UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká fakulta

HABILITAČNÍ PRÁCE

New methodologies of Machine-Learning modeling of complex chemical systems: mixtures, reactions and ligand-protein complexes

Pavlo Polishchuk, Ph.D., M.Sc.

Katedra fyzikální chemie

Olomouc 2024

Table of content

Declaration of originality:

I hereby declare that I am the sole author of the submitted thesis, which, to the best of my knowledge and belief, contains no material previously published by another person, except where due acknowledgement has been made. The author grants permission to Palacký University in Olomouc to store and display this thesis and its electronic version in the university library and on the official website.

In Olomouc 30.04.2024 ……Pavlo Polishchuk

Acknowledgment

I would like to thank my teachers, who helped me a lot in my journey in science. These are my university teacher in organic chemistry Prof. Valery Anikin; the supervisor of my PhD study Prof. Victor Kuz'min, who attracted me to the field of chemoinfromatics and inspired me a lot; Prof. Alexandre Varnek who widely opened the field of chemoinformatics for me. I'm also very thankful to my colleagues and collaborators with whom we worked on many exciting projects: Dr. Andrey Krys'ko, Dr. Anatoly Artemenko, Prof. Eugene Muratov, Dr. Olena Mokshyna, Dr. Marian Hajduch, Dr. Petr Dzubak and many others. All of them greatly contributed to the final success of our projects and many of those results will be presented below in this thesis. Separately I would like to thank my friend and a colleague Dr. Timur Madzhidov with whom we spent a lot of time discussing different scientific topics and problems, generating ideas, writing papers, etc. I'm also very thankful to my students. While I taught them, they also taught me and continue teaching every day.

Last, but not least, I express sincerest gratitude to my parents, Henadii and Inna, and friends for their invaluable support.

Introduction

My main research activities were development of new chemoinformatic approaches and their application as well as state-of-the-art approaches in the fields of drug discovery, toxicology, modeling of physicochemical properties and reactions. The developed approaches and tools cover a wide range of topics (Figure 1). All of the described tools are open-source and contribute to open-science that should support further researches and progress in corresponding areas.

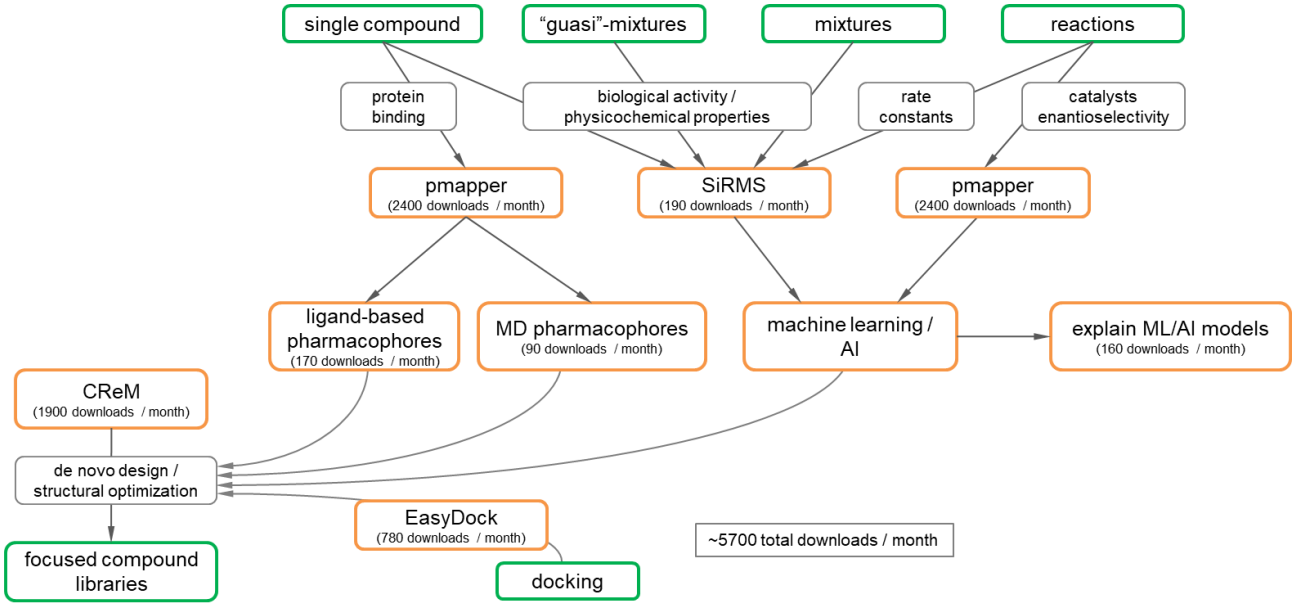


Figure 1. The main activities and developments. The number of average downloads per month for different software tools was provided based on statistics from pypistat.org taken for the period of six month from 2024-02-01 to 2024-07-30.

Quantitative structure-activity relationship (QSAR) modeling has a long history. It was developed primarily for prediction of properties of individual compounds that still remains the major direction of QSAR research. However, there are many other chemical systems which may get benefits from prediction of their properties. In the **Chapter 1** we describe the development of the molecular representation which is applicable to not only single compounds, but also to mixtures and chemical reactions. We also suggested new cross-validation protocols which give more reliable estimates of the predictive ability of models for different scenarios.

QSAR models evolved over time from simple linear models to complex non-linear ones. The latter lost one of advantages of simple models – model interpretability – and often are considered as "black boxes". In the **Chapter 2** we describe the development of new interpretation approaches. The major one is the universal interpretation approach which can retrieve contributions of fragments from any model regardless used machine learning method or descriptors. This became even more

important recently with introduction of complex neural networks to predict compound properties. We successfully demonstrated the universal nature of the approach on graph convolution neural networks. Introduction of this approach and similar ones changed the interpretation paradigm from "model ⟶ descriptor ⟶ structure" to "model ⟶ structure". We also suggested and developed the first benchmarks to evaluate interpretation approaches and model interpretability.

Pharmacophore modeling is one of the major tools in drug design. However, the majority of available tools are commercial and there only few free ones. **Chapter 3** describes implementation of 3D stereosensitive pharmacophore multiples which found them useful in ligand-based pharmacophore modeling, pharmacophore modeling based on molecular dynamics simulations and machine learning. All the developed tools are open-source and available for the scientific community.

In **Chapter 4** we tackled the problem of prediction of compound properties using 3D representation and machine learning. Currently used approaches had multiple limitations. The major ones are inability or high complexity to rationally select relevant conformers for modeling or simplistic schemes to aggregate multiple conformers into a single vector of descriptors. All these result in poor predictive performance of obtained models. We revisited the multi-instance learning (MIL) approach which can naturally treated situation where a molecule is represented by multiple conformers. We implemented several the most commonly used MIL approaches and demonstrated their superior predictive ability in combination with 3D pharmacophore multiplets relatively to 2D models in tasks of biological activity prediction. We also showed applicability of MIL approaches to prediction of enantioselectivity of catalysts.

De novo design remains one of the attracting research areas because it allows to find promising compounds without full enumeration of chemical space. The major challenge for de novo approaches is synthetic accessibility of generated molecules. In **Chapter 5** we describe a new fragment-based approach named Chemically Reasonable Mutations (CReM) which provides a clear control over synthetic complexity of generated molecules and demonstrate very competitive outputs in comparison with state-of-the-art approaches including generative neural network models.

Chapter 6 contains a few examples of computer-aided development of biologically active compounds. In particular, two projects were described in details: i) development of antagonists of the open and the closed form of integrin $\alpha_{IIb}\beta_3$, which could be promising anti-platelet agents, and ii) anti-leishmanial compounds.

Chapter 1. Machine learning modeling of properties single compounds, mixtures and reactions

The main purpose of machine learning (ML) in chemistry is to establish quantitative-structure-activity(property) relationships (QSAR/QSPR) to predict properties of new chemical entities. From the beginning the major modeling objects were single compounds and their physicochemical and biological properties. Molecules are encoded as sets of features representing their structures and machine learning algorithms establish a correlation between these features and observed property values. The established models are used to predict properties of new compounds. This helps to focus research efforts on the most promising compounds and greatly reduces costs and time required for development. Successes in prediction of properties of single compounds[1, 2] motivated researchers to extend the established approaches on other chemical systems. The major challenge was to develop appropriate representations to encode such system. Whereas there were multiple approaches to encode single molecules they could not be simply applied to represent, for example, compound mixtures or chemical reactions. In particular, for modeling of mixture properties there were suggested descriptors based on partition coefficient for mixtures[3, 4], integral additive[5, 6] and non-additive descriptors[7-9]. These approaches were frequently limited in the composition of represented mixtures, e.g. only binary mixtures or only 1:1 mixtures. They could be poorly suitable for modeling of non-additive mixture properties or they may lack of encoding of local chemical environment which may be more relevant to model a studying property.

For reaction modeling there are two groups of approaches. The former approaches use explicitly labeled reaction centers identified manually or by atom-atom mapping. In particular, they require a construction of a special graph - a condensed graph of reaction - which was used to calculate conventional fragment descriptors[10, 11]. Other group of approaches implicitly encodes information about reaction center. These approaches calculate difference between descriptors of products and reactants[12] or combine descriptors of substrates[13]. All these approaches require perfectly balanced reactions for modeling otherwise atom-atom mapping may result in erroneous outputs or feature vector may contain chemically meaningless terms. Since most of raw reaction data in the widely used databases like CAS REACT or Reaxys are not balanced, the data curation step is needed before using these modeling methods.

Here, we describe the simplex representation of molecular structure approach which was extended to enable encoding of chemical reactions and compound mixtures of arbitrary compositions. Moreover, within the same framework we were able to improve prediction of macroscopic properties of single compounds by applying a “quasi”-mixture approach which explicitly considers intermolecular interaction. The other major contribution was the development of new validation strategies to more rigorously evaluate the predictive ability of mixture and reaction models that was never done before.

1.1. Simplex representation of molecular structure (SiRMS)

Simplexes are tetraatomic fragments of fixed composition, structure, chirality and symmetry (Figure 2). The counts of identical simplexes in a molecule are simplex descriptors which can be used to find relationships between structure of studied molecules and their activities/properties. The important feature of simplexes is labeling of atoms not only by elements but also by atomic properties, such as lipophilicity, partial charge, H-bonding, etc (Figure 2). The types of used atomic parameters can be selected taking into account a studied property/activity. Property values represented by a real number, e.g. atomic charge, are split on labeled bins and atoms received a label of a corresponding bin, see an example of calculation simplexes based on partial charges in Figure 2. For those properties which are naturally encoded by discrete labels, like H-bonding, we assign corresponding labels directly, e.g. H-bond donor, H-bond acceptor, non-H-bond donor or acceptor. This results in a more comprehensive representation of a molecular object and incorporate physicochemical properties which may be important for a studied property/activity. Another unique feature of SiMRS is encoding of a molecule not only with fully connected fragments, but also by fragments were some atoms are disconnected. This gives the ability to explicitly capture contributions of distant substructures co-occurred in a molecule. Simplexes can also represent molecules at different levels of complexity. At 1D levels simplexes are combinations of atoms regardless their connectivity, at 2D level they take into account connectivity, at 3D they encode spatial arrangement of atoms and can discriminate chiral compounds which are encoded by different sets of chiral simplexes, at 4D level simplex descriptors are calculated by weighted sum of 3D simplexes of individual conformers, where weights represent Boltzmann distribution of conformers, that allows to encode ensembles of conformers. More detailed description of SiRMS approach can be found in references[14, 15].

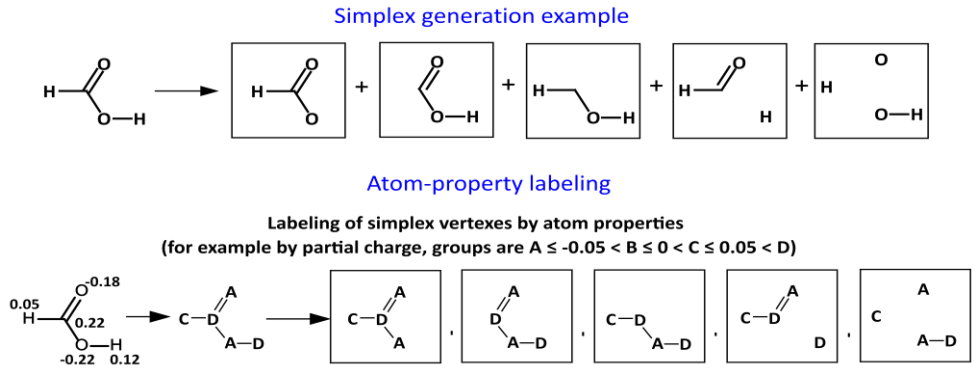


Figure 2. Example of generation of 2D simplexes for formic acid.

1.2. Modeling of non-additive properties of mixtures

To represent a mixture of compounds we consider it as a single molecular graph comprising molecular graphs of individual components. We enumerated all simplexes. Fully connected simplexes always encode a single component, whereas disconnected simplexes can represent an individual component of a mixture as well as a mixture itself (Figure 3). Here, we used the feature of SiRMS to encode disconnected fragments. Since the same simplex may occur within individual components and a mixture we label simplexes of a mixture with a special mark to distinguish them from simplexes of individual components. This representation scheme explicitly encodes possible intermolecular interactions between components that should better describes physical processes occurred in mixtures.

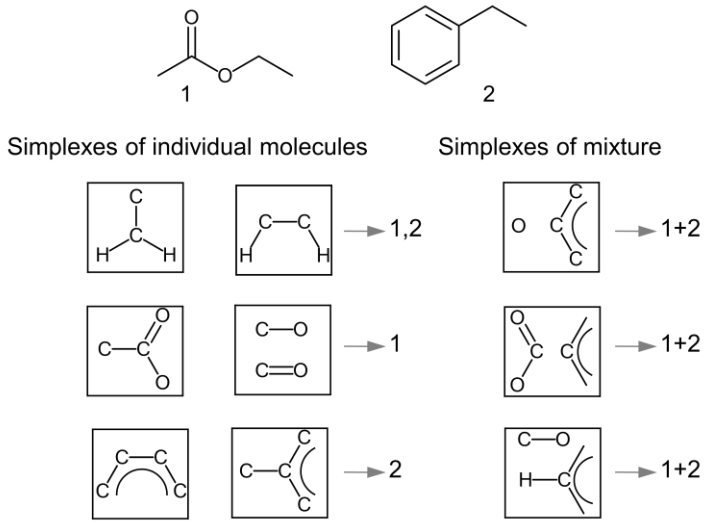


Figure 3. Scheme of calculation of simplexes for a binary mixture.

Descriptors of individual components are weighted according to their molar ratio and summarized. Mixture descriptors are multiplied on the double minimal weight according to eq 1.

$$D = \begin{cases} x_1 D_1 + x_2 D_2 \\ 2 x_1 D_{1+2} \end{cases} \quad (1)$$

where D is the descriptor value, x_1 and x_2 are molar fractions of components 1 and 2 ($x_1<x_2$ and $x_1+x_2=1$), D_1, D_2, and D_{1+2} are descriptor values for individual components 1 and 2, and for their mixture, respectively.

We suggested more rigorous validation strategies to evaluate predictive performance of mixture models to better represent different scenarios of their usage (Figure 4). These strategies are applicable if a data set is comprised from pure compounds and their mixtures where every mixture is represented by multiple data points at different concentration/molar ratio.

“Points-out”. All pure compounds and a half of randomly chosen mixture data point are going to a training set. This is the loosest strategy which evaluates the ability of a model to predict properties for new concentrations/molar ratios for a mixture for which some data points are already available.

“Mixtures-out”. In each cross-validation fold all pure compounds are always kept in a training set and whole mixtures (all data points) are randomly selected to the corresponding test set. This strategy estimates the ability of a model to predict a property for a new mixture of pure compounds.

“Compounds-out”. Within this strategy a pure compound and all its mixtures are simultaneously placed in a test set for every fold. In this case folds are created not randomly but in a supervised manner to create balanced folds if possible, because compounds may appear in different numbers of mixtures. This strategy is the strictest one and estimates the ability of a model to predict properties of mixtures comprising new compounds not available in a training set.

This differentiation of validation strategies is important because provides more relevant and reasonable estimate of model predictivity in comparison with commonly used random splitting and simulates different scenarios of model usage. The more detailed description of the SiRMS mixture approach and the suggested cross-validation protocols can be found in the reference[16].

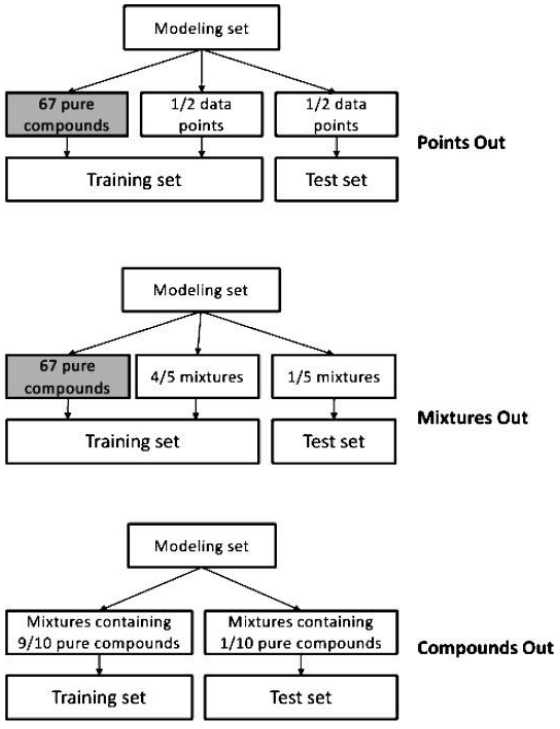


Figure 4. Strategies for validation of mixture models.

The developed workflow was applied for modeling of bubble point temperatures of binary mixtures of liquids (Figure 5). Theoretical assessment of these data could significantly reduce the costs of selection of proper agents for separation processes in industry. The dataset was compiled from Korean Data Base (KDB)[17]. It consists of 67 pure liquids and 167 of their mixtures. Each

mixture has been represented by several (7–57) points, thus, 167 mixtures have been described by 3185 data points. The matrix of mixtures was very sparse and consisted of only 167 out of possible 2211 combinations.

The models built on the entire modeling set were validated in an external test set comprising 94 new mixtures involving 66 compounds. Only 27 out 94 mixtures (632 data points) contained no new pure compounds and 67 mixtures (1386 points) contained at least one new compound. Thus, 32 external compounds were common to the modeling set, whereas other 34 were new. Four mixtures had no common compounds with the modeling set.

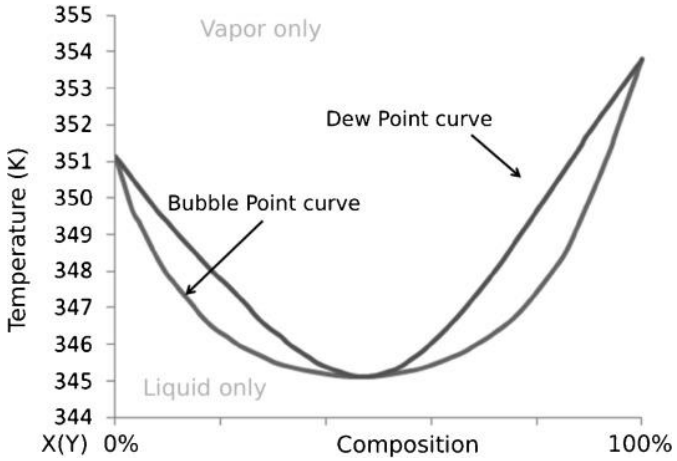


Figure 5. Vapor-liquid equilibrium curve.

Models were built using Random Forest algorithm[18]. Cross-validation was repeated several times for each validation strategy (Figure 4) using different fold compositions to get more robust estimates of model accuracy. Within the “points-out” strategy we estimated the ability of the model to reconstruct a curve from a subset of available data points. In the “mixture-out” strategy we estimated the ability to reconstruct curves for mixtures missing in the initial mixture matrix while having data points for individual components. In “compounds-out” we estimated the ability of the model to predict mixtures for unseen compound out of the training set. As it was expected the “points-out” strategy resulted in the most accurate predictions with the root mean square error 2.3K that suggests that we can reliably reconstruct partially missing bubble point curves. The “mixtures-out” strategy demonstrated reasonably high performance for both cross-validation and external test set, 5.7K and 6.6K, correspondingly. Thus, we may conclude that the model can be used to fill gaps in the existing matrix of binary mixture of liquids. The “compound-out” strategy demonstrated the lowest accuracy and had large difference between cross-validation and external test set. This indicates that reliability of prediction of bubble point temperatures for new components is relatively poor.

Table 1. Statistical characteristics of models predicting bubble point of binary mixtures built by different validation strategies.

Validation set	Parameter	point-out	mixtures-out	compounds-out
cross-validation	R^2	0.98	0.90	0.79
	RMSE, K	2.3	5.7	10.3
external test set	R^2		0.85	0.39
	RMSE, K		6.6	18.5

We compared performance of our QSPR models with COSMO-RS approach which is based on dielectric continuum models and statistical thermodynamic. COSMO-RS could also reproduce curves with high accuracy and showed performance on the "mixture-out" set of 27 test set mixtures (R^2=0.84, RMSE = 6.6K) comparable to that showed by our QSPR model. On the "compound-out" test set represented by 67 mixtures COSMO-RS showed lower accuracy (R^2=0.78, RMSE = 13.0K), which was still better than performance of our model.

1.3. Modeling of rate constants of chemical reactions

Chemical reactions are transformations of one or more reactants into one or more products. Reactants and products can be considered as individual mixtures and, thus, they can be encoded using approaches applicable to mixtures. This may solve the problem with imperfectly balanced reactions which is particularly important for approaches relied on atom-atom mapping, which are the most widely used now.

To adopt the approach described in the previous section to represent reactions as mixtures we made following extensions:

1) the number of atoms in a fragment (simplex) can be variable, usually from 2 to 6;
2) to avoid combinatorial explosion we enumerate only fully connected fragments and fragments having only two disconnected components;
3) we improved the representation approach to be applicable to mixtures with arbitrary number of components and molar ratios.

The general workflow of generation of mixture descriptors for reaction consists of three steps (Figure 6):

I. We enumerate all subgraphs consisting of 2 to 6 atoms. For the mixture of three components A, B and C, we generate fragments of individual species including atoms of only A and B, as well as mixture fragments including atoms of two (AB, BC, AC) or three (ABC) components. For components containing less than 2 atoms (e.g., component C), individual descriptors are not generated. In this way additionally to individual components we encode potential bi- and tri-

molecular interactions between components in a mixture of reactants or products. Encoding of a higher order interactions was not considered because their probability is extremely low.

Each type of fragments is considered as an individual descriptor and its count weighted by the occurrence of a corresponding component is the descriptor value.

II. The feature vectors of individual components are summed up and result in a final feature vector $D_S = A + B + C$. Similarly, summation of the vectors of mixture simplexes AB, BC, AC and ABC results in D_M vector representing intermolecular interactions.

