Table 3.

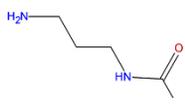
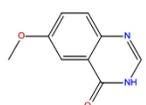
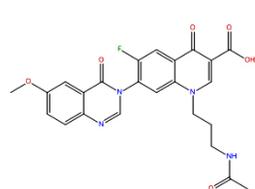
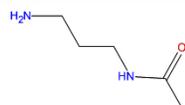
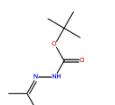
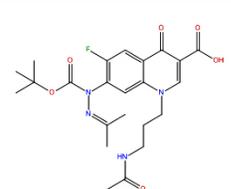
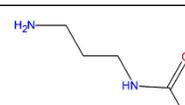
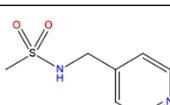
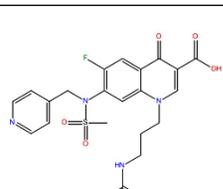
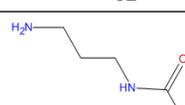
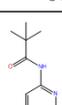
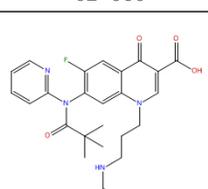
The predicted activity values ($pMIC_{pred-combi}$) for newly generated set of 304 unknown fluoroquinolones are in the range between ($-1.1041 < pMIC_{pred-combi} < 1.7989$) and the corresponding $MIC_{pred-combi}$ values (anti-logarithm of $pMIC_{pred-combi}$) are in the range ($0.0159 < MIC_{pred-combi} < 12.7087$). We opted to choose approximately 5% (15 compounds) from a pool of 304 combinatorially-generated compounds as possible most active compounds with activity values in the range $0.0159 < MIC_{pred-combi} < 0.0970$ (lower MIC value, higher activity). The corresponding $pMIC_{pred-combi}$ values for these 15 compounds which were used for assessing the applicability domain, are in the range ($1.0133 < pMIC_{pred-combi} < 1.7989$). Chemical structures of these compounds, corresponding synthetic codes (synthID = *reactant1-reactant2*)

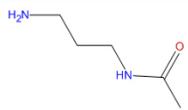
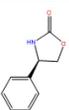
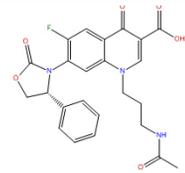
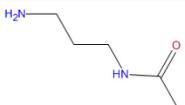
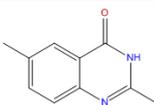
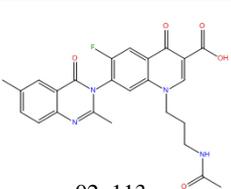
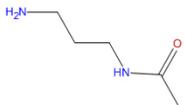
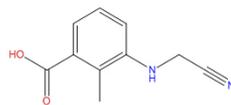
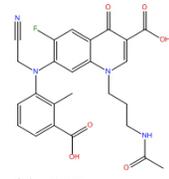
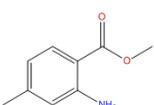
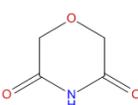
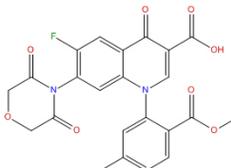
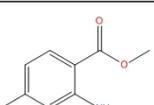
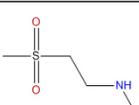
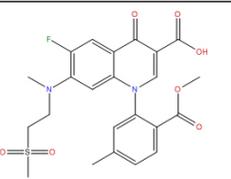
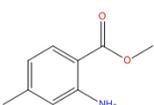
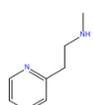
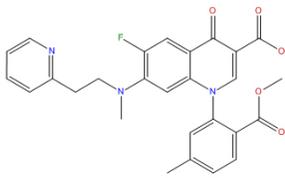
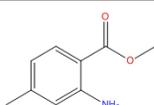
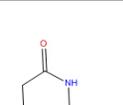
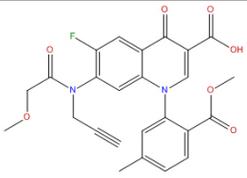
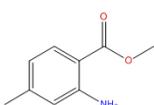
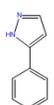
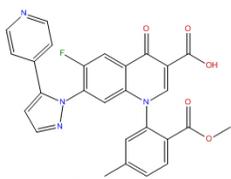
tant2) and predicted activity values ($pMIC_{pred-combi}$, $MIC_{pred-combi}$) are available in Table 4.