III. Concatenation of D_S and D_M vectors results in the feature vector of the whole mixture (SiRMS−mix).

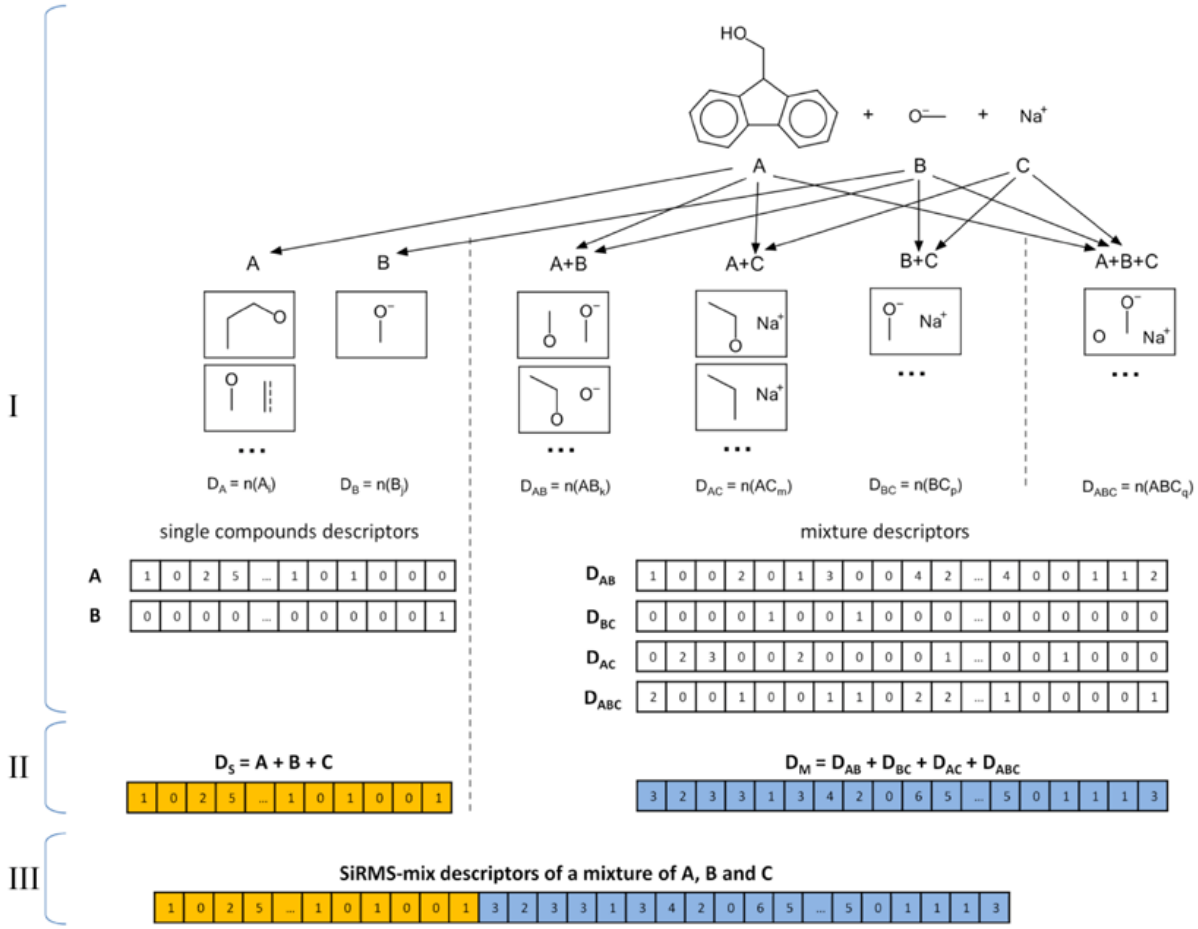


Figure 6. Generation of descriptors for a mixture of three components.

Since a chemical reaction can be represented as an ensemble of two mixtures: a mixture of starting materials (reactants) and a mixture of products, the reaction feature vector can be computed as their combination. Two different ways of combining mixture feature vectors into reaction feature vector have been investigated: (i) their concatenation and (ii) by calculation of the difference between product and reactant mixture descriptors (Figure 7).

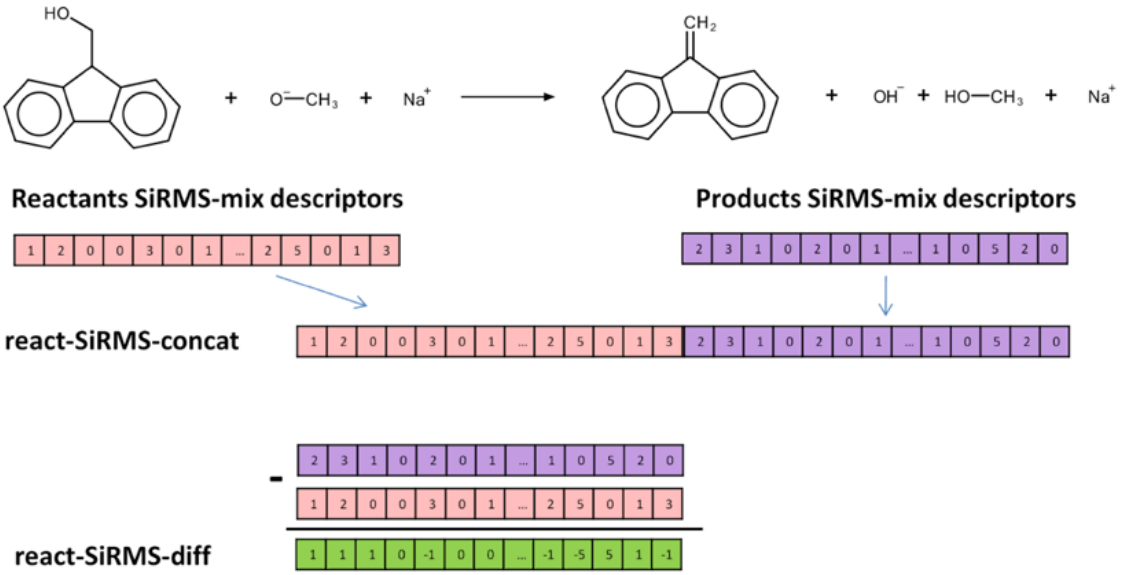


Figure 7. Reaction descriptor vectors based on the concatenated product and reactant mixture descriptors (*react-SiRMS-concat*) and on their difference (r*eact-SiRMS-diff).*

The suggested approach was applied to model rate constants of 313 E2 reactions carried out at different temperatures in pure solvents which were collected from literature[19]. An E2 reaction proceeds in a single step with a single transition state. It results in a formation of a π-bond due to synchronous trans-elimination of a leaving group (L) in the presence of a base (B-) needed to tie in the hydrogen atom (Figure 8). The dataset involves 90 distinct substrates and 60 distinct products, the most representative of them are listed in Figure 9. Among the most representative leaving groups one can mention bromide and chloride anions occurred in 101 and 93 reactions, respectively, as well as *p*-tosylate and trimethylamine which occurred in 35 reactions each. The other seven leaving groups are occurred in very few reactions. Overall, 23 bases were detected, the most representative of them were methoxide occurred in 59 reactions, ethoxide (38 reactions), tert-bytoxide (30), thiophenyl (30), triethylamine (24), bromide (20), chloride (14) and hydroxide (14) ions and piperidine (10).

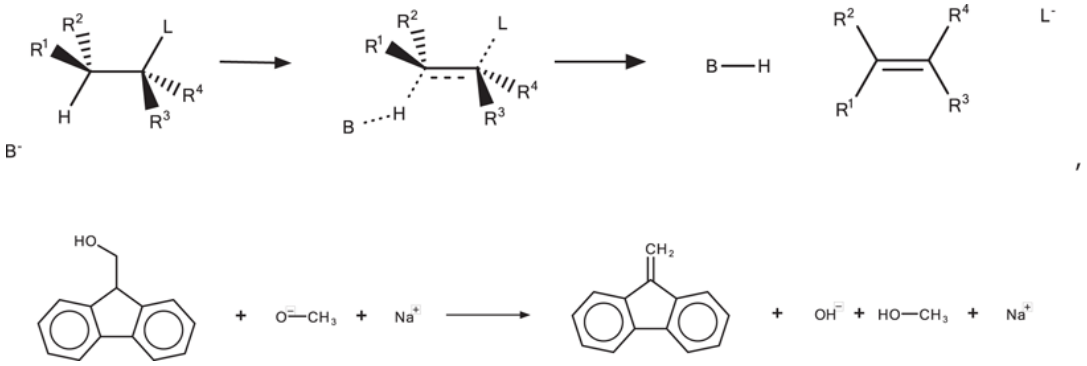


Figure 8. A bimolecular elimination reaction. (*top*) Schematic representation of the E2 reaction mechanism, where B^- is a base and L is a leaving group. (*bottom*) An example of an E2 transformation of (9*H*-fluoren-9-yl)methanol into 9-methylene-9*H*-fluoren, where CH_3O^- (from sodium methylate) is a base and hydroxide ion is a leaving group.

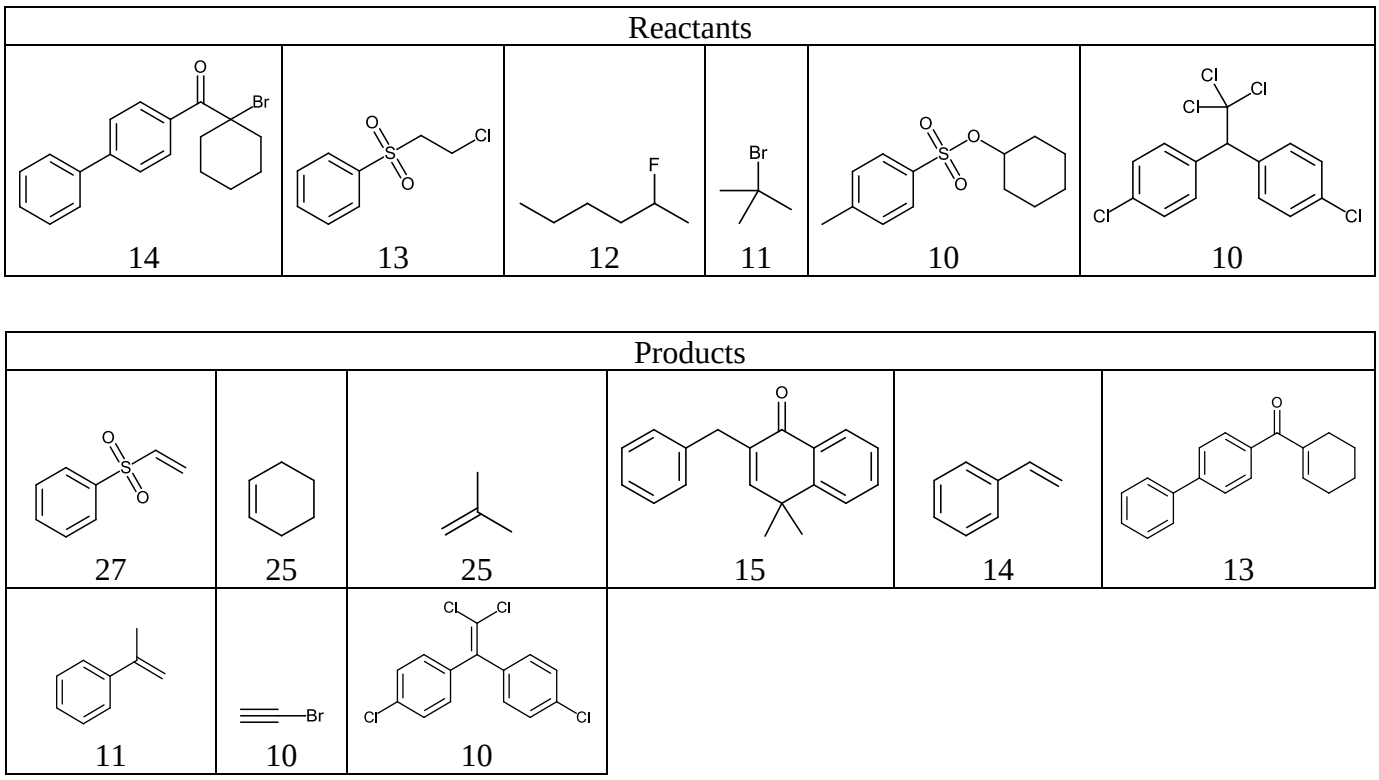


Figure 9. The most frequently occurred substrates and products in the dataset of E2 reactions.

To build models we used Random Forest approach. For encoding reactions we used simplexes included from 4 to 6 atoms. The atoms were labeled either by elements (*elm*) and partial atomic charges (*chg*) using the standard binning scheme (Figure 2).

For validation of models we introduced for the first time a more rigorous "product-out" cross-validation strategy which was inspired by the validation strategies developed for mixture model assessment described in the previous section. Within the "product-out" strategy we remove all reactions comprising a particular product to a test set for a particular fold. Since the number of reactions associated with individual products varied a lot, we implemented a special Monte Carlo protocol which stochastically created a given number of balanced folds. We repeated it ten times to get a more robust estimate of model predictive performance. For comparison purposes we used the conventional cross-validation strategy where reactions were split on folds randomly ("reaction-out").

In order to discard reactions dissimilar to those in the training set, the "Fragment Control" applicability domain approach has been used[20]. The "Fragment Control" discards any test set reaction containing fragments which don't occur in the training set reactions. An applicability domain was applied to the test set reactions at each fold followed by assembling the results for all folds. In such a way, statistical parameters were calculated for the entire set. Data coverage was assessed as a ratio of the number of reactions accepted by applicability domain to the total number of reactions.

For comparison we used reaction fingerprints and condensed graph of reaction approaches. A reaction fingerprint is the difference between count-based fingerprints of products and reactants. we used three types of reaction fingerprints developed by Schneider et al.[12] and implemented in RDKit[21]: (i) atom pairs representing two particular atoms with the specified number of non-hydrogen neighbor atoms separated by up to three bonds[22], (ii) Morgan fingerprints identical to extended-connectivity fingerprints with radius 2[23] and (iii) topological torsions representing four consecutively linked non-hydrogen atoms with the specified number of π-electrons and the number of non-hydrogen neighbor atoms[24]. The other comparable approach is a condensed graph of reaction (CGR). CGR is a special molecular graph merging product and reactants in a single entity.[25] Formed and cleaved bonds (dynamic bonds) are labeled with special marks. From CGR conventional fragment descriptors can be computed. We calculated augmented atoms and sequences with length varying from 1 to 8 atoms using ISIDA Fragmenter tool[10].

We found that *react-SiRMS-diff* descriptors performed a little bit better than corresponding *react-SiRMS-concat* but in the majority of cases the difference was negligible (Table 2). Combination of simplexes labeled by charge and element resulted in comparable performance to individual labeling schemes. As expected the "product-out" cross-validation strategy demonstrated lower performance than the "reaction-out" strategy. However, taking into account applicability domain of models the performance of the "product-out" validation strategy substantially increased at cost of lower coverage. For example, for *react-SiRMS-diff* R^2_{AD} (*chg+elm*) increased from 0.42 to 0.74 at the cost of lowering the coverage from 100% to 15%. Such a big lost in the data coverage can be explained by high structural diversity and relatively small size of the data set due to which the test set objects often contain the fragments absent in the training set. This suggests that the models cannot extrapolate too far reliably, however within the chosen applicability domain model performance is reasonably high.

Table 2. Cross-validation performance of models predicting rate constants of E2 reactions. (AD means performance taking into account applicability domain)

#	Model	SiRMS atom labeling	Validation strategy	R^2	RMSE	R^2_{AD}	$RMSE_{AD}$	Coverage
4)	react-SiRMS-diff	chg	reaction-out	0.62	0.87	0.68	0.83	0.80
5)		chg+elm		0.68	0.81	0.74	0.74	0.76
6)		elm		0.69	0.78	0.74	0.74	0.85
7)		chg	product-out	0.36	1.14	0.64	0.86	0.22
8)		chg+elm		0.42	1.08	0.74	0.75	0.15
9)		elm		0.47	1.03	0.64	0.90	0.38
10)	react-SiRMS-concat	chg	reaction-out	0.63	0.86	0.67	0.84	0.79
11)		chg+elm		0.67	0.81	0.73	0.76	0.75
12)		elm		0.69	0.79	0.73	0.75	0.83
13)		chg	product-out	0.35	1.15	0.62	0.89	0.21
14)		chg+elm		0.39	1.11	0.71	0.80	0.14
15)		elm		0.43	1.07	0.59	0.90	0.37
16)	Atom pairs FP		reaction-out	0.61	0.89	0.62	0.87	0.97
17)			product-out	0.35	1.14	0.41	1.07	0.64
18)	Morgan FP		reaction-out	0.67	0.82	0.70	0.70	0.92
19)			product-out	0.40	1.10	0.67	0.81	0.33
20)	Topological torsions FP		reaction-out	0.60	0.90	0.62	0.88	0.94
21)			product-out	0.34	1.15	0.51	1.03	0.45
22)	ISIDA/CGR		reaction-out	0.69	0.79	0.74	0.74	0.88
23)			product-out	0.41	1.09	0.61	0.90	0.16

Reference models demonstrated similar trends in performance change depending on the validation protocol and taking into account applicability domain (Table 2). The best reference models based on ISIDA/CGR (#20, Table 2) and Morgan fingerprints (#16, Table 2) demonstrated poorer performance than the best SiRMS model (#5, Table 2). The corresponding R^2_{AD} and coverage for "product-out" cross-validation scheme were 0.61/16% (p-value = 0.0002) and 0.67/33% (p-value = 0.0080) versus 0.74/15%. This confirms that the suggested mixture-based encoding of reaction is competitive to state-of-the art methods and can be used in future applications. More details about the mixture-based reaction representation and modeling results can be found in the paper[26].

1.4. "Quasi"-mixture modeling for prediction of macroscopic properties of single compounds

Macroscopic properties of pure compounds, e.g. critical temperature, volume and pressure, are strictly determined by intermolecular interactions. Commonly used QSPR approaches use either 2D topological descriptors[27] or 3D descriptors computed using quantum chemical calculations for a single molecule[28] or their combination[29, 30]. Those descriptors calculated from individual molecules implicitly encode possible intermolecular interactions. We suggested the approach where possible intermolecular interactions were explicitly encoded by 2D descriptors that should improve the predictive ability of models[31].

A pure compound can be represented as a “quasi”-mixture of two identical components in ratio 1:1. Further, we can apply to this “quasi”-mixture the same representation approach as for mixtures with different components (Figure 10). An individual molecule is encoded by connected and disconnected fragments (simplexes) whereas a mixture is represented by disconnected simplexes only. This explicitly supply to a model possible intermolecular interactions of a pure compound in a condensed phase.

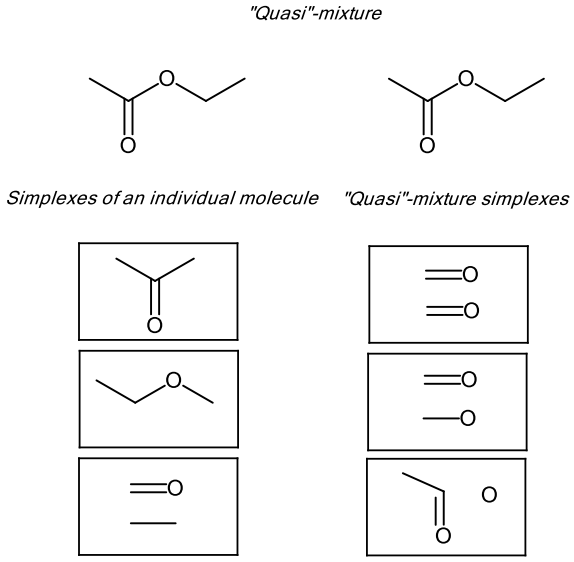


Figure 10. Example of “quasi”-mixture representation.

To evaluate the suggested approach we collected three data sets on critical temperature, volume and pressure. The experimental data were taken from the comprehensive handbook[32]. Wrong or incomplete data were curated using NIST Webbook database. Considered compounds belonged to various classes, such as saturated and unsaturated hydrocarbons, aromatic hydrocarbons and their derivatives, heterocyclic compounds, alcohols, ethers, esters, various halogenated compounds, etc. The experimental T_c values were available for 407 compounds, P_c – for 382 compounds, V_c – for 309 compounds. Collected data cover the large range of values (for critical temperature from about 100 K to about 900 K, critical pressure from about 10 to about 90 bar, critical volume from about 100 to about 1000 cm3/mol).

As a reference modeling approach we applied the common SiRMS approach encoding single molecules (Figure 2). Models were built using Random Forest method. To estimate the predictive ability of models we used 3×5-fold cross-validation. All models had high predictive performance (Table 3). Determination coefficients differed insignificantly for single molecule and “quasi”-mixture approaches. However, difference in RMSE was statistically significantly. “Quasi”-mixture models improved RMSE for T_c, V_c and P_c on 7%, 15% and 6%, respectively.

Table 3. Statistical performance of models with 95% confidence interval calculated for 3 × 5-fold cross-validation.

		T_c	V_c	P_c
single molecule	R^2	0.87 ± 0.04	0.96 ± 0.01	0.88 ± 0.03
	RMSE	40.1 ± 0.7 (K)	28.0 ± 0.10 (cm^3/mol)	4.45 ± 0.07 (bar)
"quasi"-mixture	R^2	0.89 ± 0.02	0.96 ± 0.01	0.90 ± 0.02
	RMSE	37.2 ± 0.7 (K)	24.0 ± 0.12 (cm^3/mol)	4.17 ± 0.05 (bar)
RMSE improvement		7%	15%	6%

1.5. Summary

We developed mixture representation within the SiRMS approach. Its main feature is explicit encoding of intermolecular interactions. This representation was successfully applied to model properties of not only conventional mixtures (bubble point curves[16]) but also chemical reactions (rate constants[26]) and individual compounds (critical properties[31]). We showed that the developed approach result in models with comparable or better performance than state-of-the-art methods that confirmed its applicability.

To improve estimation of the predictive ability of models we suggested new cross-validation protocols and demonstrated their applicability on several tasks. These protocols more rigorously evaluate model performance and better correspond to real use cases in comparison to random fold split. The protocols were further applied in other studies[33, 34] and inspired the development of analogous approaches. In particular, for reaction modeling there were suggested "transformation-out" and "solvent-out" cross-validation protocols[35] or the "everything-out" strategy for mixture modeling[36].

We made an open-source implementation of the SiRMS approach including mixture-based descriptors which is available as a Python package *sirms* - https://github.com/DrrDom/sirms.

Chapter 2. Approaches to interpretation of machine learning models

Machine learning models were developed from simple and linear models to more complex nonlinear ones able to process large heterogeneous data sets. The latter usually have the greater predictive ability but they frequently lack interpretability to understand the reasons of model decisions. This became extremely important recently with wide introduction of deep learning approaches in chemoinformatics. These complex models are usually considered as "black boxes" which are hardly or non-interpretable. Understanding the reasons of model decisions can be useful in knowledge-based validation of models and increase confidence that the model is right for right reasons. Another aspect of model interpretation is to use information about favorable and unfavorable structural motifs revealed by a model to guide next design steps in optimization of compound properties.

The commonly using interpretation approaches calculate contributions of individual descriptors. This creates one of their major limitations. They are applicable only to models built on clearly interpretable descriptors. Otherwise it is impossible to understand interpretation output. However, models trained on complex non-interpretable or hardly interpretable descriptors can demonstrate comparable or better predictive abilities. This became most pronounced recently with introduction of graph convolution neural networks and similar end-to-end modeling approaches which take a molecule as an input, create embedding of a molecule inside the network and establish a correlation between embedding and a target property. Such models could not be interpreted using conventional approaches.

Here we describe the developed approach to interpret Random Forest models, which is model-specific, and the universal interpretation approach (UIA), which can be applied to any QSAR/QSPR models regardless of machine learning methods or descriptors used. The latter changed the paradigm in interpretation of machine learning models, because we could skip the intermediate step of calculation of contribution of descriptors and directly estimate contributions of individual atoms or group of atoms. Therefore, interpretability of descriptors becomes not a necessary requirement. We demonstrated this ability of UIA in multiple studies and adopted this approach to interpretation of graph convolution models that confirmed its universal applicability.

With introduction of deep learning in QSAR/QSPR studies multiple special interpretation approaches were suggested and adopted. However, as it was shown in pilot studies not all of them are reasonable to apply to interpretation of QSAR/QSPR models. Therefore, to progress the field further there is a need to have appropriate benchmarks relevant to chemoinformatics tasks. To address this issue we suggested the first benchmark for QSAR/QSPR interpretation approaches.

2.1. Interpretation of Random Forest models

Random Forest (RF) belongs to "black box" models. This is an ensemble of decision trees built using three principles [18]:

1. Every tree is trained on a bootstrap subset of training set molecules.
2. Only a random subset of descriptors is considered to select an optimal split in each node.
3. Every tree is grown till maximum depth.

Thus, every tree is overfitted on a random subspace of training set examples and descriptors. Aggregation of predictions made by such weak learners gives a great boost to predictive performance. Predictions are made by averaging in the case of regression models and by majority voting in the case of classification ones. RF models are widely used in QSAR modeling due to their robustness and few tuning parameters to which models are sensitive: the number of trees and the number of randomly chosen variables considered at each node split (m_{try})[37]. Samples which were not used for training of an individual tree create an out-of-bag (OOB) set which is used for estimation of model robustness and its predictive ability.

Individual trees are interpretable by design (Figure 11). They are recursively constructed by applying simple rules, like "IF x < 5 THEN left node ELSE right node". In a leaf node the average property value of training set examples reached this node is assigned as a predicted value to test examples (in classification this is the most frequently occurred label). Therefore, the prediction rule is a combination of rules of individual nodes which are on the path from the root to the leaf node and it can look like "IF $S_1 \leq 3$ AND $S_2 \leq 2$ THEN $pIC_{50} = 8.1$" (Figure 11). This explains the decision of a model and can provide a deeper insight in the case of interpretable descriptors. However, in the case of an ensemble of randomized trees this interpretation approach does not work because there is no straightforward way to combine different rule sets or determine contributions of individual descriptors.

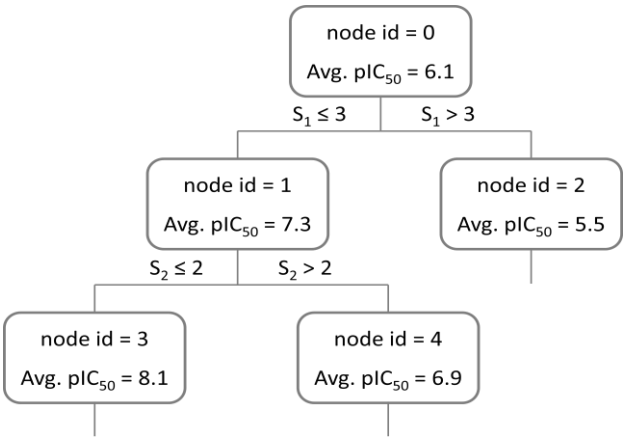


Figure 11. Example of a decision tree.

For a long time there was only one approach for interpretation of RF models suggested by Leo Breiman in its seminal work – variable importance[18]. Values of an individual variable are shuffled and predictions for the out-of-bag samples are made. Variable importance is calculated as a decrease in prediction accuracy. The bigger the drop in prediction accuracy the more important the variable. To get a more robust estimate, this procedure is applied several times to calculate average importance. The lack of the importance measure – it says nothing about the direction of variable influence, positive or negative.

We suggested an approach[38] to calculate descriptor contributions which can be interpreted in a way similar to coefficients of ordinary linear regression model and overcome limitations of variable importance suggested by Breiman. The developed approach is applicable only to regression models. The procedure of the calculation of descriptor contributions is based on two features: 1) in each node only one descriptor is used for splitting, 2) the difference between mean activity values in parent and child nodes can be considered as a predicted activity change caused by this descriptor (local increment).

In each node of a tree we calculate average activity values for training set examples reached it (Figure 12):

$$A_{mean} = \frac{1}{n}\sum_{i=1}^{n} A_i \qquad (2)$$

A_{mean} – average activity values of training set molecules in a node; A_i – activity of i-th molecule; n – the number of training set molecules in a node.

Each tree node, except the root, has an associated rule according to which compounds fall into this node. The difference between mean activity values in the child (A_{mean}^{child}) and parent (A_{mean}^{parent}) nodes represents a local increment (LS) of the contribution of the corresponding descriptor (Figure 12).

$$LS^{child} = A_{mean}^{child} - A_{mean}^{parent} \qquad (3)$$

Thus, summation of the average activity value in the root node and all local increments in nodes on the path from the root to the leaf node results in a predicted activity value of a compound reached this leaf node.

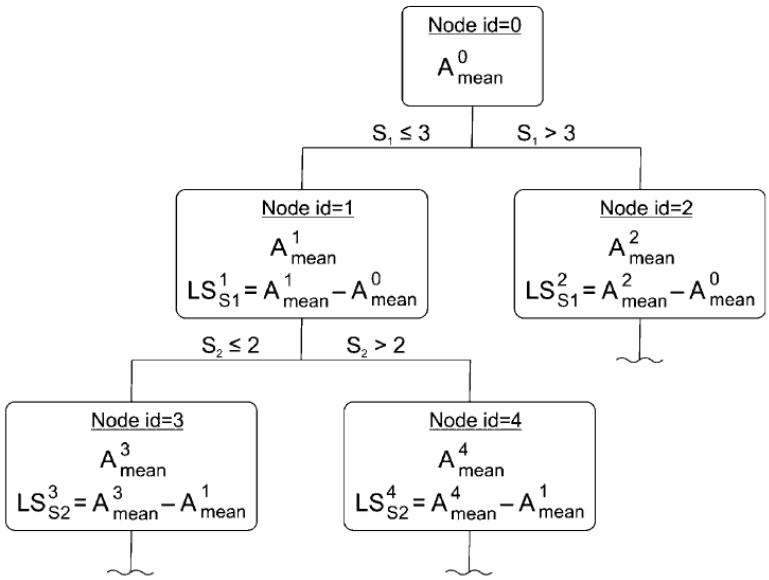


Figure 12. Example of calculation of local increments of descriptor contributions where Amean: mean activity value of training set compounds in the node; S: descriptor; LS: local increment in a compound activity caused by descriptor S.

To determine the overall contribution of an individual descriptor for a particular compound, one should summarize local increments of this descriptor included in the rules of nodes, which contain given compound in all trees. The final sum is divided by the overall number of trees in the forest.

$$C_{M,k} = \frac{1}{T}\sum_{j=1}^{m} LS_{k,j} \tag{4}$$

$C_{M,k}$ is a contribution of k-th descriptor in activity of a molecule M; $LS_{k,j}$ is a local increment of k-th descriptor associated with j-th node, which contain the molecule M; m is total number of nodes in all trees containing molecule M; T is the total number of trees in RF model.

Intercept (the absolute term, A_0) is calculated by averaging of average activity of training set molecules in root nodes ($A_{root,t}$) of all trees (T).

$$A_0 = \frac{1}{T}\sum_{t=1}^{T} A_{root,t} \tag{5}$$

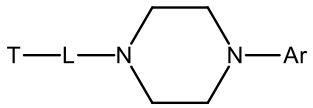
It can be easily demonstrated that the sum of the absolute term (A_0) and all local increments ($C_{M,k}$) of all descriptors (K) for a particular compound M will be equal to the activity value predicted by the forest ($A_{M,pred}$).

$$A_{M,pred} = A_0 + \sum_{k=1}^{K} C_{M,k} \tag{6}$$

Thus, descriptor contributions are additive for an individual compound, but their actual values can be different for different compounds.

The calculated descriptor contributions can be analyzed in the same way as regression coefficients of linear models. In the case of fragmental descriptors their contributions can be equally distributed among all atoms belonging to corresponding fragments and the calculated atomic contributions from different fragments can be summed up to get the final atom contribution which is easy to depict on a structure for the subsequent analysis.

We demonstrated the applicability of this interpretation approach on the task of modeling of ligand affinity for 5-HT_{1A} receptors. We collected the set of 347 ligands of 5-HT_{1A} receptor with associated pK_i values. All compounds belonged to derivatives of arylpiperazines of the general formula:



where Ar was various aryl groups; L – polymethylene linker chain with 1-6 carbon atoms; T – diverse terminal groups mainly represented by amide and imide moieties comprising a hydrophobic part. For more detailed description of the dataset one may refer to the paper [38].

For modeling we used 2D simplex descriptors which are the counts of tetraatomic fragments of fixed topology and composition [14, 15] (Section 1.1). Therefore, we could recalculate atomic contributions from calculated contributions of descriptors. 7800 descriptors were calculated in total. For the comparison purpose we built a partial least square (PLS) model which is easily interpretable through regression coefficients[39]. These regression coefficients represent contribution of individual descriptors which can be recalculated into atomic contribution by the procedure described above. We used 5-fold cross-validation to estimate predictive of models.

To build the PLS model we removed highly correlated descriptors and performed variable selection that remained 72 simplex descriptors. The final model has the reasonable predictive ability R^2_{5CV}=0.64. To build the RF model we tuned the number of trees and variables used in each split by evaluating out-of-bag predictions. The final RF model was built using m_{try} 2500 and consisted of 750 trees. It has comparable predictive ability to the PLS model, R^2_{5CV}=0.70.

First, we compared calculated descriptor contributions from the RF models with their importance calculated by the permutation procedure suggested by Breiman. We expected that descriptors having high positive or negative contributions would have high importance score. To verify this hypothesis we calculated correlation between importance scores and absolute contribution values of corresponding descriptors. They demonstrated very good concordance, $R_{Pearson}$ = 0.98 and $R_{Spearman}$ = 0.90, thus, supporting the validity of the developed interpretation approach.

Second, we recalculated atom contributions from PLS and RF models and calculated contribution of individual fragments Ar, L and T by summation of contributions of corresponding atoms. Comparison of average contributions of individual fragments demonstrated good concordance between estimates obtained from two models. The correlation was 0.88 (Figure 13). This also supports the conclusion about validity of the developed interpretation approach.

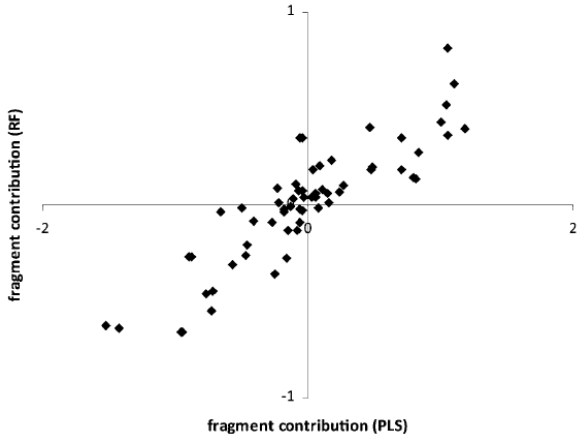
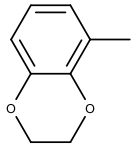
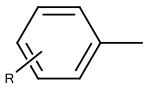
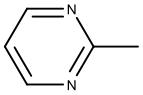
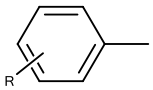


Figure 13. Calculated average contributions of molecular fragments affinity for 5-HT_{1A} receptors from PLS and RF models (only fragments which occurred 3 or more times are displayed; $R_{Pearson}$=0.88).

We also ranked fragments of each group Ar, L and T according to their calculated contributions and compared the revealed trends of structure-activity relationships with expert knowledge and available pharmacophore hypotheses. We found good correspondence between the expected relationship and the relationships retrieved from the models. For Ar fragments the *ortho*-substituted phenyl groups with p-electron donating substituents were the most favorable whereas *para*-substituted phenyl had large negative contribution (Table 4). These observations are supported by findings of other authors that substituents in *ortho*-position, which are able to form H-bonds, are favorable for activity of 5-HT_{1A} receptor ligands [40, 41] and substituents in *para*-position are unfavorable due to steric clashes [42].

Table 4. Ar fragment contribution to affinity for 5-HT_{1A} receptors calculated from PLS and RF models.

Ar		Average PLS contribution (range)	Average RF contribution (range)
		0.84 (0.62 – 1.06)	0.27 (-0.17 – 0.58)
	2-OCH_3	0.18 (-0.08 – 0.35)	0.24 (-0.28 – 0.60)
	2-Cl	0.03 (-0.13 – 0.13)	0.04 (-0.23 – 0.27)
	3-Cl	-0.04 (-0.09 – 0.01)	0.07 (-0.26 – 0.37)
	H	-0.06 (-0.21 – -0.03)	-0.02 (-0.32 – 0.29)
	3-CF_3	-0.09 (-0.31 – 0.01)	0.11 (-0.36 – 0.55)
		-0.11 (-0.42 – -0.05)	0.04 (-0.09 – 0.18)
	2-CH_3	-0.66 (-0.98 – -0.37)	-0.04 (-0.40 – 0.26)
	4-F	-0.73 (-1.03 – 0.41)	-0.55 (-0.92 – 0.39)
	4-NO_2	-0.94 (-1.24 – 0.08)	-0.66 (-1.01 – 0.31)
	4-Cl	-0.96 (-1.18 – -0.02)	-0.66 (-0.87 – 0.14)

For the linker L we observed a clear trend that 4 and more methylene groups in the chain are an optimal length resulted in high affinity for 5-HT_{1A} receptors that corresponds to experimental data [43]. For terminal groups T it was previously established that they have positive steric influence on the affinity for 5-HT_{1A} receptors [42, 44]. Our findings fully correspond to these observations. Relatively small hydrophobic groups (adamant-1-ylcarbonylamino, 1,3-dioxo-tetrahydro-pyrrolo[1,2-c]imidazol-2-yl, 1,3-dioxo-tetrahydro-imidazo[1,5-a]pyridine-2yl and phtalimidyl) have positive contributions whereas larger groups, like 3-benzhydrylidene-2,5-dioxo-pyrrolidin-1-yl and 3-fluoren-9-ylidene-2,5-dioxopyrrolidin-1-yl had negative contributions. Overall, interpretation outputs were in agreement with pharmacophore hypotheses of ligands of 5-HT_{1A} receptors [45-47]. Active molecules should have positively charged nitrogen distant from aromatic and carbonyl groups on 4.9Å and 4.3Å, correspondingly (Figure 14). This confirms that the optimal linker moiety should include at least 4 methylene groups.

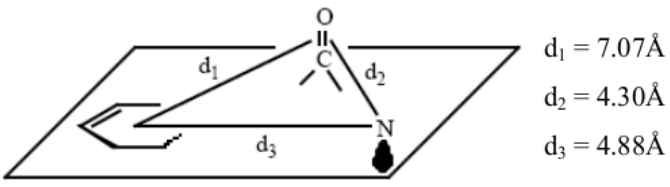


Figure 14. Tricentric pharmacophore model of 5-HT_{1A} ligands [47].

Additionally we studied interpretation of a random model. We applied Y-scrambling and built RF model. It demonstrated no predictive ability, R^2_{OOB} = -0.17. Interpretation of this model using the

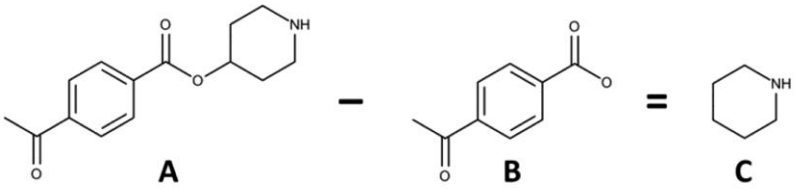
developed approach resulted in fragment contributions which did not correlate with contribution calculated from the previous model (R^2 = -0.02).

The suggested approach was implemented as a free program (http://www.qsar4u.com/pages/rf.php) and makes interpretation of RF models straightforward. The analysis of calculated contributions of descriptors is similar to the analysis of regression coefficient in linear models. In the case of using fragment descriptors this allows to retrieve contributions of substructures and to establish the most important ones that can guide further structural optimization of these compounds. Later, this approach was extended to classification models by Palczewska et al and implemented in rfFC package for R.[48]

2.2. The universal approach to structural interpretation of QSAR models

All interpretation approaches developed before used the following paradigm – calculate or retrieve contributions of descriptors from a model and then, if possible, map them back on a structure to get information about contribution of individual fragments in a molecule. We call it "model⟶descriptors⟶structure" paradigm[49]. This creates the major obstacle to perform structural interpretation – the necessity to use interpretable descriptors, mainly fragmental descriptors, whose contributions can be transferred to a structural level.

We suggested an interpretation approach, which is agnostic to machine learning methods and descriptors[50]. It designated the appearance of a new interpretation paradigm ("model⟶structure") where contributions of structural moieties are calculated directly from a model skipping the step of calculation of contributions of descriptors and their interpretation or transferring on a structural level. The idea behind is to virtually remove a fragment of interest from a molecule, predict activity/property for the new structure with a removed fragment and subtract this value from the predicted activity/property of the initial molecule. In such a way we "mask" a part of a molecule from a model and calculate the difference between predicted activity/property values (Figure 15). In such a way we can estimate contribution of arbitrary groups of atoms (even not directly connected) or even individual atoms. The only limitation of this approach is the ability of descriptors/models to encode structures consisting of disconnected parts, which can appear after removal of a linker group.



Interpretation	$Activity_{pred}(A)$	$Activity_{pred}(B)$	Contribution(C)
Structural	$f(A) = x$	$f(B) = y$	$W(C) = x - y$
Physico-chemical	$f(A_E, A_H, A_D, A_{HB}) = x$	$f(A_E, A_H, A_D, B_{HB}) = y$	$W_{HB}(C) = x - y$

Figure 15. The scheme of structural and physicochemical interpretation of QSAR models. A and B are descriptors of compounds A and B. Subscripts E, H, D, HB designate subset of descriptors encoding electrostatic, hydrophobic, dispersive and H-bonding interactions.

This approach is applicable to regression as well as classification tasks. In the case of regression models the predicted property/activity value can be directly used for calculation of a difference (contribution). Thus, calculated contributions have the same units as a modeling property. A positive contribution designates that a fragment increases the modeling property/activity and a negative contribution is observed if a fragment decreases it. In the case of classification models we used as a predicted value the probability to belong to the active class. Thus, the difference (contribution) indicates how much a particular fragment increases the probability of a molecule to become active and calculated contributions are always within the range from -1 to 1.

The suggested structural interpretation approaches was validated on multiple data sets representing regression and classification tasks. Here, we will illustrate how it works on some of them. For more detailed description and examples one may refer to our publications [50, 51].

To compare the suggested structural interpretation approach with well-established ones we built RF, PLS, SVM and GBM (Gradient Boosting Method) models for the classical data set, which was used by Free and Wilson to estimate contributions of individual fragments using linear models[52]. The data set was small and consisted of 29 compounds (Figure 16). To build models we used 2D simplex descriptors (Figure 2). Due to the small size of the data set five-fold cross-validation performance of all models was poor, R^2_{5CV} = 0.26-0.43. The consensus model calculated by averaging predictions of individual models was also poor, R^2_{5CV} = 0.38. However, this was expected and we interpreted these models. Calculated contributions of individual fragments were compared with those obtained by Free and Wilson in their seminal paper[52]. For all models, linear (PLS) and non-linear ones (RF, GBM, SVM), we observed good correspondence of calculated contributions with those ones calculated by Free and Wilson from their linear model (Figure 16). This illustrates convergence of different interpretation approaches that additionally supports validity of interpretation outputs.

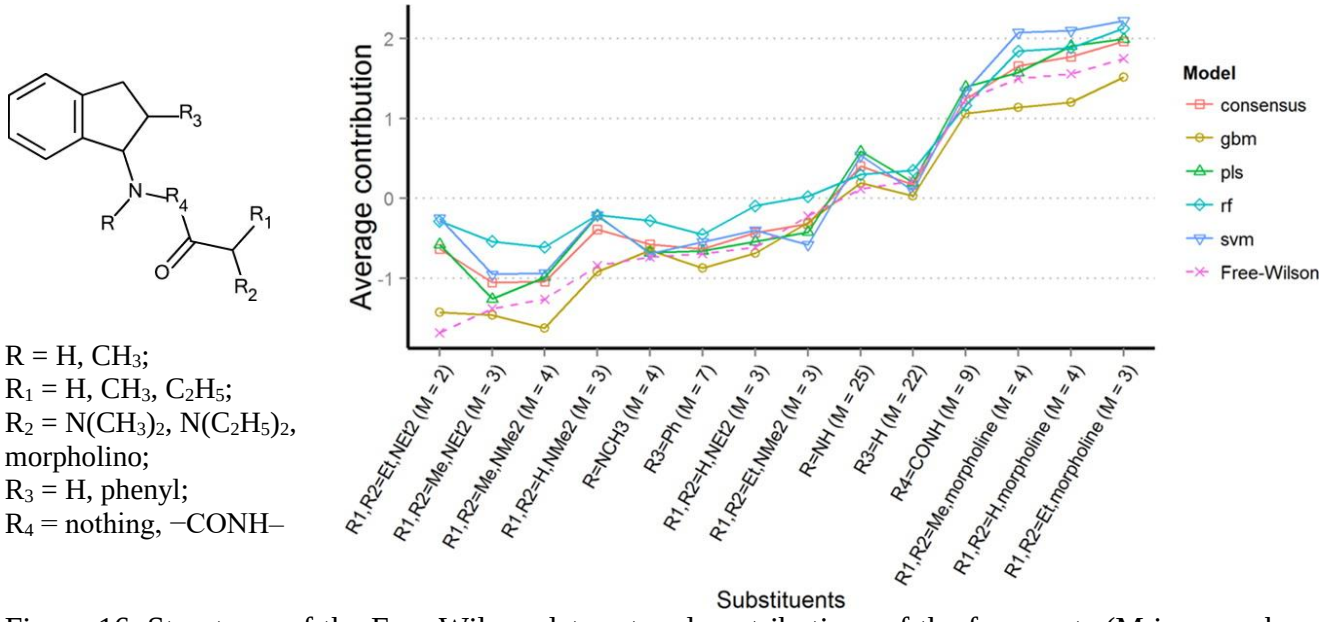


Figure 16. Structures of the Free-Wilson data set and contributions of the fragments (M is a number of molecules comprising a particular fragment).