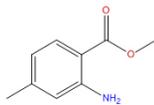
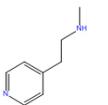
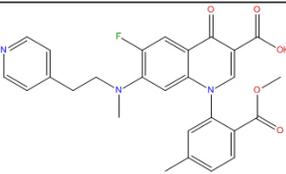
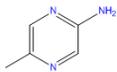
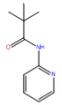
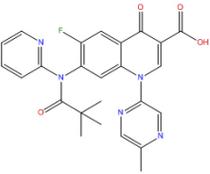
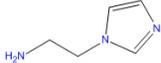
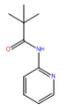
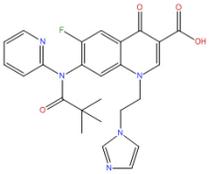
The applicability domain (AD) of the five-parameter linear model (Williams plot) was assessed employing the well known leverage approach (Figure 5). Training set objects (*Assay3*, 46 compounds with experimental activity values) used in the model development are presented as black solid dots, whereas the selected objects from the combinatorial set (*CombiData*, 15 compounds) as gray solid rectangles labeled with the corresponding synthetic code (synthID).

The analysis of AD for the training set objects shows that two compounds labeled with the codes *11* and *36* are outliers. Compound *11* is a typical *X*-outlier ($h > h^* = 0.391$), whereas compound *36* can be classified as an *Y*-outlier if the cut-off value for standard deviation is set as $\pm 2.0\sigma$. Beside this conclusion, the prediction for these compounds was quite good as presented in Table 3. According to h^* cut-off value of 0.391, it is apparent that

Table 4. Chemical structures of the selected compounds (*CombiData*, 15 compounds), corresponding synthetic codes (synthID = *reactant 1-reactant 2*) and predicted activity values ($pMIC_{pred-combi}$, $MIC_{pred-combi}$). The most promising 11 chemicals are signed with bold IDs.

ID	<i>reactant1</i>	<i>reactant2</i>	synthID	$pMIC_{pred-combi}$	$MIC_{pred-combi}$
1				1.1041	0.0787
	02	009	02-009		
2				1.2783	0.0527
	02	018	02-018		
3				1.0133	0.0970
	02	060	02-060		
4				1.5306	0.0295
	02	071	02-071		

ID	reactant1	reactant2	synthID	pMIC pred-combi	MIC pred-combi
5				1.2725	0.534
	02	108	02-109		
6				1.1009	0.0793
	02	113	02-113		
7				1.3287	0.0469
	02	117	02-117		
8				1.7989	0.0159
	31	033	31-033		
9				1.25242	0.0559
	31	038	31-038		
10				1.0813	0.0829
	31	059	31-059		
11				1.7783	0.0167
	31	077	31-077		
12				1.7964	0.0160
	31	079	31-079		