We studied how different models and descriptors affect interpretation outputs. We built PLS, SVM and RF models for the solubility data set consisting of 1033 compounds (regression task) and SVM and RF models for the mutagenicity data set consisting of 4361 compounds (classification task) using 2D simplex and all 2D Dragon descriptors [50]. 2D Dragon descriptors included topological, constitutional, connectivity, informational, 2D autocorrelations, molecular properties and others. We chose Dragon descriptors because they were commonly used in chemoinformatics and some of them were non-interpretable making models trained on them non-interpretable as well.

Table 5. QSAR models of solubility (logS) [50].

Model	SiRMS		Dragon	
	R^2_{CV}	RMSE	R^2_{CV}	RMSE
PLS	0.84	0.82	0.91	0.60
RF	0.88	0.71	0.91	0.62
SVM	0.87	0.72	0.92	0.59

Table 6. QSAR models of mutagenicity [50].

Descriptors	Model	Balanced accuracy
SiRMS	RF	0.817
	SVM	0.800
Dragon	RF	0.816
	SVM	0.793

All models demonstrated the high predictive ability (Table 5 and Table 6). Thus, we expected that interpretation of these models would result in similar outputs. The only issue of Dragon

descriptors was inability to calculate them for disconnected structures. This made impossible to calculate contributions of linker moieties. Thus, for models trained on Dragon descriptors we estimated contributions of terminal groups only. In all cases we observed very good correspondence of calculated fragment contributions across models trained using different descriptors and machine learning methods (Figure 17 and Figure 18). The estimated contributions also corresponded to expert knowledge. Thus, differently substituted aromatic groups decreased solubility of compounds whereas ionogenic and polar groups (-$N(CH_3)_2$, CH_2OH, -SO_2NH_2, etc) improved it or affect solubility only slightly (Figure 17). For interpretation of mutagenicity models we calculated contributions of known toxicophores and detoxicophores (Figure 18). As expected, toxicophores had positive contribution values and thus increased probability of compounds to be mutagenic. In particular, high contributions of alkyl-O-N=O, which were close to 1, indicated that introduction of this group into a molecule will make it mutagenic with very high probability. Such fragments can be considered "activity triggers" or "emerging patterns", because substantially change activity of compounds. On the other hand, some fragments such as aryl-SO_2NH_2 can substantially decrease probability of a compound to be mutagenic (Figure 18).

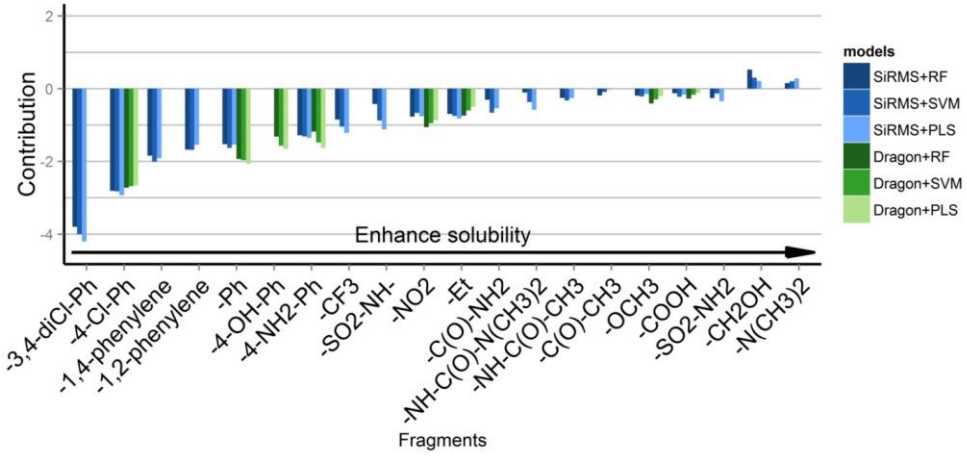


Figure 17. Average contributions of fragments to compound solubility calculated from models trained on simplex (SiRMS) and Dragon descriptors.

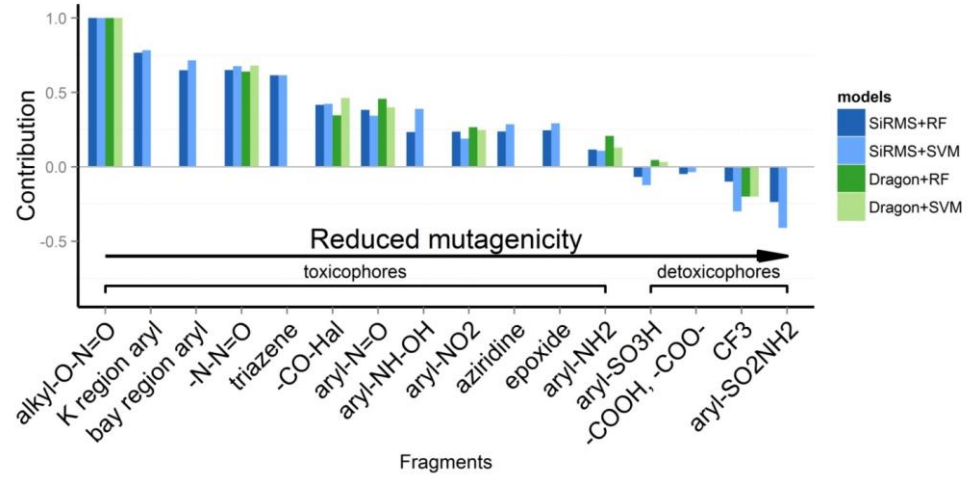


Figure 18. Average contributions of fragments to compound mutagenicity calculated from models trained on simplex (SiRMS) and Dragon descriptors.

We implemented interpretation of QSAR models in the automatic modeling pipeline available as an open-source Python package *spci* with graphical user interface (https://github.com/DrrDom/spci) and the complimentary R package *rspci* (https://github.com/DrrDom/rspci) which facilitates analysis of interpretation output. We also provide a basic web-application https://spci.imtm.cz.

2.3. Physicochemical interpretation of QSAR models

The suggested approach to structural interpretation answers the question "how a fragment influences modeling property/activity", but it does not answer the question "what is a reason of the observed contribution of a fragment". To answer the latter question we suggested the extension of the interpretation scheme[51]. If we encode molecules by descriptors representing different physicochemical properties, e.g. lipophilicity, partial atomic charges, hydrogen bonding, etc, we can retrieve contribution of fragments in terms of individual physicochemical properties. To achieve this we virtually remove a fragment in terms of a particular type of descriptors only, e.g. H-bonding, predict activity for a molecule with a "masked" fragment in terms of H-bonding and calculate the difference between predicted activity for the initial molecule and a "masked" one. The difference will represent the contribution of a removed fragment in terms of their H-bonding ability (Figure 15). This physicochemical interpretation imposes some restrictions to interpretation – it is only applicable to models built using descriptors with clear physicochemical meaning and interpretation, but it may provide deeper insights on the underlying property.

Simplex descriptors are perfectly suitable for such physicochemical interpretation, because they allow to label atoms by different physicochemical properties. Usually, we use (i) partial atomic charges to represent electrostatic interactions, (ii) lipophilicity to represent hydrophobic interactions, (iii) refractivity to represent dispersive interactions and (iv) H-bonding. Thus, every compound is represented by four sets of simplexes descriptors which are concatenated.

Here, we will represent one example of application of physicochemical interpretation for QSAR models of permeability of the blood-brain barrier (BBB). For more examples one may refer to the paper[51]. The data set was collected from 178 compounds which permeate BBB and 143 non-permeable ones. All compounds were checked to pass BBB mainly by passive diffusion. GBM, RF and SVM models were built using 2D simplex descriptors labeled by partial charge, lipophilicity, refractivity and H-bonding. A consensus model was created by averaging predictions of three individual models. All models have comparable predictive performance estimated by five-fold cross-validation (balanced accuracy = 0.75-0.77).

Fragments that represent rings and common functional groups were chosen for interpretation of QSAR models. The overall fragment contributions calculated from different models were in a good

agreement with each other ($R_{Pearson}$ = 0.78-0.90). Thus for analysis we used contributions calculated from the consensus model. The retrieved structure-activity relationship trend was the following: CF_3, phenyl and halogens mostly enhanced permeability, whereas thiazole, amide, carboxy, nitro and many other lower permeability of compounds (Figure 19).

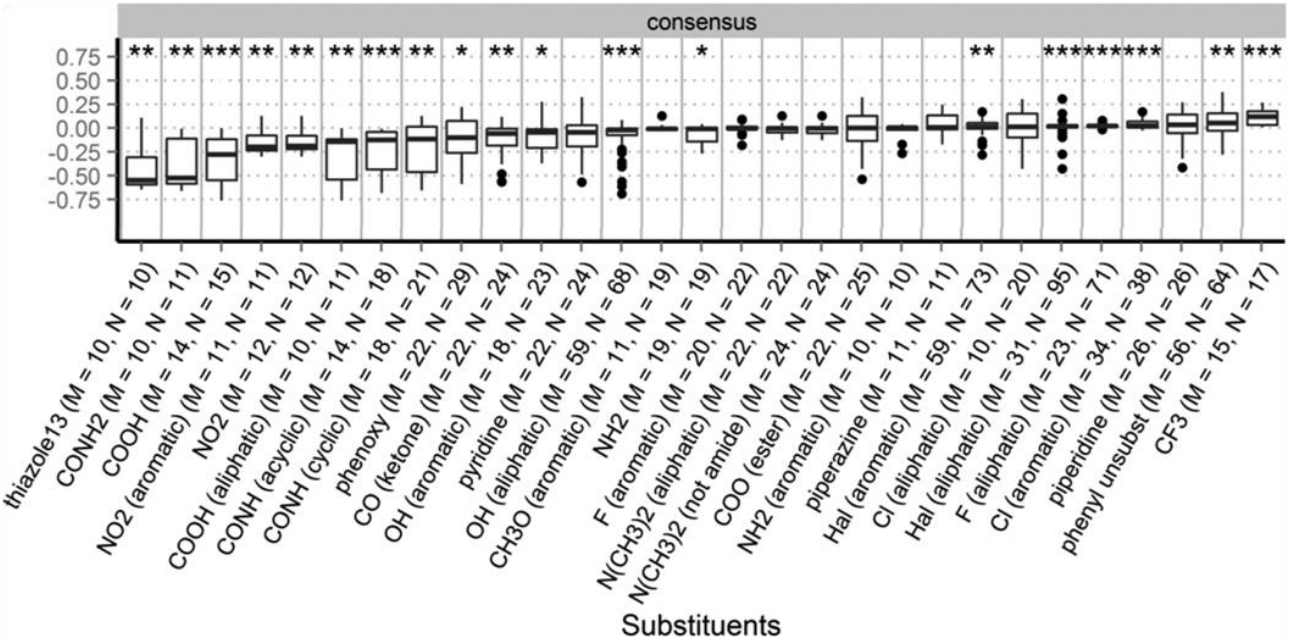
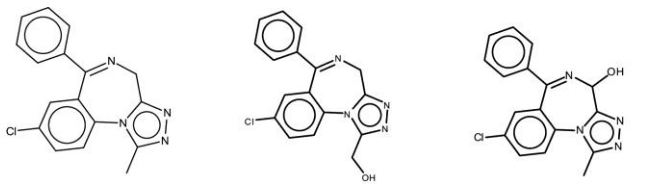
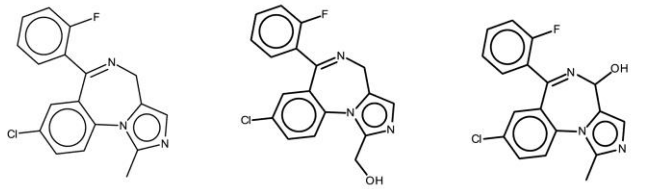


Figure 19. Distribution of fragment contributions calculated using the consensus BBB model. Only fragments occurring in at least 10 compounds are shown. Numbers in brackets: M is the number of compounds containing a fragment, and N is the number of fragments across the whole data set (some compounds have several identical fragments, and their contributions were estimated separately). Asterisks refer to statistical significance calculated by the two-sided Wilcoxon rank test (p value): ***, $p < 0.001$; **, $p < 0.01$, *, $p < 0.05$.

Fragments with wide distributions of contributions have a particular interest. This means that the contribution of such fragments is highly context dependent. For example, contributions of aliphatic hydroxy group are close to zero for the majority of compounds, but there were multiple outliers. The deeper analysis revealed that these were benzodiazepine derivatives. Derivatives bearing a hydroxyl group were indeed less permeable than the parent compounds (Figure 20). Thus, models correctly recognized the effect of the hydroxyl group in these cases and interpretation could reveal this influence. This is an example of "activity triggers" – groups which have large contributions and can substantially change activity of compounds. Although this particular case can be also captured by matched molecular pairs analysis, for more complex and diverse data sets, where the number of matched molecular pairs will be low or none, interpretation of QSAR models can bring a lot of new information which cannot be retrieved by other methods.



	alprazolam	1-hydroxyalprazolam	4-hydroxyalprazolam
contribution of OH		-0.69	-0.56
obs BBB class	1	0	0



	midazolam	1-hydroxymidazolam	4-hydroxymidazolam
contribution of OH		-0.35	-0.61
obs BBB class	1	0	0

Figure 20. Consensus contributions of hydroxyl groups to BBB permeability in benzodiazepine derivatives. Addition of a hydroxyl group to alprazolam and midazolam makes compounds non-permeable through the blood-brain barrier that was correctly captured by QSAR models.

Physicochemical interpretation of the consensus QSAR model revealed that possible formation of hydrogen bonds is the most important factor in low permeability of compounds containing thiazole, nitro, and carboxylic groups (high negative H-bonding contributions, Figure 21). This can be explained by strong interactions of such groups with water medium and the necessity of desolvation before passage through a membrane. At the same time, CF_3 group (a moderate positive H-binding contribution) is unlikely to form H-bonds, and this is preferable for BBB permeability. The carbamoyl group has a large negative contribution of electrostatic factors, which means it may have an unfavorable distribution of partial atomic charges. These findings are in a good agreement with earlier established rules and accumulated knowledge. A number of studies have indicated a negative effect of a large number of H-bond donors/acceptors, which should be less than 3. At the same time, the topological polar surface area should be less than 80$Å^2$.[53-56]

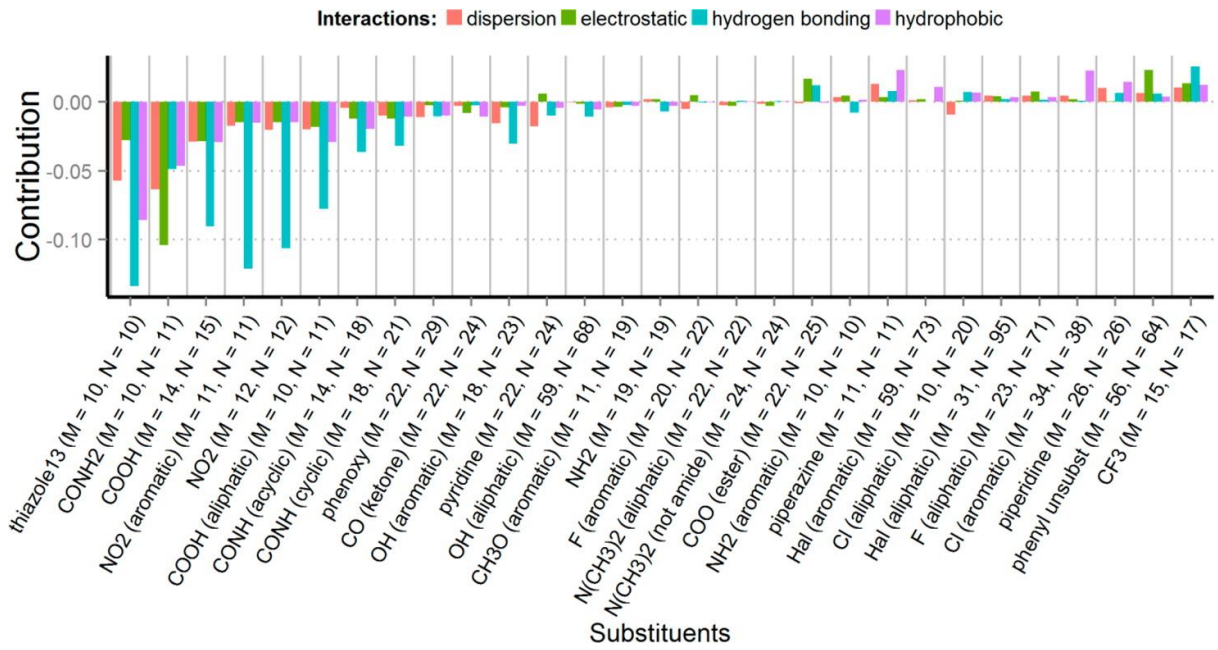


Figure 21. Median fragment contributions of different physicochemical factors estimated from the consensus BBB model. Only fragments occurring in at least 10 compounds are shown. Definitions of *M* and *N* were given in the Figure 20 caption.

The described approach was contributed to the open-source Python package *spci* (https://github.com/DrrDom/spci).

2.4. Deconvolution of influence of chemical context on fragment contributions

As we showed in the previous sections the contributions of fragments can be context dependent. Here is another example. The contributions of ester groups to BBB permeability of pseudocacaine differ depending on their context (Figure 22). Context dependence of fragment contributions is a natural attribute of modeling of non-additive properties and it should be observed regardless of a model or an interpretation approach used. Wide distributions of calculated fragment contributions were also observed for RF and PLS models of 5-HT_{1A} ligands (Table 4). Visual inspection to identify specific context features explaining high dispersion of fragment contributions is very tedious and laborious for large data sets. Therefore, it is important to develop an automatic pipeline which may assist in the task of deconvolution of the effect of structural context.

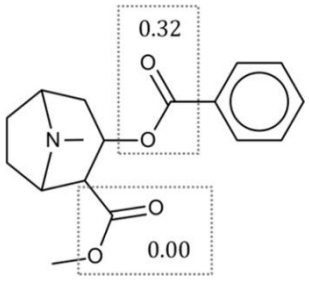


Figure 22. Contribution of hydroxyl groups of pseudocacaine in permeability of the blood-brain barrier.

We suggested an approach which can be used on top of the outputs of interpretation approaches calculating contributions of individual fragment, like those ones described above. When one aggregates contributions of the same fragment in different molecules the following patterns can be observed. If a distribution of fragment contributions is narrow this can mean that the context is the same or it does not influence the contribution of the fragment. If the distribution is wide there can be two possible cases – a wide distribution with a single peak or with several ones. Both these cases may indicate significant context dependence, but only in the second case this can be deconvoluted relatively easy. We suggested a protocol to analyze distribution of fragment contributions which consists of several steps[57] (Figure 23):

1) we establish existence of several peaks (clusters) in a distribution by applying Gaussian Mixture Modeling (GMM)
2) we choose a cluster (usually with the highest average contribution) and label all compounds comprising fragments from this cluster as "active" whereas all other compounds from other clusters as "inactives"
3) apply SMARTSminer [58] to the newly created data set to automatically establish generic patterns encoded by SMARTS which discriminate "actives" from "inactives".

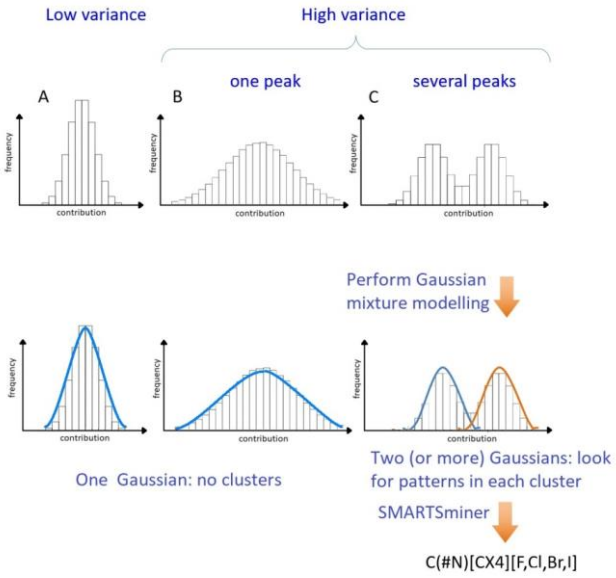


Figure 23. Workflow of the analysis of the context-dependence of fragment contributions.

We demonstrated applicability of the developed approach on the data set of 1984 compounds with measured toxicity ($pIGC_{50}$) against *Tetrahymena Pyriformis*. SVM, RF and GBM models, which were built using 2D simplex descriptors, had comparable predictive performance according to five-fold cross-validation as well as the consensus model obtained by averaging predictions of individual models, R^2_{5CV} = 0.73–0.77. Therefore, for further analysis we used only the consensus model. To get fragments we exhaustively fragmented all compounds by breaking up to three single

acyclic bonds. Fragments which were occurred less than 10 times were removed. For the remaining 311 fragments we calculated contributions using the previously described universal interpretation approach and analyzed their distributions by GMM (Figure 24). For further analysis we chose 118 fragments, where two or more clusters were detected, and 39 fragments had narrow distributions.

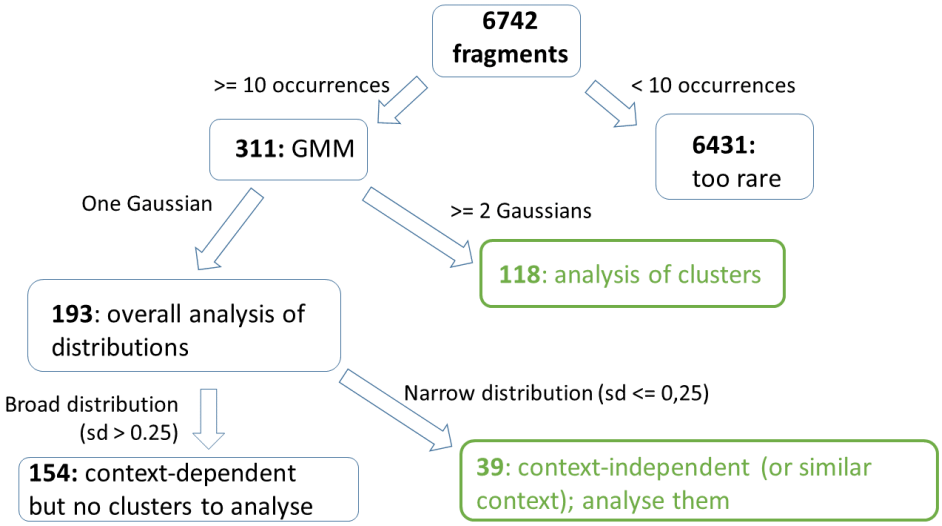


Figure 24. Decision tree illustrating the workflow for the analysis of fragments contibution to *Tetrahymena Pyrofirmis* toxicity. Green boxes contain fragments to be analysed. The upper green box contains the fragments of the major interest to this study since several clusters were found in their distributions. The lower green box contains fragments having narrow distributions with no clusters (standard deviation <= 0.25).

The largest average contributions (around 1.0–2.0) amongst 39 fragments having narrow distributions corresponded to various aromatic fragments, e.g. 4-bromophenyl, benzoate group, etc. This can be explained to a large part by their high lipophilicity and hence the implicit relationship to non-polar narcosis. However, some of them, such as benzaldehyde derivatives, can be reactive.

Halogen atoms except fluorine had large median contributions, F (0.25) < Cl (0.52) < Br (0.72) < I (0.91). However, in the case of Cl, Br and I a long right tail was observed on distribution diagrams, which was identified as a separate cluster by GMM (Figure 25). Application of SMARTSminer to indentify patterns distinguishing compounds in two clusters revealed that halogen atoms had much greater contributions if they occurred in the activated environments: α-haloketones, esters, amides or alkenes. Halogens with lower contributions from the first clusters were other aliphatic or aromatic derivatives (Table 7).

The findings for bromine and chlorine fragments were in accordance with the understanding that activated halogens (e.g. adjacent to an ester or other unsaturation) are electrophilic in nature and will have a strong influence on toxicity [59]. Specifically, the reactivity of α-haloactivated compounds occurs as a result of their reactivity in Phase II enzymes. It is mediated by a S_N2-type of transition state with the partially negative charged sulfur atom from the thiol groups of glutathione S-

transferases. It was noted that the halo-substituted compounds of this type were one of eight classes of S_N2 electrophiles [60].

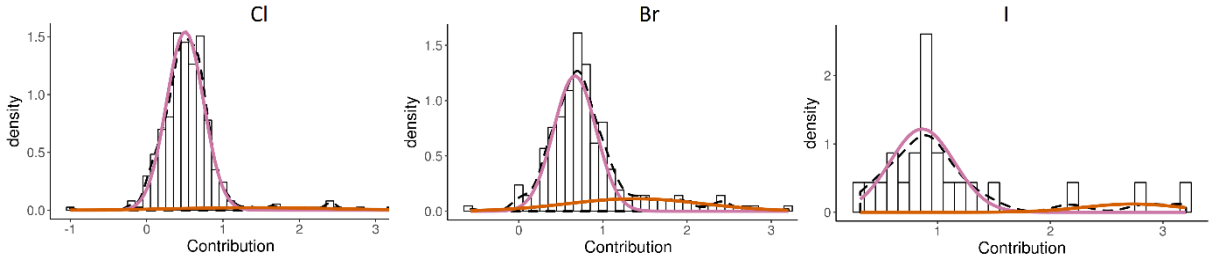
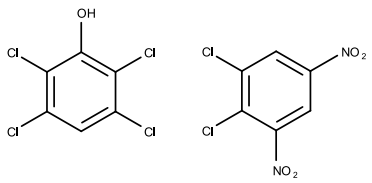
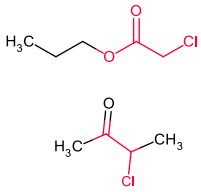
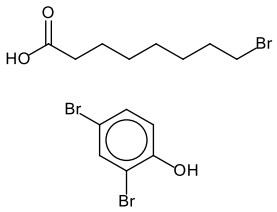
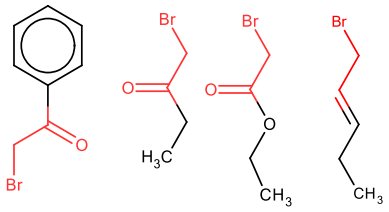
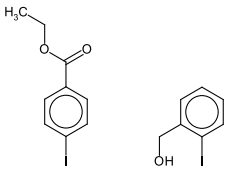
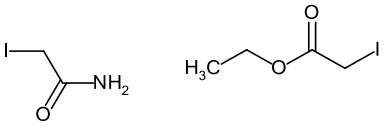


Figure 25. Distributions of contributions of halogens (Cl, Br, I) with regard to their toxicity to *Tetrahymena pyriformis*. Histograms and dashed lines represent observed distributions of fragment contributions. Solid colored lines represent Gaussians detected by GMM.

Table 7. Examples of SMARTS patterns and molecules corresponding to each cluster detected by GMM for chlorine, bromine and iodine (Figure 25). SMARTS patterns matched in structures are colored in red. For iodine no SMARTS patterns were identified due to the small number of compounds in clusters.

		Cluster 1	Cluster 2
Cl	Mean contribution ± standard deviation	0.47 ± 0.24	1.16 ± 1.02
	Coverage	95%	5%
	SMARTS		A[CD3H0](CCl)=[OX1-0]; C(Cl)[CD3H0]
	Examples		
Br	Mean contribution ± standard deviation	0.64 ± 0.24	1.69 ± 0.79
	Coverage	77%	23%
	SMARTS		A[CX4][CX3]=[C,O]; Br[CX4][CX3]
	Examples		
I	Mean contribution ± standard deviation	0.83 ± 0.27	2.72 ± 0.46
	Coverage	87%	13%
	SMARTS		
	Examples		

Similar results to halogen atoms were obtained for other fragments represented by several clusters. Thus, it was identified that methylcarbonyl groups had contributions mainly within the range -0.5 to 0.5, but if they were conjugated with a double or triple bond or a halogen atom the contribution of this group substantially increased. More examples can be found in the article [57].

SMARTSminer was also applied directly to the whole data set in order to directly find possible toxicophoric patterns without QSAR modeling. Since the SMARTSminer approach is only suitable for working with classification tasks, two subsets of compounds were selected based on thresholds: the "negative" set with 500 compounds having $pIGC_{50}$ <= 2.5 and the "positive" set with 406 compounds having $pIGC_{50}$ >= 5. No patterns were found by running SMARTSminer with the chosen settings for positive and negative support values, 0.7 and 0.3 correspondingly. This means that there were no patterns detected on at least 70% of molecules from the "positive" set and at most in 30% on molecules from the "negative" set. Decreasing positive and negative thresholds to 0.6 and 0.2, respectively, helped retrieve about 100 patterns. They mostly matched aromatic and some heteroaromatic substructures which were abundant in the "positive" set of compounds (Figure 26) and less frequent in the "negative" set.

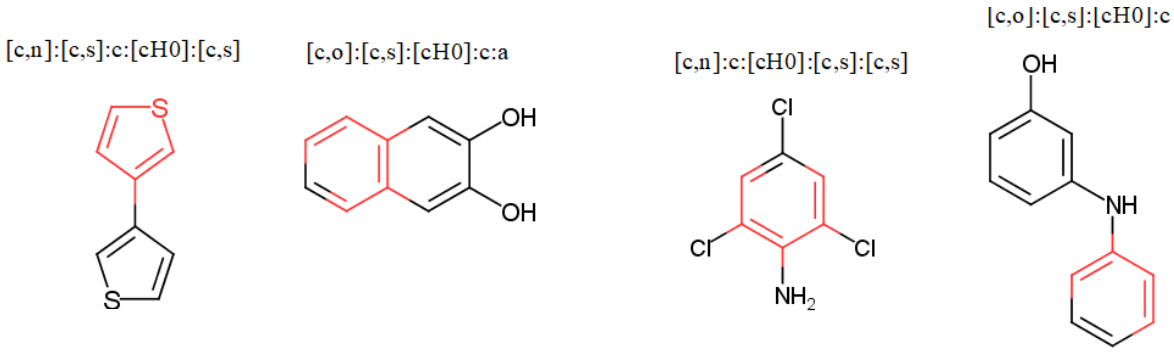


Figure 26. Top-ranked discriminative patterns found by SMARTSminer to discriminate high from low toxicity compounds and examples of matched compounds from the „positive" set.

Further decrease of support values did not help; numerous general patterns matching mainly aromatic substructures were found. The patterns identified by our approach could not be found by SMARTSminer directly because all of them have very low positive support values (<0.1). Poor performance of SMARTSminer can be explained by the high structural diversity of the compounds in the data set and different, or mixed, mechanisms of toxic action. This additionally highlighted the unique ability of interpretation of QSAR models to reveal structure-activity relationships in complex and diverse data sets.

The described approach was implemented in R language and contributed to the R package *rspci* (https://github.com/DrrDom/rspci).

2.5. Benchmarking interpretation of QSAR models

Wide application of deep neural networks for QSAR modeling stimulated interest to interpretation approaches to understand and explain model decisions. Modern deep learning approaches frequently use end-to-end modeling, take a molecular graph or SMILES as input, create own internal representation (embedding) and establish a correlation with a modeling property. This makes many of commonly used approaches not applicable and, thus, requires specific interpretation approaches developed for these methods [61]. There are multiple *post hoc* interpretation approaches developed specifically for neural network models, e.g. Layer-wise Relevance Propagation (LRP)[62], DeepLift[63], CAM[64], GRAD-CAM[65], etc. Some deep learning approaches are interpretable by design, e.g. attention-based neural networks[66]. Weights assigned by the attention layer can be interpreted as importances of corresponding features. In graph-based networks these features can be atoms and thus importance of individual atoms within a molecule can be established. Machine learning agnostic interpretation approaches can be applied to modern deep learning models too. Examples of such feature-based approaches include Integrated Gradients[67], Shapley values[68, 69], which return a contribution of individual descriptors. The latter approaches belong to the old paradigm "model→descriptor→structure" with the consequent limitations – these approaches are applicable to models built with interpretable descriptors only.

Despite the fact that multiple interpretation approaches have been developed and new ones constantly appear there are no suitable benchmarks to evaluate their applicability to interpretation of QSAR models. Often authors demonstrate applicability of their interpretation approaches on well-studied end-points like lipohilicity, solubility or toxicity where relevant patterns are well known [70]. Interpretation is frequently performed for pre-defined motifs or on a very limited number of considered examples [50, 70, 71]. For example, authors visually inspect a subset of molecules and compare calculated contributions with expert knowledge. Such non-systematic evaluation can be biased by a human expert and the choice of inspected molecules. Real data sets may have hidden biases which are difficult to control, some properties may depend on multiple factors or the response can be caused by different mechanisms of action. All these issues complicate proper validation of interpretation approaches based on real-world examples.

Synthetic data sets are more reasonable to evaluate interpretation approaches. They are designed in such a way that end-point values are pre-defined according to some logic according to true patterns chosen by a designer. These data sets can be used to investigate the ability of models to capture the introduced logic and the ability of interpretation approaches to retrieve it.

Two recent studies used artificial data sets. The study of Sheridan was mainly focused on comparison of interpretation of models built using different conventional descriptors and machine learning methods [72]. The author used similarity maps [73] for model interpretation which provides atom

contributions (colors) and which is a special case of the universal interpretation approach[50]. Besides real-world datasets, two artificial, "idealized" ones were utilized. Both represented simple additive properties: heavy atom counts and the number of negative charges in compounds at pH 7.4. Regarding the impact of descriptors (D) and models (Q) on interpretation quality the author concluded, that "… one has to have a very high cross-validated predictivity to recover those expected colors, and not all D/Q combinations are suitable". Another study utilized Integrated Gradient interpretation method[67]. Authors used graph convolutional models and their interpretation produced atomic contributions. They created 16 synthetic classification data sets. Compounds were retrieved from ZINC database and satisfied to particular positive or negative SMARTS patterns combined by Boolean operators, e.g. compounds were labeled as active if they comprise a naphtyl group and no amino groups. The goal of the study was to investigate the ability of models to retrieve atoms corresponding to these positive and negative patterns and it was demonstrated that models could not always recognize true atoms correctly. Unfortunately, authors did not provide data sets to enable comparative studies. Moreover, those data sets represented only one possible scenario of structure-property relationship where the property of compounds depended on local chemical context encoded by SMARTS that is not always relevant to actual structure-activity relationships observed in drugs and biologically active compounds where distant groups can have a cooperative effect which cannot be easily encoded by a single SMARTS pattern.

We developed synthetic data sets with pre-defined patterns determining end-point values and controlled possible biases in data sets[74]. Creation of these data sets will enable systematic evaluation of interpretation approaches to validate their ability to retrieve structure-property relationships captured by QSAR models, because calculated contributions of atoms or fragments can be compared with expected values determined by the incorporated logic ("ground truth"). We developed regression and classification data sets, which represent different logics and levels of end-point complexity.

1. Simple additive end-points, where specific contributions were assigned to individual atoms and the sum of atom contributions determined a compound property value. This represents "idealized" additive properties. There were three data sets: (i) *N data set* where the end-point was the count of N atoms; (ii) *N-O data set* – the end-point was the count of N atoms minus the count of O atoms; (iii) *N+O data set* – the count of N plus O atoms divided by two, but the number of N and O atoms for each molecule was the same (this is the special case to study interpretability if there are correlated features in a data set).

2. Additive end-points depending on a local chemical context, where contributions were assigned to groups of atoms and the sum determined the property value of a compound. This is related to molar refractivity or lipophilicity modeling, where group contribution methods are successfully applied[75, 76]. There were two data sets where end-points were: (i) the presence of at least

one amide group matched by SMARTS NC=O (classification task) and (ii) the count of such amide groups (regression task).

3. Pharmacophore-like settings, where compounds were labeled as “active” if they had a specific 3D pattern. For simplicity we chose a two-point 3D pharmacophore where an H-bond acceptor and an H-bond donor were on the distance of 9-10Å. This case was the most similar to real problems, where property depends on distant features and their mutual position in 3D space.

Compounds were chosen by randomly picking molecules from the curated ChEMBL23 database to get 10000 molecules in every data set, which were split on train and test sets (70/30). Structures of data sets were additionally controlled to do not have obvious correlated patterns to reduce a possible hidden bias. All six data sets were balanced (the number of “active” and ‘inactive” molecules was equal) or had the distribution of end-point values close to the normal distribution.

Structural interpretation retrieves contributions of particular atoms or fragments. To quantitatively measure interpretation performance we suggested to use several metrics. AUC is an integral metric which demonstrates how well relevant patterns are ranked over others within a particular molecule. We calculate AUC values for individual molecules that shows how well true atoms ranked in a molecule. To get the final score we averaged AUC values for all considered molecules. In QSAR interpretation context this metric was first used by McCloskey et al[67].

The weakness of AUC is that it is an integral measure, which accounts for both relevant and irrelevant patterns. In practice it is more reasonable to focus only on relevant features tanked on top. To address this, we proposed top-n score which is calculated as follows and should be more stringent:

$$\text{top - n score} = \frac{\sum_i m_i}{\sum_i n_i} \tag{7}$$

where n_i is the total number of “positive” atoms in the i-th molecule, m_i is the number of “positive” atoms in n_i top ranked atoms according to their calculated contributions. For instance, if a molecule has two true patterns with expected contributions +1 and interpretation retrieved only one of them among top two contributing patterns, the molecule will contribute n = 2 and m = 1 to the equation above. Top-n is an integral characteristic of a data set and varies from 0 to 1 (perfect interpretation) and it is closely related to early enrichment, where only top scored patterns are considered to calculate performance.

Additionally, we suggested to calculate root mean square error (RMSE) of predicted contributions for each molecule and averaged them across molecules in a data set to estimate deviation of calculated contributions from the expected values. This is less important metric, because proper ranking is more practically valuable than exact estimation of contributions which are generally unknown in real cases.

We applied this benchmark suite to study the previously suggested universal structural interpretation approach to calculate atom and fragment contributions.

For benchmarking we built RF, SVM, GBM and PLS models based on different 2D RDKit fingerprints: (i) atom-pair fingerprints, which enumerate pairs of particular atoms at the topological distance from 1 to 30 (AP); (ii) Morgan fingerprints, which enumerate atom-centered substructures of radius 2 (MG2); (iii) RDK fingerprints, which enumerate all possible substructures with atom count from 2 to 4 (RDK) and (iv) topological torsion fingerprints, which enumerate all possible linear substructures with four atoms (TT). AP, MG2 and RDK fingerprints were also used in their binary (bit vector) form of length 2048 (bAP, bMG2, bRDK). We also built graph convolution (GC) models, which use end-to-end modeling and for which we specifically implemented the same interpretation approach and contributed it to the deepchem repository.[77]

There was a strong or moderate correspondence between AUC and top-n interpretation measures and model predictive abilities for almost all data sets (Figure 27 and Figure 28). This supports the hypothesis that interpretation performance depends on the model prediction accuracy. In general, atom-pairs descriptors followed by Morgan fingerprints resulted in models with the best interpretation performance. The dependence of interpretation performance of models from the machine learning methods was not obvious. Therefore, we can make a conclusion that descriptors are more important than machine learning methods for high quality interpretation of QSAR models.

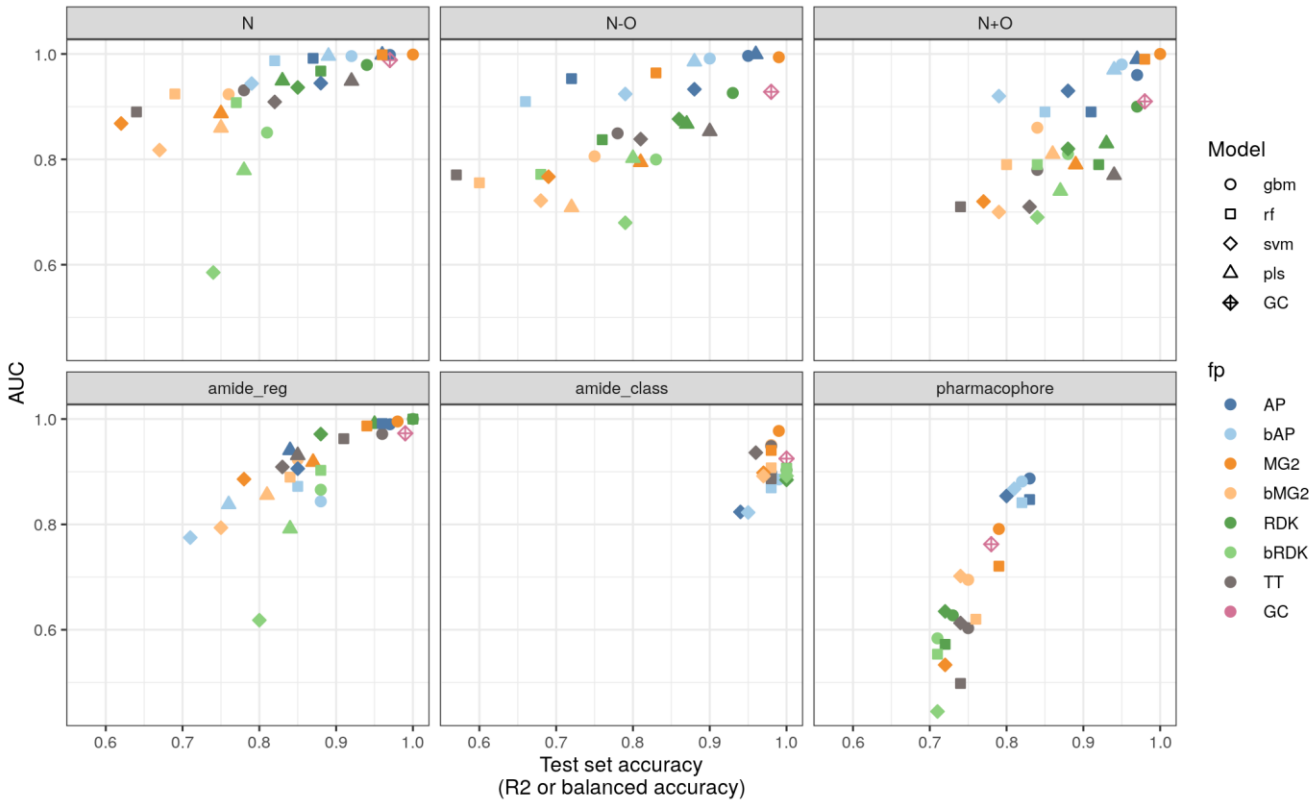


Figure 27. Interpretation performance (AUC) vs. model prediction accuracy.

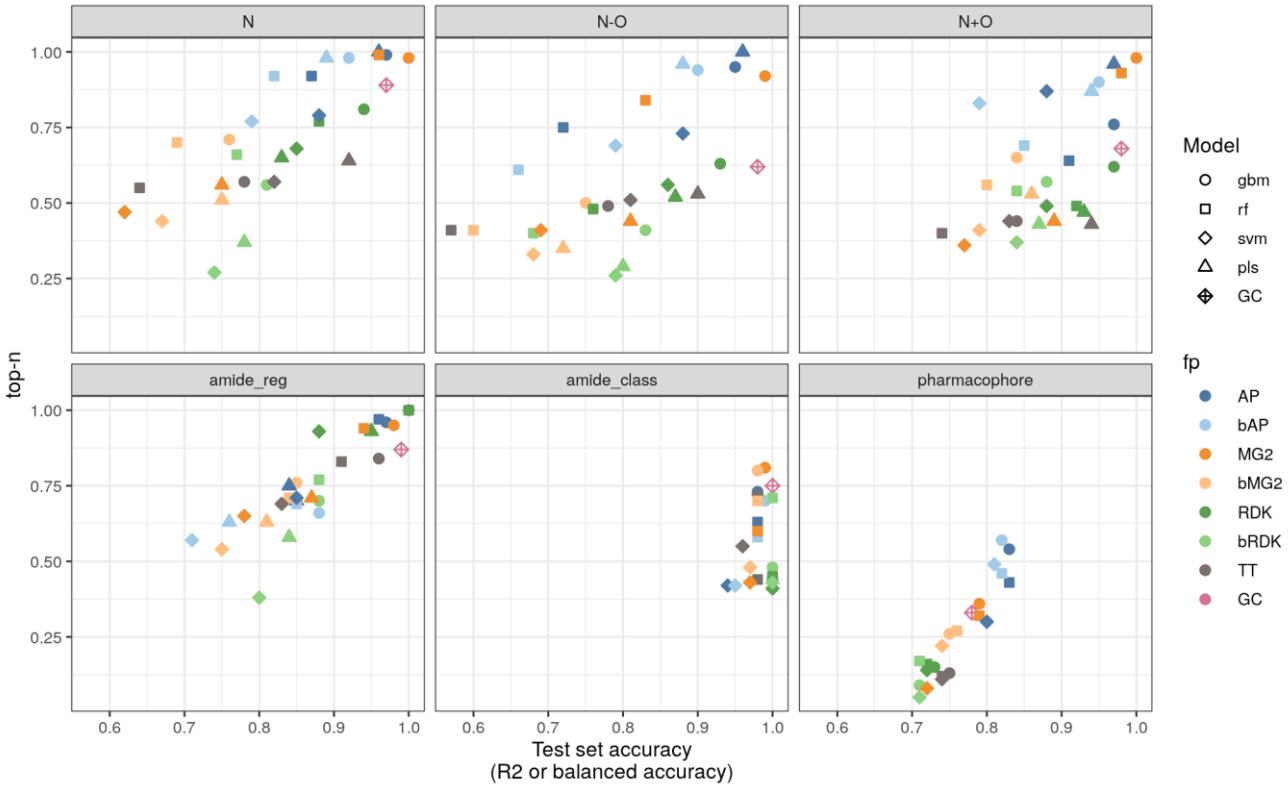


Figure 28. Interpretation performance (top-n scores) vs. model prediction accuracy.