ID	reactant1	reactant2	synthID	$pMIC_{pred-combi}$	$MIC_{pred-combi}$
13				1.2556	0.0555
	31	120	31–120		
14				1.1392	0.0726
	34	071	31–071		
15				1.0436	0.0904
	35	071	35–071		

four compounds from the combinatorial set (02–071, 31–033, 34–071 and 35–071) are typical *X*-outliers. Except these compounds, the other 11 combinatorially generated compounds are in the boundaries of AD with activity values in the range ($1.0133 < pMIC_{pred-combi} < 1.7964$) and the corresponding MIC values in the range ($0.0160 < MIC_{pred-combi} < 0.0970$). According to the frequency of appearance of the substituents in this compound set, one could conclude that the building-blocks (02 and 31) attached into position 1 of the main 6-fluoroquinolone core, are of significant importance for activity (Table 4). These fragments belong to the class of linear amides and planar

aromatic amines, respectively. Incorporated into the main 6-fluoroquinolone scaffold, they apparently increase the possibility of hydrogen-bonding as well as inter-molecular π - π stacking interactions with the planar aromatic/heteroaromatic residues of the duplex mycobacterial DNA.³⁸ The building-blocks incorporated into position 7, belong to different chemical groups (planar aromatic/heteroaromatic rings, carbonyl groups, aliphatic/aromatic esters, methylsulfonyl groups). The presence of these groups on the main core increase the possibility of establishing hydrogen-bonding as well as electrostatic interactions and π - π stacking interactions with the amino acid residues of the GyrA subunit's binding pocket and the DNA aromatic/heteroaromatic parts, respectively.^{39,40} Thus we believe that some of the compounds from our generated set of 15 compounds (Table 4, 11 compounds signed with bold IDs) could be proficiently used as templates for defining new SAR rules. Moreover, the AD offers a good insight into the drug-like chemical space of substituted 6-fluoroquinolones.

4. Conclusion

A classical virtual combinatorial chemistry approach was applied for combinatorial generation of a large set of new structurally-similar 6-fluoroquinolone analogs with unknown activity values. The combinatorial explosion (virtual combinatorial library) was filtered using the well known drug-likeness filtering algorithms (Lipin-

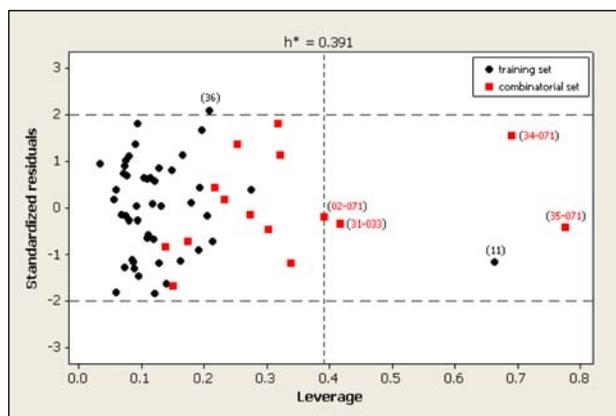


Figure 5. Graphical representation (Williams plot) of the five-parameter MLR model's applicability domain (AD) together with selected combinatorially-generated compounds.

ski, Veber, CombiVL).^{20,21} The retained compounds (304) defining the drug-like chemical sub-space (*CombiData*) were treated chemometrically employing a previously built five-parameter MLR model for activity predictions ($pMIC_{pred-combi}$). The selected most important molecular descriptors describe some of the inter-molecular interactions between the 6-fluoroquinolones and enzymatic subunit (GyrA) and mycobacterial DNA. Furthermore, these parameters/descriptors together with the predicted activity values ($pMIC_{pred-tr}$, $pMIC_{pred-combi}$) were used for defining the model applicability domain (AD). The obtained AD offers a good graphical representation, outliers identification, navigation within the drug-like chemical space, as well as a structural insight into the possible ligand-(DNA)-enzyme interactions. Our future work will be directed to use of selected compounds in three dimensional models of complexes between GyrA and the ligand.