The analysis of classification models for the amide data set showed that all models have nearly perfect predictivity (balanced accuracy on the test set ≥ 0.94), however, many of them had relatively poor interpretation results (AUC = 0.82-0.98, top-n score = 0.41-0.81) (Figure 27 and Figure 28). This can be explained by the presence of multiple true patterns in "active" molecules, making major impact on interpretation performance. If there are multiple identical true groups in a molecule, masking of one group does not change much the predicted class probability due to the presence of the remaining groups and therefore the contribution is close to zero. This was confirmed by the fact that interpretation performance for compounds containing only one true pattern was very high whereas for compounds with multiple true patterns it was dropped substantially (Table 8). This issue will be common for all similar interpretation approaches using "masking" technique if identical atoms/fragments are "masked" one by one and not simultaneously.

Table 8. Interpretation performance of selected models for the classification amide data set calculated for subsets of molecules having different number of amide groups.

count of amide groups	GBM / MG2			GBM / AP			GC		
	mean AUC	top-n	mean RMSE	mean AUC	top-n	mean RMSE	mean AUC	top-n	mean RMSE
all	0.98	0.81	0.12	0.90	0.73	0.14	0.92	0.75	0.14
0	-	-	0.03	-	-	0.02	-	-	0.02
1	1	0.98	0.12	0.96	0.89	0.17	1	0.98	0.2
2	0.94	0.69	0.4	0.77	0.58	0.42	0.81	0.56	0.36
3	0.9	0.65	0.51	0.75	0.6	0.53	0.66	0.44	0.52
4	0.87	0.62	0.57	0.6	0.45	0.57	0.53	0.39	0.57
5	0.8	0.44	0.58	0.57	0.47	0.58	0.54	0.33	0.57
6	0.66	0.55	0.67	0.49	0.39	0.67	0.61	0.48	0.67

The classification amide data set most notably demonstrated the weakness of AUC as the measure of interpretation accuracy. AUC indicates the ability of models to rank true pattern high. However, there is a small difference in AUC values between models which score true patterns on top and those which score true patterns high but not exactly on top. In the latter case it may be difficult to identify true patterns among other highly scored ones. For example, AUC for GBM and RF models trained on Morgan fingerprints were 0.98 and 0.94 respectively, but corresponding top-n scores were 0.81 and 0.60 indicating that probability to find true patterns among the top scored ones was much higher for the GBM model. Therefore, the proposed top-n score is a more practically relevant and robust measure of interpretation accuracy.

We specifically studied interpretation of models having correlated features using the designed *N+O* data set. In this case the count of N and O atoms were the same for all molecules. There are two possible outputs for this data set because to get perfect predictions a model can capture only one of correlated features or assign comparable contributions to both of them. In the former case interpretation may be incomplete and misleading. Therefore, it is useful to know how different models behave in this situation. For count-based fingerprints RF and GBM models based on decision trees frequently prioritized one of two true correlated features but both of these patterns were ranked higher than non-relevant atoms. PLS and SVM models assigned mostly comparable contributions to N and O atoms but not always discriminate them well from other atoms. GC model also assigned comparable contributions to N and O atoms, but it differentiated them quite well from other non-relevant ones. So, these properties of models should be taken into account in future interpretation studies.

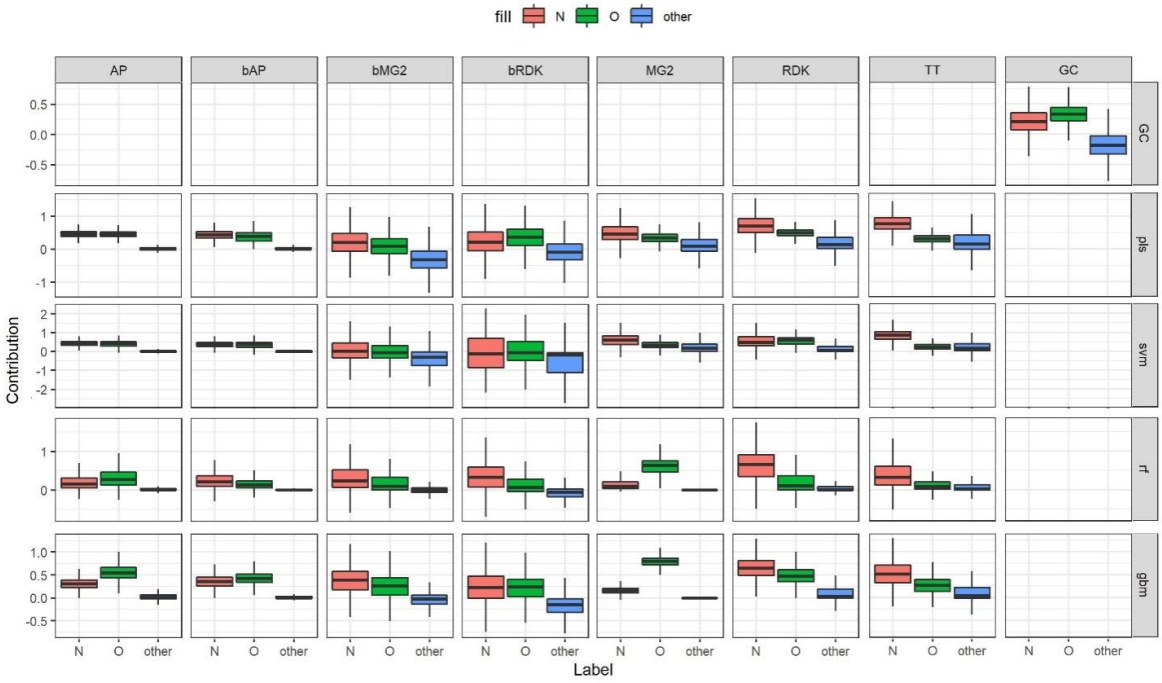


Figure 29. Contributions of nitrogen, oxygen and other atoms for models trained on the *N+O* data set.

The pharmacophore dataset was the hardest task and models achieved just moderate balanced accuracy, 0.71-0.83. Thus, it was expected that interpretation accuracy may be relatively low. The predictive ability of conventional models mostly depended on descriptors type. The model accuracy decreased in the following row: atom pairs > Morgan fingerprints > topological torsion > RDK fingeprints. For this data set the correlation between the model predictive ability and interpretation accuracy was the most pronounced (Figure 30). Several models had poor interpretation, average AUC values close to 0.5 (random ranking) or even lower, however the predictive ability of these models was moderate ($R^2_{test} \geq 0.71$). This suggests that models which are usually considered as acceptable according to their predictive ability may result in ranking ability of patterns close to random choice. Top-n scores were also low. Even for the most predictive models with balanced accuracy >0.8 the top-n scores were 0.43-0.57. This means that only about 50% of true pharmacophore centers can be identified within top 2 atoms ranked by their contributions.

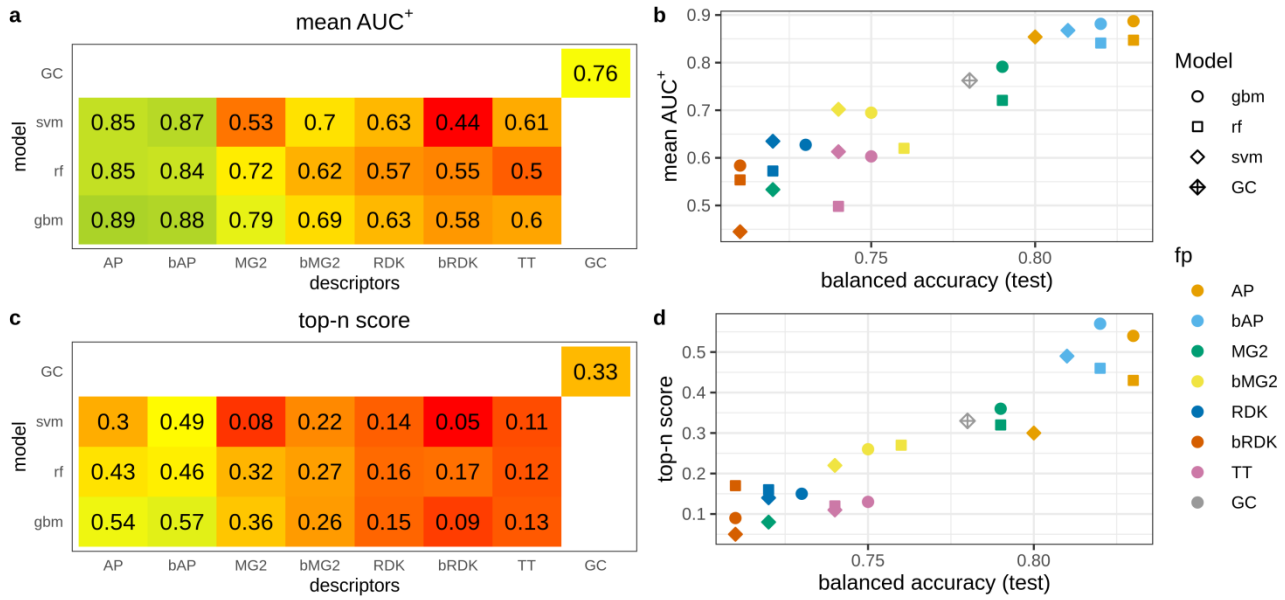


Figure 30. Interpretation performance of models trained on the pharmacophore data set.

To overcome the issue of low performance of atom-based interpretation we suggested to calculate contributions of fragments. This may help to identify fragments comprising relevant atoms and, thus, approximately locate true centers in molecules. For many applications this information may be enough. For example, identification of fragments containing sites of metabolism may help to modify corresponding molecules to avoid metabolic transformations even without knowledge about an exact site of metabolism. To perform fragment-based interpretation we exhaustively fragmented molecules by breaking up to three single acyclic bonds and kept only fragments of the size up to 7 heavy atoms and containing not more than 40% of the total number of heavy atoms in a molecule. We calculated the average percentage of correct pharmacophore centers comprised in top 2 fragments with the highest contributions. This metric directly corresponds to top-n score for atoms, because each “active” molecule had only two true centers (atoms). Fragment-based interpretation could substantially improve accuracy to detect true pharmacophore centers for models which had poor atom-based interpretation performance. However, performance for models with the highest atom-based interpretation accuracy (GBM/AP and GBM/bAP) was not increased that much (Figure 31). So, fragment-based interpretation may be preferable for interpretation of models with the moderate or poor predictive ability.

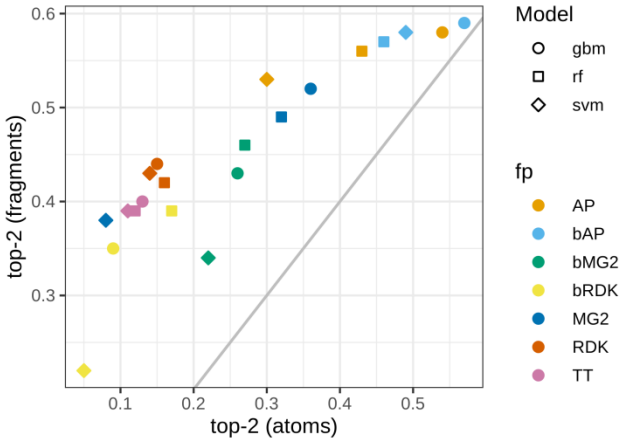


Figure 31. Top-2 score for atom- and fragment-based interpretation of models trained on the pharmacophore data set.

While interpretation of conventional models resulted in mostly explainable performance, interpretation performance of GC models raised a question. GC was among models with the highest predictive ability and one may expect highly accurate interpretation results. However, in some cases there was a large difference between interpretation performance (mainly in top-n scores) for GC and conventional models having the comparable predictive ability. This was observed for regression tasks. For example, in the case of N-O data set top-n score for GC model was 0.62 while the model accuracy was very high, R^2_{test} = 0.98 (Figure 32). Conventional models with lower predictive performance had much higher interpretation accuracy (e.g. RF/AP has R^2_{test} = 0.72 and top-n score = 0.75) and, thus, conventional model better captured the true structure-property relationship.

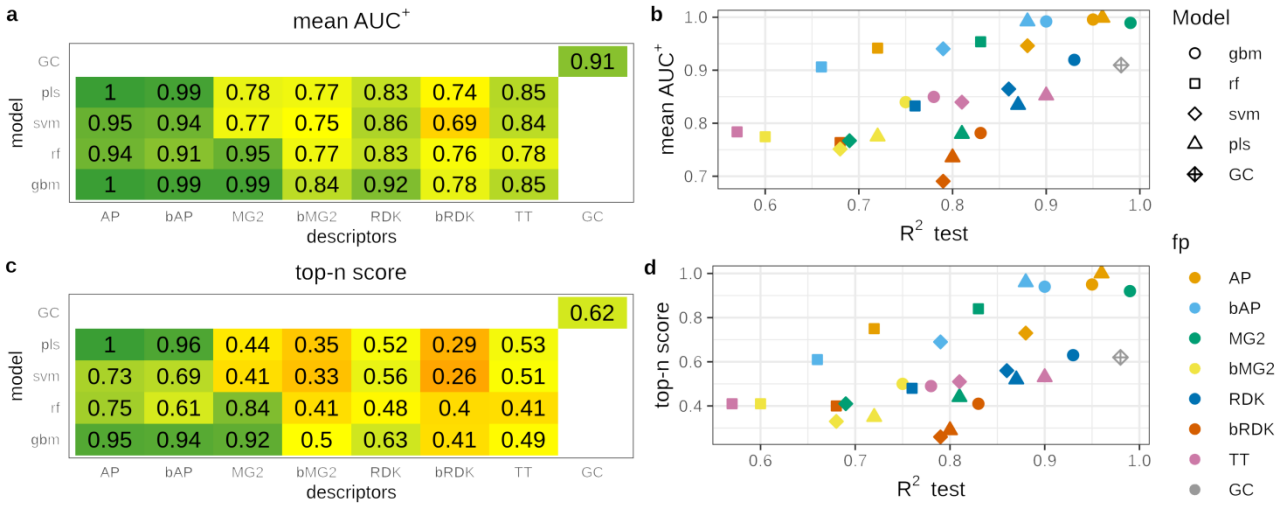
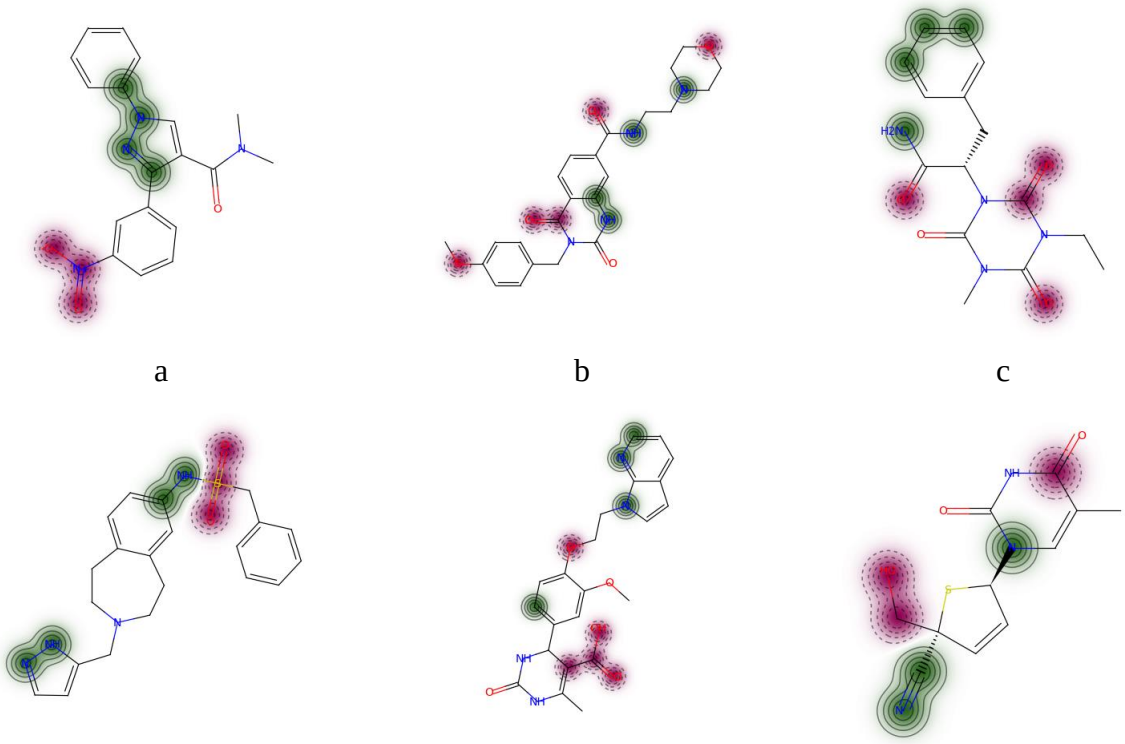


Figure 32. Interpretation performance of models trained on the N-O data set. Only performance to recognize positively contributed patterns (N atoms) are shown. Results for negatively contributed patterns (O atoms) were highly similar and omitted for clarity.

The analysis of interpretation results of GC model for the N-O data set revealed several sorts of misinterpreted patterns. Atoms attached to nitrogens were ranked on top (green) (Figure 33a-f). Sometimes nitrogens in nitro groups were misinterpreted as negatively contributing (Figure 33a). Aromatic carbons were frequently recognized as positive, though they were far from any nitrogen (Figure 33c,e).



a b c

d e f

Figure 33. Top-scored (green) and bottom-scored (pink) atoms by the GC model for the N-O data set. The number of highlighted top and bottom atoms is equal to the total number of „positive“ (nitrogen) and „negative“ (oxygen) atoms in corresponding molecules.

We can suggest two possible explanations:

1. The implemented interpretation approach is not fully suitable for GC models. It is difficult to prove whether a particular interpretation approach is suitable or not for a particular model, but comparison of interpretation performance with other models may shed some light. For classification tasks in contrast to regression we did not observe large discrepancy in top-n scores between GC and conventional models of similar predictive performance. This indirectly confirms validity of the chosen interpretation approach and its applicability to GC models. In future it would be reasonable to compare results obtained in this study with interpretation results of “orthogonal” interpretation approaches.

2. Hidden bias in data sets. It is impossible to control all possible biases in data sets. Conventional models which had high predictive ability comparable to the corresponding GC models demonstrated the much better ability to identify true patterns. Therefore, an explanation may lie in

the nature of GC models which learn sophisticated internal representation of molecular structure to find correlation with a given end-point. Thus, GC models may construct overly complex embeddings correlated by chance with true ones and the end-point. We demonstrated on the example of N+O data set, that GC models assign comparable contributions to correlated features. Thus, it can be difficult to distinguish true patterns from correlated ones. In other words, GC models may "amplify" a hidden bias and capture it. This ability of GC models can bring some advantages if the underlying true patterns are really highly complex. However, for tasks of biological activity prediction GC models did not demonstrate systematically better predictivity than conventional models[78]. Thus, constructing complex internal representations may not be necessary to capture relevant structure-activity relationships.

The benchmark data sets as well as supplementary scripts facilitating the analysis of model interpretation are distributed as an open-source repository - https://github.com/ci-lab-cz/ibenchmark. Although the benchmark suite was published quite recently it was already used by other researchers to investigate performance of the developing interpretation approaches.[79] However, there are a lot of opportunities to improve the benchmark. More complex patterns can be implemented, e.g. similar to those used by McCloskey et al[67], or more complex 3D pharmacophore patterns or a combination of pharmacophore patterns, to better reproduce real case scenarios.

2.6. Summary

This chapter described the evolution of development of approaches to interpretation of QSAR models. From development of model-specific approaches we passed to the universal structural interpretation approach which is applicable to any model regardless used machine learning method and descriptors. The validity of this approach was demonstrated in numerous studies. We also confirmed the universal applicability of the approach by implementing it for interpretation of graph convolution models. The development of this universal approach designated the shift from conventional the "model→descriptor→structure" interpretation paradigm to the new one, "model→structure". In the latter case contributions of structural motifs are calculated directly from models skipping the step of calculation of contribution of descriptors and their interpretation. Switching to the new paradigm makes all models interpretable and solves the long lasting belief that there is a trade-off between model predictivity and interpretability.[80] Thus, we argue that true "black box" QSAR models actually exist.

The developed universal approach for structural interpretation was further enhanced to explain which physicochemical properties contribute to the effect of individual fragments and to identify chemical contexts which are important for high contributions of considered fragments. We also

implemented the pipeline helping to automatically identify important chemical context which can explain discrepancy of calculated contributions of identical fragments.

We suggested and developed the first benchmark to evaluate approaches to interpretation of QSAR models. Using this benchmark we demonstrated that interpretation accuracy strongly depends on the prediction accuracy of models. However, in some specific cases even highly predictive models can result in poor interpretation. We also demonstrated in multiple examples that interpretation outputs are not sensitive to a used interpretation method or a model. Models built with different descriptors and machine learning methods mainly result in similar interpretation outputs if predictivity of models is comparable. However, interpretation performance may depend to some extent on descriptors used.

All approaches were implemented in open-source software to support further research in the emerging field of interpretation of machine learning models in chemoinformatics.

Chapter 3. Pharmacophore representation and modeling

Pharmacophore models are widely used in drug design due to their simplicity, interpretability and speed.[81] They demonstrated the competitive ability to discover hits[82], scaffold hopping[83], activity profiling abilities[84] and other applications[85]. Despite wide applicability of pharmacophore modeling the majority of available software are commercial and do not have free academic license[85, 86]. The only freely accessible programs are Pharmer[87] and Pharao[88] which provide virtual screening capabilities. There are also a few free web-applications. Pharmit is a web-based virtual screening tool of various databases powered by Pharmer[89]. PharmaGist is a web-server to create ligand-based pharmacophore models from a set of input ligands[90]. PharmMapper was designed to perform search among pharmacophore models retrieved from PDB ligand-protein complexes in order to identify possible targets for new compounds[91]. Recently much interest was attracted to pharmacophore modeling using molecular dynamics (MD) simulations that was encouraged by accessibility of high-performance computing systems to researchers. Molecular dynamic simulations may provide more comprehensive information about protein-ligand interactions and help to identify potential hot spots and pharmacophore centers in apo-proteins[85]. while many approaches were suggested and implemented only few tools were made publicly available as a Python module, e.g. PyRod[92], or as a web-application, e.g. Pharmmaker[93].