5. Abbreviations

ATP, Adenosine triphosphate; MIC, Minimal Inhibitory Concentration; CombiChem, Combinatorial Chemistry; CombiVL, Lipinski/Veber rules combination; SAR, Structure-Activity Relationship; QSAR, Quantitative Structure-Activity Relationship; CV LOO, Cross Validation Leave-One-Out; AD, Applicability Domain;

6. Acknowledgments

Authors thank Agency of Research of R. Slovenia (ARRS) for the financial support through the Grants P1-0017 and 1000-07-310016. We are sincerely grateful to Dr. Marjana Novič for valuable insights, discussion and her continuing support of this research.

7. References

1. L. C. Du Toit, V. Pillay, M. P. Danckwerts, *Respir. Res.* **2006**, *7*, 118.
2. Y. Zhang, K. P. -Martens, S. Denkin, *Drug Discov. Today* **2006**, *11* (1/2), 21–27.
3. Y. Zhang, *Annu. Rev. Pharmacol. Toxicol.* **2005**, *45*, 529–564.
4. J. J. Champoux, *Annu. Rev. Biochem.* **2001**, *70*, 369–413.
5. H. Peng, K. J. Mariani, *J. Biol. Chem.* **1993**, *268*, 24481–24490.
6. C. Levine, H. Hiasa, K. J. Mariani, *Biochim. Biophys. Acta* **1998**, *1400*, 29–43.
7. R. J. Reece, A. Maxwell, *Crit. Rev. Biochem. Mol. Biol.* **1991**, *26*, 335–375.
8. D. C. Hooper, *Drugs* **1999**, *58* (suppl. 2), 6–10.
9. G. Anquetin, J. Greiner, P. Vierling, *Curr. Drug Targets: Infect. Disord.* **2005**, *5*, 227–245.
10. J. H. Block, J. M. Beale, eleventh ed., Whilson and Gisvold's Textbook of Organic Medicinal and Pharmaceutical Chemistry, Lippincott Williams & Wilkins, **2004**.
11. C. M. Dobson, *Nature* **2004**, *432*, 824–828.
12. R. S. Bohacek, C. McMartin, W.C. Guida, *Med. Res. Rev.* **1996**, *16*, 3–50.
13. The Practice of Medicinal Chemistry, Third Edition, Edited by C. G. Wermuth, Academic Press/Elsevier, Amsterdam, Netherlands, **2008**, Chapter 25.
14. E. Besalú, R. Ponec, J. V. de Julián-Ortiz, *Mol. Divers.* **2003**, *6*, 107–120.
15. <http://www.cambridgesoft.com/software/ChemBioOffice>
16. M. Congreve, R. Carr, C.W. Murray, H. Jhoti, *Drug Discov. Today* **2003**, *8*, 876–877.
17. Array Biopharma, Optimizer® Building Blocks, <http://www.arraybiopharma.com/OptimerBuildingBlocks/Default.asp>
18. T. Schwalbe, D. Kadzimirisz, G. Jas, *QSAR Comb. Sci.* **2000**, *24*, 6, 758–768.
19. T. Wieland, *J. Math. Chem.* **1997**, *21*, 141–157.
20. C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Deliver. Rev.* **1997**, *23*, 3–25.
21. D. F. Veber, S. R. Johnson, H.-Y. Cheng, B. R. Smith, K. W. Ward, K. D. Kopple, *J. Med. Chem.* **2002**, *45*, 2615–2623.
22. J. Jaén-Oltra, M. T. Salabert-Salvador, F. J. García-March, *J. Med. Chem.* **2000**, *43*, 1143–1148.
23. P. J. Hajduk, M. Bures, J. Praestgaard, S. W. Fesik, *J. Med. Chem.* **2000**, *43*, 3443–3447.
24. J. Jaworska, N. Nikolova-Jeliazkova, T. Aldenberg, *ATLA-Altern. Lab. Anim.* **2005**, *33*, 445–459.
25. J. Jaworska, M. Comber, C. Van Leewen, C. Auer, *Environ. Health Persp.* **2003**, *111*, 1358–1360.
26. H. Liu, E. Papa, P. Gramatica, *Chem. Res. Toxicol.* **2006**, *19*, 1540–1548.
27. T. I. Netzeva et. al., *ATLA-Altern. Lab. Anim.* **2005**, *33*, 155–173.
28. L. Eriksson, J. Jaworska, A. P. Worth, M.T.D. Cronin, R. M. McDowell, P. Gramatica, *Environ. Health Persp.* **2003**, *111*, 1361–1375.
29. N. Minovski, M. Vračko, T. Šolmajer, *Mol. Divers.* **2010**, *14*(2), doi: 10.1007/s11030-010-9238-5.
30. M. Oblak, M. Randić, T. Šolmajer, *J. Chem. Inf. Comp. Sci.* **2000**, *40*, 994–1001.
31. D. Ciubotariu, V. Gogonea, M. Medeleanu, Van Der Waals Molecular Descriptors. Minimal Steric Difference. In: M. V. Diudea (Ed.), QSPR/QSAR Studies by Molecular Descriptors. Nova Science Publishers, Inc., Huntington, New York, **2001**, 281–361.
32. T. I. Oprea, *Curr. Opin. Chem. Biol.* **2002**, *6*(3), 384–389.
33. T. I. Oprea, J. Gottfries, V. Sherbukhin, P. Svensson, T. C. Kühler, *J. Mol. Graphics* **2000**, *18*, 512–524.
34. P. Seneci, S. Miertus, *Mol. Divers.* **2000**, *5*, 75–89.
35. T. Langer, G. Wolber, *Pure Appl. Chem.* **2004**, *76*(5), 991–996.
36. I. Huc, J.-M. Lehn, *Proc. Natl. Acad. Sci.* **1997**, *94*, 2106–2110.