Our motivation was to develop open-source tools for pharmacophore modeling, which may complement existing ones and extend applicability of pharmacophore models. We developed 3D pharmacophore representation which enables alignment-free comparison of pharmacophores and takes into account their stereoconfiguration. Based on this representation we implemented a fully automatic ligand-based modeling protocol which can handle with large data sets and may identify pharmacophores for different binding modes. Using the same representation we enhanced MD pharmacophore analysis and suggested a new scoring scheme for virtual screening of ligands using MD pharmacophores.

3.1. pmapper - 3D stereosensitive pharmacophore multiplets

Our development of 3D pharmacophore multiplets[94] was inspired by the work of Mason et al[95] who suggested to encode pharmacophores by sets of all possible combinations of 3 or 4-point pharmacophore features where each distinct triplet/quadruplet activates a fixed bit in a bit string. This enables fast comparison of pharmacophores and compute similarity between them.

We consider pharmacophores as complete graphs with vertices labeled by the pharmacophore feature type and edges representing distances between features in 3D space. To distinguish enantioneric pharmacophores we implemented special treatment of stereoconfiguration. The full

algorithm is depicted in Figure 34. Pharmacophore features (N-bond donor/acceptor, positively or negatively charged center, aromatic and hydrophobic groups) are assigned to a conformer of a molecule using SMARTS patterns adapted from the work of Koes at el[87]. These features create a complete graph where distances between all pairs of features are binned with a chosen binning step (1Å by default). Afterwards we enumerate all possible quadruplets and for each quadruplet we generate a canonical signature. The quadruplet signature is a tuple consisting of two parts: one encoding content and topology of a quadruplet (canonical graph signature) and the other encoding stereoconfiguration.

Content and topology encoding

The quadruplet is considered a complete graph. One round of the Morgan-like algorithm[96] is applied to generate canonical identifiers of features taking into account its surroundings. New feature labels consist of the current label of the considered feature and lexicographically sorted labels and binned distances to all other features in a quadruplet (Figure 34). The new feature identifiers are lexicographically sorted and the obtained tuple represents a canonical graph signature of the pharmacophore quadruplet which encodes its content and topology.

Stereoconfiguration encoding

All quadruplets can be divided into five classes based on the canonical feature identifiers (Figure 35) determined on the previous step. Capital letters below denote distinct feature labels and do not designate particular feature types.

a) AAAA system where all features have identical canonical identifiers. This means that four features have identical labels and pairwise binned distances (features create a regular tetrahedron). A quadruplet belonging to this system is achiral.
b) AAAB system where three features have identical canonical identifiers (A) and one feature has a different one (B). This system corresponds to trigonal pyramid and is achiral.
c) AABC system where two features have identical canonical identifiers (A) and two features have different ones (B and C). This system is achiral because there is a plane of symmetry going through B and C features and the center of AA distance.
d) AABB system where pairs of features have identical canonical identifiers (A and B). This system can be chiral or achiral depending on distances between pairs of vertices. The achiral one would have a plane of symmetry whereas the chiral one represents the case of axial chirality.
e) ABCD system where all features have distinct canonical identifiers. This system is chiral.

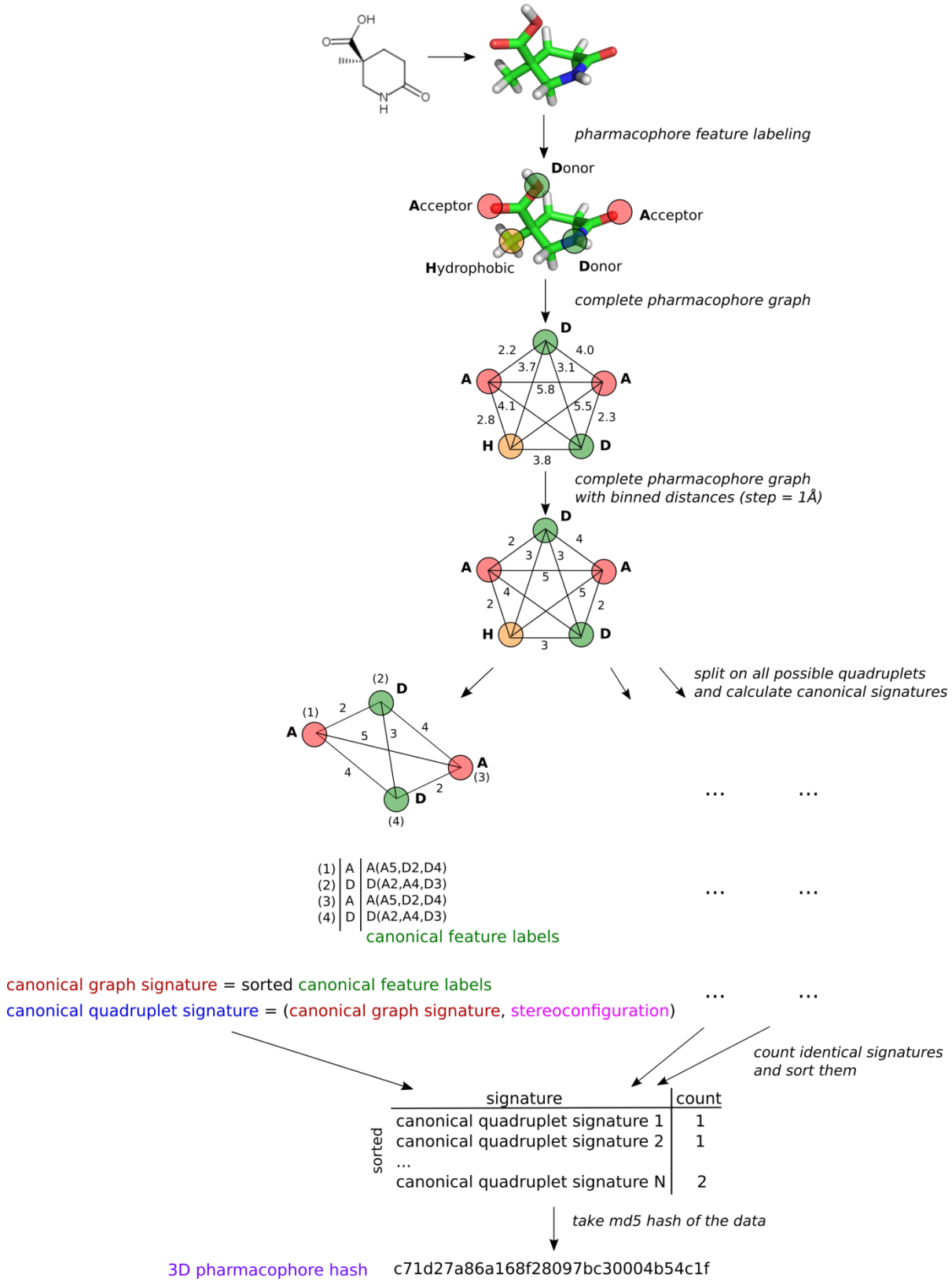


Figure 34. An example of generation of 3D pharmacophore representation.

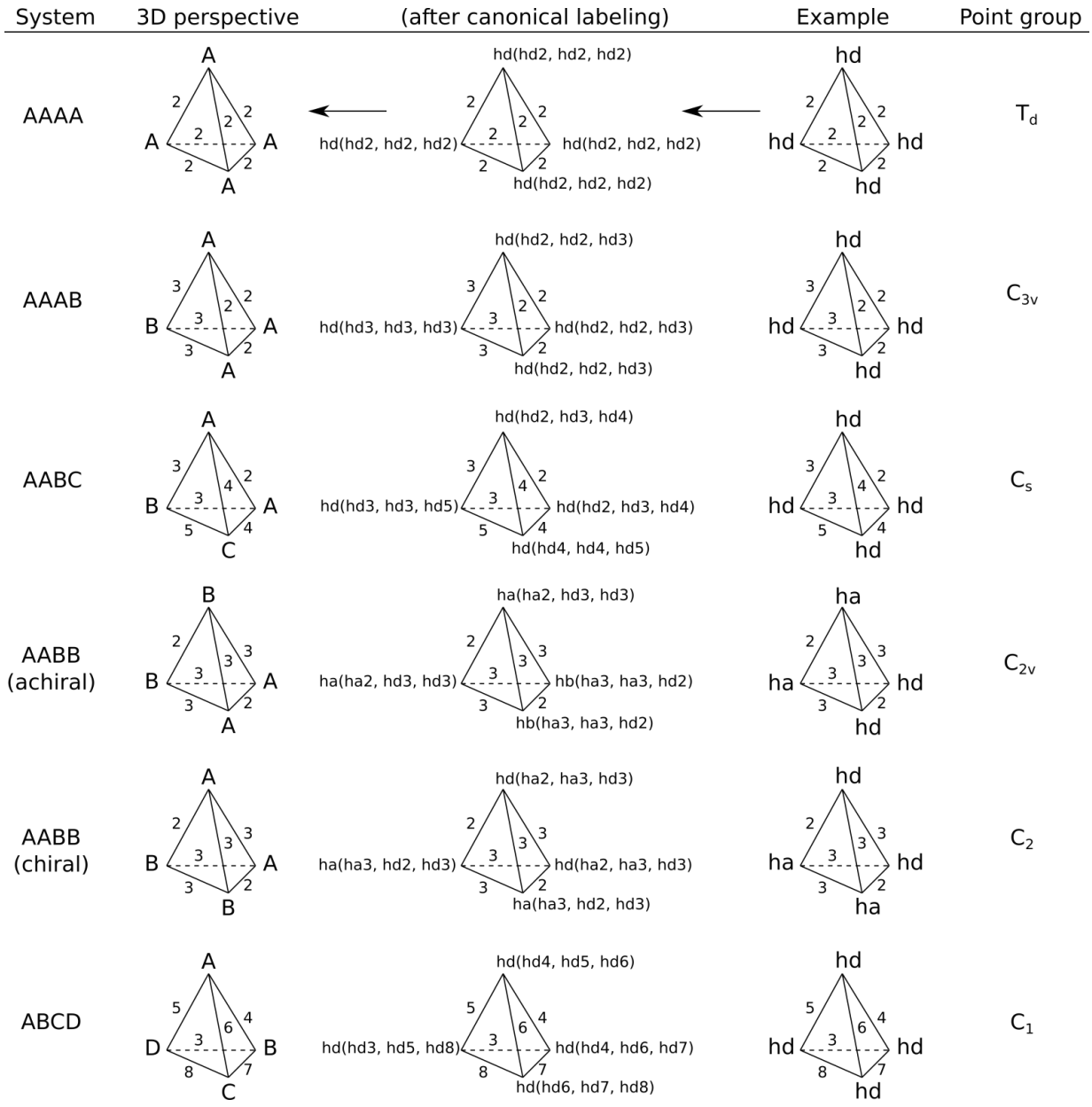


Figure 35. Basic chirality systems. Labels A, B, C and D designate distinct canonical feature labels, not a particular type of a pharmacophore feature. Labels “ha” and “hd” designate hydrogen-bond acceptor and donor features, correspondingly. Numbers on edges designate binned distances between features.

The general workflow to determine configuration sign is depicted in Figure 36. All quadruplets belonging to AAAA, AAAB or AABC systems are assigned configuration sign 0. Quadruplets belonging to AABB and ABCD classes can be achiral if all features lie in the same plane. Therefore, first angles between all edges and corresponding planes are calculated for a quadruplet and the minimal angle is choosing as a measure of deviation of a quadruplet from planarity. To tolerate small deviations from planarity quadruplets having minimal angle less that a pre-defined threshold are considered planar and assigned configuration sign 0. For the remaining quadruplets configuration signs (-1 or +1) are assigned based on the sign of the scalar triple product. To calculate scalar triple

product all vertices are ranked by lexicographic ordering of features based on their canonical identifiers. If two features have the same identifiers (in the case of AABB system) then to break tie i) two features having signature A are placed in random order, ii) the feature B having the shortest distance to the first feature A is placed third and the remaining feature B is placed fourth (Figure 36). Four ranked features are represented in 3D space by three vectors connecting the top ranked feature with remaining ones as depicted in Figure 36. The scalar triple product is calculated and its sign determines the configuration of the quadruplet.

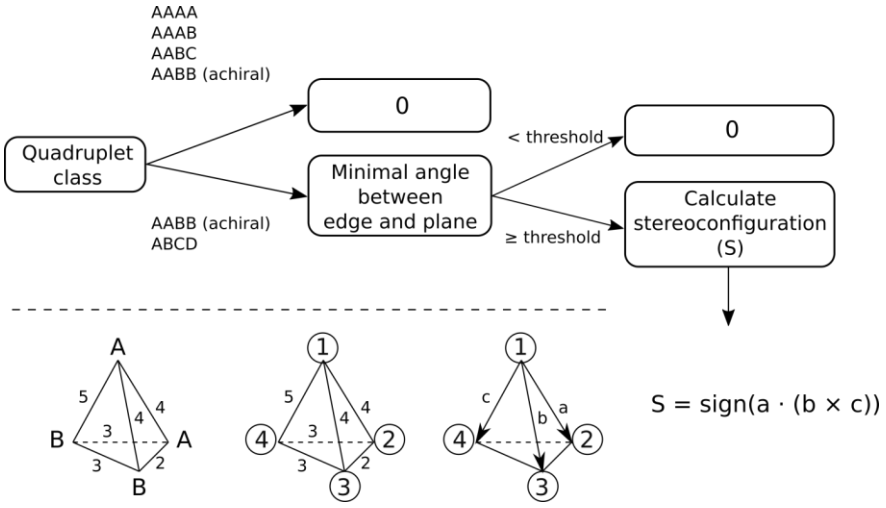


Figure 36. The workflow to determine configuration sign of quadruplets. Labels A and B designate canonical feature identifiers. Numbers on edges designate binned distances. Numbers in circles are ranks of vertices. Labels a, b and c designate vectors in 3D space.

There are special cases of trapeze-like and parallelogram-like quadruplets belonging to AABB class that requires a specific treatment of stereoconfiguration determination (Figure 37). These types of quadruplets are indistinguishable by canonical graph signatures and by the stereoconfiguration sign calculated as described above. Therefore, the calculated stereoconfiguration sign (S) is modified by summing with signum function of cosine of the angle B-A-A-B multiplied by 10 (10 for trapeze-like and -10 for parallelogram-like).

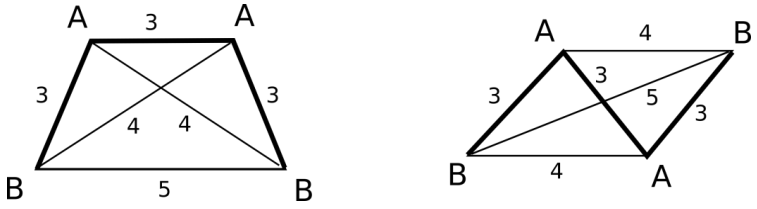


Figure 37. Trapeze-like and parallelogram-like pharmacophore quadruplet belonging to AABB class and not distinguishable by graph canonical signatures.

To represent a pharmacophore one can use binary or count-based representation. In binary representation each distinct quadruplet is hashed and activates a predefined number of bits in a fixed-length bit string. In count-based representation all occurrences of identical quadruplet signatures are summed up and create a feature vector, where each item has its own name and a count. This feature vector can be further converted to MD5 hash to store it more compactly in a database and to be used for fast searching of identical or similar pharmacophores. We assume that identical or very similar pharmacophores within a chosen binning step will have identical hashes (feature vectors). Thus, identification of similar pharmacophores can be as simple as comparison of two hashes.

Bit strings and feature vectors can be used as descriptors for machine learning models. There is also an option to enumerate not only quadruplets but triplets or doublets. In this case the obtained representation will not distinguish enantiomeric pharmacophores, but the number of elements in a feature vector will greatly reduced. The SMARTS patterns used for pharmacophore feature labeling can be fully customized and a user can create own features with new definitions. The described 3D pharmacophore representation was implemented as an open-source Python package *pmapper* - https://github.com/DrrDom/pmapper.

3.2. Automated ligand-based pharmacophore modeling protocol

We implemented a fully automatic pipeline which takes a data set of ligands as input and return a set of pharmacophore models with their validation scores. Its main features are speed, the ability to identify pharmacophores for different binding modes and automatic validation of generated models on the hold-out set of compounds. The pipeline consists of several stages.

Training and test sets formation

A data set of active and inactive compounds can be large and it can be computationally infeasible to use all available compounds for model development. Therefore, a representative subset of active and inactive compounds should be selected for model training.

Two strategies of training set creation were implemented. The first strategy assumes that all active compounds have the same binding mode. Active and inactive compounds are clustered separately using Butina clustering[97] and 2D pharmacophore triplets both implemented in RDKit[21] (Figure 38). Centroids of each cluster of active and inactive compounds having at least 5 compounds are selected to the training set. The number of compounds in the training set depends on the number of clusters which can be tuned by selection of different clustering cutoff values. All remaining compounds form the test set for external validation.

The second strategy assumes that active compounds have different binding modes. In this case active and inactive compounds are clustered jointly using Butina clustering and 2D pharmacophore triplets. We assume that compounds from an individual cluster may have similar binding modes

relatively to compounds from other clusters. From each cluster 5 active and 5 inactive compounds are randomly chosen and form the individual training set. Clusters containing less than 5 active compounds are ignored. Centroids of clusters obtained from clustering of only inactive compounds (similarly to the first strategy) are added to each training set in order to better represent inactive compounds (Figure 38). Duplicated inactive compounds in each training set are removed. Multiple training sets are created within this strategy and all of them are used for development of individual models. All compounds not included in a particular training set create a complementary test set which is used for external validation of selected models.

As a result a single training set is created within the strategy I and multiple training sets within the strategy II.

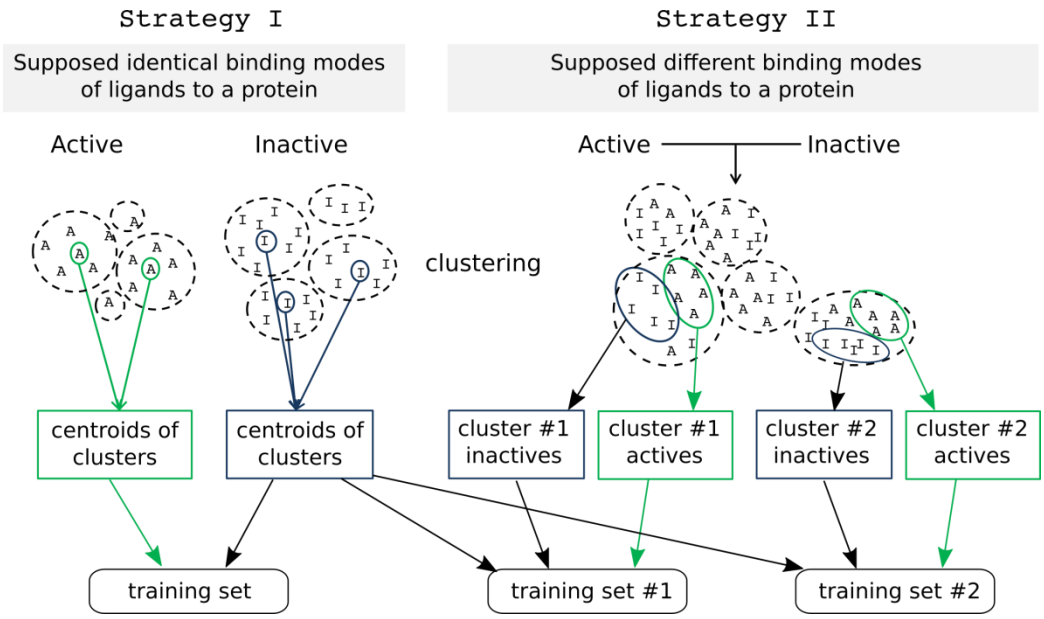


Figure 38. Two strategies of training set compound selection.

Model development and selection

For all compounds we enumerate all possible stereoisomers and generate up to 100 conformers by RDKit and optimize them in MMFF94 force field[98] (Figure 39). Models are generated iteratively from the simplest ones to more complex. The main underlying assumption is that identical 3D pharmacophore hashes correspond to similar 3D pharmacophores. Thus, we do not need to superpose 3D pharmacophores to identify whether they match each other or not as it is performed in all other pharmacophore modeling approaches[86].

At the beginning 3D pharmacophore hashes of all possible 4-point pharmacophores are calculated for training set compounds. Duplicated hashes obtained for the same compound are removed. Occurrence of hashes among active and inactive compounds is calculated followed by calculation of F-score for every 4-point pharmacophore hypothesis. For the training set formed

within the strategy I the pharmacophore models having F0.5 score greater or equal 0.8 are selected for the next iteration. This will focus on selection of more precise models rather than those ones which cover a larger number of active compounds. For the training set formed within the strategy II the pharmacophore models having F2 score equal to 1 are selected. If there are no such models than models having recall equal to 1 are selected. A different criterion for the strategy II was chosen because training sets have a smaller number of active compounds and it is assumed that these compounds have the same or similar binding modes, therefore it would be expected to find the model which covers all active compounds from the training set. Top 100 models for each strategy are selected independently for the next iteration.

On the next iteration 5-point pharmacophores are generated adding one feature to the selected 4-point pharmacophores. Hashes and their occurrences are calculated again and the best performing models are selected for the next iteration. This procedure continues until generated pharmacophores meet the abovementioned criteria. If there are no pharmacophore models that satisfy criteria after the current iteration, the models selected on the previous iteration are selected as final ones and are validated on an external test set (Figure 39). The described procedure generates the most complex pharmacophores which match preferably active compounds in the training set and avoid matching inactive ones.

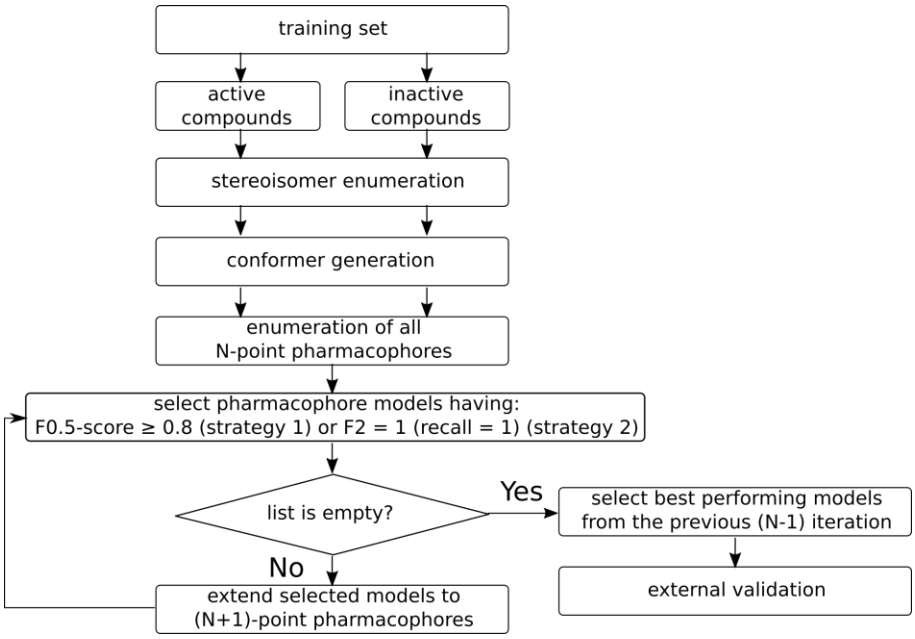


Figure 39. Overall workflow of pharmacophore model generation.