37. T. I. Oprea, *J. Comput. Aid. Mol. Des.* **2000**, *14*, 251–264.
38. M. M. Hann, A. R. Leach, G. Harper, *J. Chem. Inf. Comp. Sci.* **2001**, *41*, 856–864.
39. J. G. Hedde, F. M. Barnard, L. M. Wentzell, A. Maxwell, *Nucleos. Nucleot. Nucl.* **2000**, *19(8)*, 1249–1264.
40. J. T. Smith, *Eur. J. Clin. Microbiol.* **1984**, *3*, 347–350.

Povzetek

Uporabili smo klasični pristop kombinatorične kemije (*CombiChem*), s katerim smo generirali 5590 novih, strukturno-podobnih 6-fluorokinolonskih analogov z virtualno sintezo z izbranimi primarnimi (43) in sekundarnimi (130) aaminski substituenti. Tako dobljeno virtualno kombinatorično knjižnico smo filtrirali z uporabo našega niza filtrov »podobnosti-z-učinkovinami« (drug-likeness) z vgrajenimi Booleovimi algebrskimi operatorji (True/False) za redukcijo/selekcijo niza spojin. Preostali (304) fluorokinolonski analogi z opcijo True definirajo »učinkovinam podobni« kemijski prostor (*CombiData*). Kvantitativne relacije med strukturo in učinkovitostjo (QSAR) na teh virtualnih 304 6-fluorokinoloni z neznanimi vrednostmi za biološko aktivnost smo izračunali z vnaprej razvitim pet-parametrskim modelom na osnovi multiple linearne regresije (MLR) na nizu spojin z eksperimentalno določenimi aktivnostmi ($R_{tr} = 0.8417$, $R_{tr-cv} = 0.7884$). Dobljene vrednosti za aktivnosti neznanih spojin skupaj z rezultati iz modela smo uporabili za definicijo domene aplikabilnosti (AD). Tako dobljena domena aplikabilnosti nudi dobro grafično reprezentacijo in vpogled v relacije struktura-aktivnost za ta niz virtualnih molekul, kar je možno uporabiti za načrtovanje novih 6-fluorokinolonov s potencialno boljšo biološko aktivnostjo.