Virtual screening

To speed up screening it is performed in several steps that include fingerprint screening, isomorphic embedding and hash comparison. On the first step a hashed fingerprint of a query pharmacophore and dataset molecules' pharmacophores are generated by *pmapper*. Fingerprints are used as a Bloom filter to quickly discard irrelevant pharmacophores: a candidate molecule is only

relevant for further screening if activated bits in a query pharmacophore fingerprint are the subset of activated bits of the compound.

On the next step topology and content of a pharmacophore model is compared with a candidate molecule pharmacophore. A pharmacophore represented by a complete graph with binned distances is checked to be a subgraph (using VF2 subgraph isomorphism algorithm) of a candidate molecule pharmacophore. On the last step 3D pharmacophore hashes of a query pharmacophore model and corresponding subgraphs of candidate pharmacophores are compared for identity in order to determine whether they have identical topology and stereoconfiguration.

Case study

For evaluation of the developed pipeline we collected three data sets of inhibitors of acetylcholinestarase (AChE), inhibitors of cytochrome P450 3A4 (CYP450 3A4) and antagonists of adenosine 2a receptor (A2a) from ChEMBL database[99] (Table 9). These targets were chosen because there were enough amount of data about active and inactive compounds for model development and there were many 3D protein-ligand complexes in Protein Data Bank for validation of obtained models.

Table 9. Data sets used for ligand-based pharmacophore modeling.

Data set	Number of actives	Number of inactives	Total number of compounds
AChE	176 ($pIC_{50} \geq 8$)	1070 ($pIC_{50} \leq 6$)	1246
CYP450 3A4	138 ($pIC_{50} \geq 7$)	548 ($pIC_{50} \leq 5$)	686
A2a	293 (pK_i/pK_d/$pIC_{50} \geq 7$)	279 (pK_i/pK_d/$pIC_{50} \leq 5$)	574

We did not observe large difference for different tolerance values (0, 5 and 10 degrees) which should help to tolerate deviation of quadruplets from planarity and used the default value 0 for the further study. We built pharmacophore models using different clustering threshold (0.3, 0.4, 0.5) and both strategies I and II. We also combined hit lists of individual models into a consensus prediction. All results were obtained for test sets which were not used for model training.

For comparison purposes we used similarity search based on 2D pharmacophore fingerprints implemented in RDKit. Each active compound from a whole data set was used as a query and all remaining compounds were ranked according to Tanimoto similarity to that reference to build ROC curve and calculate area under curve (AUC) value. A compound with the highest AUC value was selected from each data set for comparison with output of pharmacophore models. This is the most rigorous comparison scenario.

Performance of individual models built using strategy I and II were similar. However, the former sometimes had better recall values (the percentage of known active compounds retrieved by a model) and the latter had better precision (the percentage of actives among all compounds retrieved

by a model) (Figure 40). In many cases we observed early enrichment. Consensus models outperformed individual pharmacophore models obtained by both strategies in terms of recall values; however, precision of consensus models was poorer than for individual ones.

In many cases model performance was very close or identical to the best similarity search results (Figure 40). In some cases, e.g. for the CYP450 3A4 data set, performance of individual models obtained with the strategy I was better than similarity search (points are above ROC). However, models obtained with the strategy II had performance similar or slightly worse than the best similarity search results (points lay on a ROC or below). This can be explained by the fact that these models (strategy II) were trained on more congeneric subsets of compounds representing a smaller subspace of available ligands relatively to models trained on a diverse subset of compounds (strategy I). Consensus models outperformed similarity search results for the CYP450 3A4 and AChE data sets.

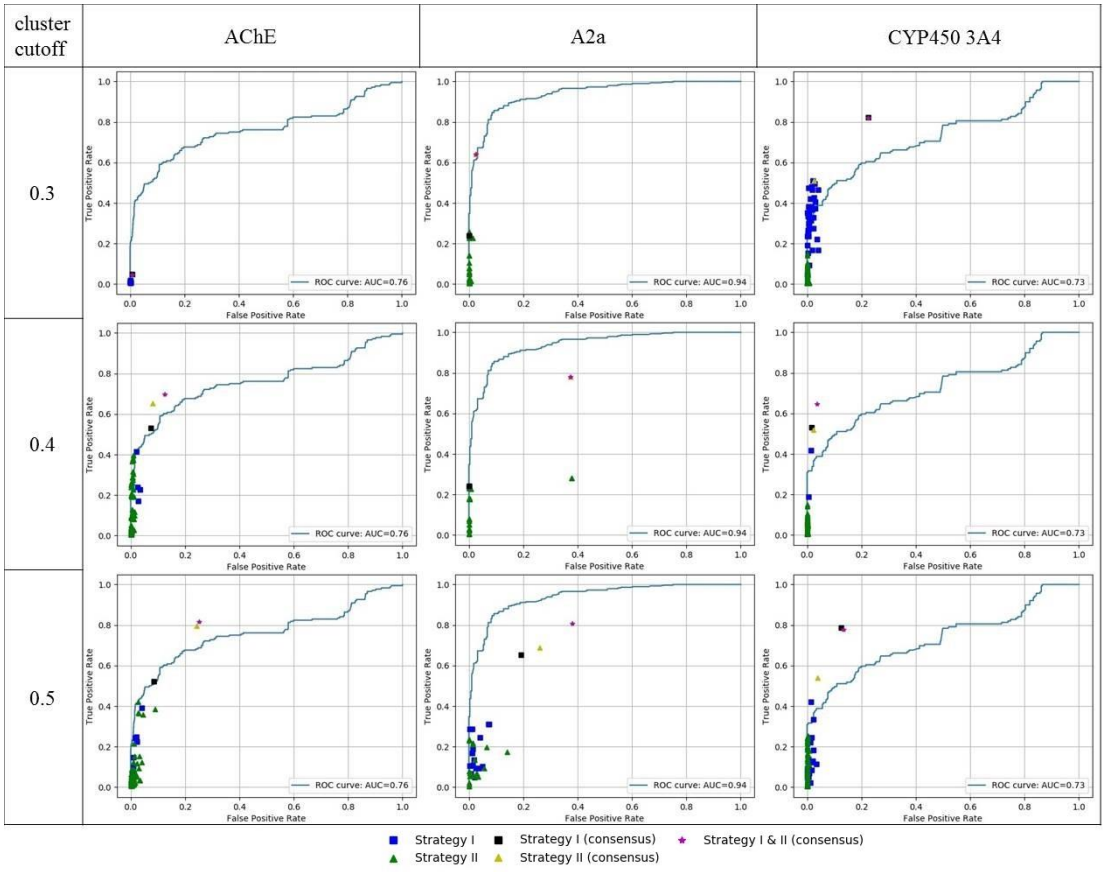


Figure 40. Performance of 3D ligand-based pharmacophore models and comparison with the best results of similarity search based on 2D pharmacophore fingerprints.

To estimate the validity of the obtained 3D ligand-based pharmacophores we used them to screen 3D poses of available ligands taken from ligand-protein PDB complexes. In total 9 antagonists of adenosine 2a receptors, 10 inhibitors of AChE and 26 inhibitors of CYP450 3A4 which were not present in the corresponding sets used to train the models were collected from PDB. 5 antagonists of adenosine 2a receptors, 2 inhibitors of AChE, 3 inhibitors of CYP540 3A4 were

found by at least one model. Donepezil, AChE inhibitor, fits the 3D pharmacophore model which contains features corresponding to observed ligand-protein interactions. Two benzene rings are involved in interaction with Trp86 and Trp186 and the carbonyl oxygen forms H-bond with the backbone of Phe295 residue (Figure 41). The hydrophobic feature of one of CYP450 3A4 pharmacophore models matches the phenyl group of ritonavir which fits the hydrophobic pocket of the protein formed mainly by phenylalanine and leucine residues. Another hydrophobic feature and H-bond acceptor match the thiazolyl ring which coordinates the heme. The hydroxyl group matching H-bond donor and acceptor features forms H-bond with the Ser119 side chain (Figure 41). The A2a pharmacophore model matched most of observed ligand-protein interactions for the ligand in the 5OLZ complex. H-bond donor matches the amino group which forms H-bonds with Glu169 and Asn253. H-bond acceptor features match two nitrogen ring atoms. One of them forms H-bond with Asn253 and another one with Glu159 through a water molecule. Phenyl group matching a hydrophobic feature of the pharmacophore is in the pocket close to Met177, Trp246 and Leu249 (Figure 41).

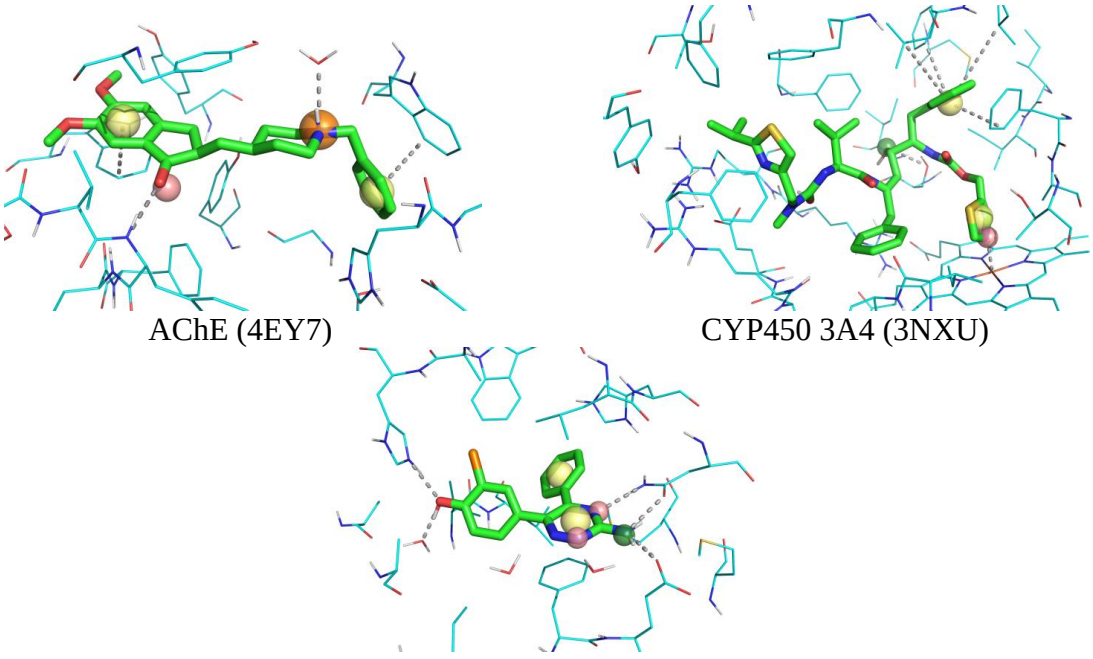


Figure 41. Examples of compounds matching corresponding 3D pharmacophore models developed in this study. Red sphere denotes H-bond acceptor, yellow sphere – hydrophobic/aromatic feature, orange sphere – H-bond donor and positively charged center, green sphere – H-bond donor and H-bond acceptor feature.

It is interesting to note that in all cases features of ligand-based pharmacophore models matched observed protein-ligand interactions. This means that even in the absence of structural information about a protein ligand-based models can reveal features responsible for binding and receptor recognition. This result additionally confirms validity of the developed pipeline. This

success is also linked to the RDKit conformer generator which generated relevant conformers that helped to identify key features and their spatial position. The tool *psearch* is open-source and available as a Python package - https://github.com/meddwl/psearch.

3.3. Pharmacophore modeling based on molecular dynamics of protein-ligand complexes

Molecular dynamics (MD) simulations can capture a wide variety of important biomolecular processes, including conformational changes, ligand binding, and protein folding. These simulations can predict how biomolecules will respond to perturbations such as mutation, phosphorylation, protonation, or the addition or removal of a ligand[100-102]. The major prerequisites for emerging of MD simulations of drug discovery were rapid increase in the number of experimental data about structures of proteins and protein-ligand complexes and wide accessibility to high-performance computing. Many approaches were developed to identify and refine pharmacophore centers in apo-proteins through identification of hot spots and hydration sites[92, 103-105]. Approaches for ensemble pharmacophore modeling based on MD simulations were suggested. One of them is Common Hit Approach (CHA)[106] which demonstrated good performance in virtual screening. Molecules were ranked according to the number of matched representative pharmacophore models retrieved from an MD trajectory of a protein-ligand complex. Ensemble modeling frequently outperformed individual models retrieved from X-ray structures of protein-ligand complexes and models occurred most frequently during MD simulation. This suggests high importance of considering all variety of pharmacophore models.

One of the issues of MD pharmacophore modeling is the very large number of models which are retrieved from every snapshot of an MD trajectory. Therefore, it is needed to rationally select a representative subset MD pharmacophores to facilitate their analysis and processing. Within CHA the authors retrieved 20 000 pharmacophore models. To select representative pharmacophores they grouped all models according to the number and types of pharmacophore features. The energy of ligand conformations corresponding to each pharmacophore model was calculated with the Merck Molecular Force Field (MMFF). A conformer with median energy was identified within each group and the corresponding pharmacophore model was selected as representative. The spatial arrangement of features was ignored because pharmacophore models were grouped only by the type and the number of pharmacophore features. Therefore, dissimilar pharmacophores with different geometry but the same set of features can get to the same group, which will not correspond to a single representative model.

We suggested to apply the previously developed 3D pharmacophore hashes to quickly remove identical or similar models from large sets of 3D pharmacophores[107]. We retrieved all snapshots from an MD trajectory using MDTraj[108], for every snapshot we got a pharmacophore model using PLIP[109]

which was encoded by a 3D pharmacophore hash using *pmapper*[94] and kept only models with distinct pharmacophore hashes. This resulted in a representative subset of pharmacophore models in a single pass (Figure 42).

We also suggested a new scoring scheme for ligands, Conformer Coverage Approach (CCA), based on multiple pharmacophore models retrieved from MD simulations. Within CCA compounds are ranked according to the percentage of conformers matching at least one representative pharmacophore model (Figure 42). We assumed that compounds whose conformers fit more frequently to pharmacophores observed within MD simulations of a protein-ligand complex may have more favorable binding due to the less decrease of binding entropy. In the ideal case if all conformers of a ligand match some of observed MD pharmacophores, it means that flexibility of a ligand matches flexibility of a protein very well and ligand should not lose too many conformational degrees of freedom upon binding. Of course, this is a simplification, but it can be valid to certain extent.

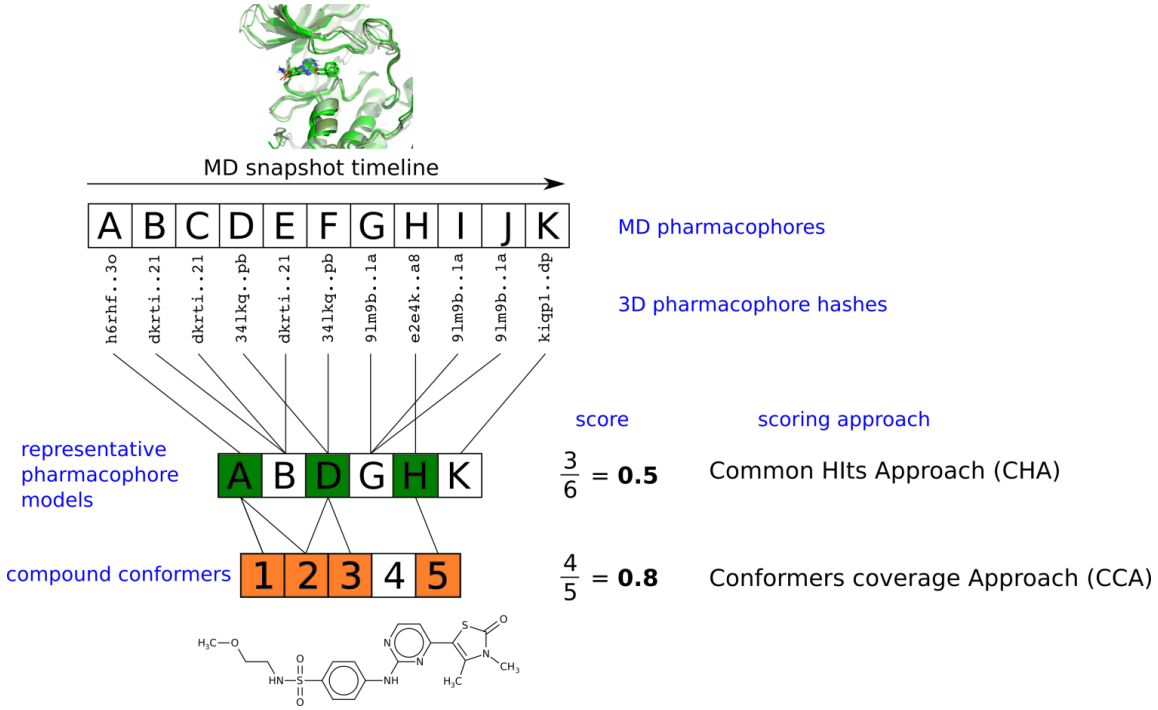
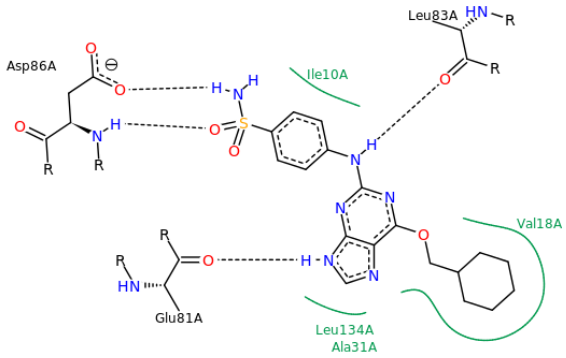


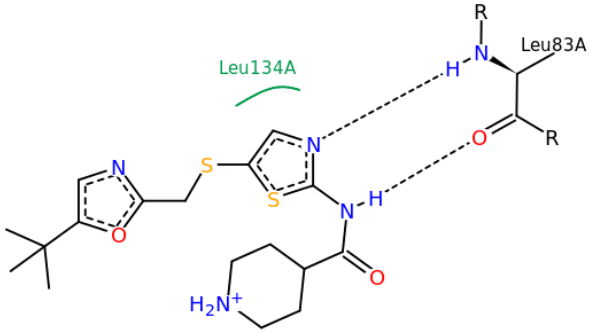
Figure 42. Compound scoring schemes based on the proposed Conformers Coverage Approach and the previously developed Common Hits Approach. Distinct representative pharmacophore models were selected among all MD pharmacophores based on their 3D pharmacophore hashes.

For validation of the suggested approach we selected four complexes of cyclin-dependent kinase 2 (CDK2) and its ligands (PBD: 2C6O, 2FVD, 2XMY, 5D1J) (Figure 43). MD simulations were performed as described in the protocol[107]: structures were minimized and equilibrated in explicit solvent and production simulations were performed for 50 ns at 310 K. To evaluate performance of virtual screening we used the data set of known inhibitors and decoys from DUD-E[110]. After a thorough check all duplicates were removed from the validation set. Also ligands

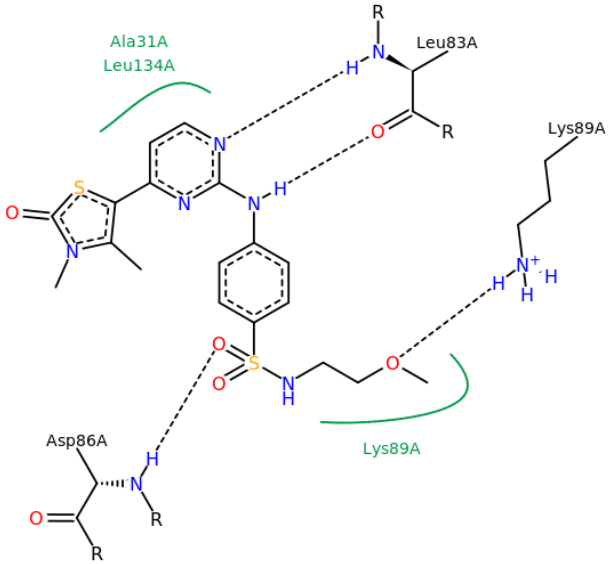
presented in selected four complexes were removed from the DUD-E data set to avoid the overestimation of model performance. The final data set contained 473 active compounds and 27853 decoys. For compounds with an undefined configuration of stereocenters or double bonds all possible stereoisomers were enumerated. For each stereoisomer of a compound at most 100 conformers were generated using RDKit and optimized in MMFF. Conformers with RMSD less than 0.5Å were discarded as redundant.



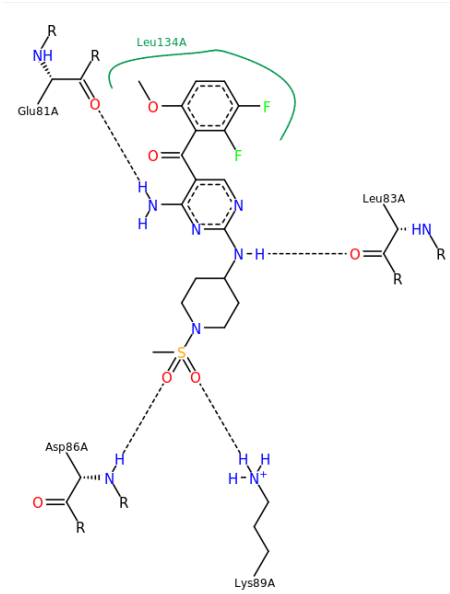
IC_{50} = 5 - 8.1 nM [111-113]
2C6O



IC_{50} = 38 - 46 nM [114, 115]
5D1J



K_i = 0.11 nM [116]
2XMY



K_i = 3 nM [117]
2FVD

Figure 43. Protein-ligand interaction charts of four selected CDK2 complexes.

2500 frames were extracted from each MD trajectory of four complexes and the corresponding number of structure-based pharmacophore models was derived. 3D pharmacophore hashes were calculated for each pharmacophore to identify highly similar ones. By design, the pharmacophores with identical hashes should have root mean squared distance (RMSD) within the chosen binning step. In order to verify this hypothesis, we aligned all pairs of pharmacophore models with identical sets of features and calculated best fit RMSD values. As expected, pharmacophores having identical hashes have a distribution of RMSD values from 0 to 0.93Å across all four protein targets whereas

RMSD values for pairs of pharmacophores having different hashes were distributed in a wider range, from 0.01 to 4.96Å (Figure 44). This indicates an important feature - identical 3D pharmacophore hashes always correspond to similar pharmacophores but not always similar pharmacophores have identical hashes. This means that by removing pharmacophores with identical hashes we achieve the main purpose to reduce the number of redundant pharmacophores although keeping some redundancy among remaining representative pharmacophores.

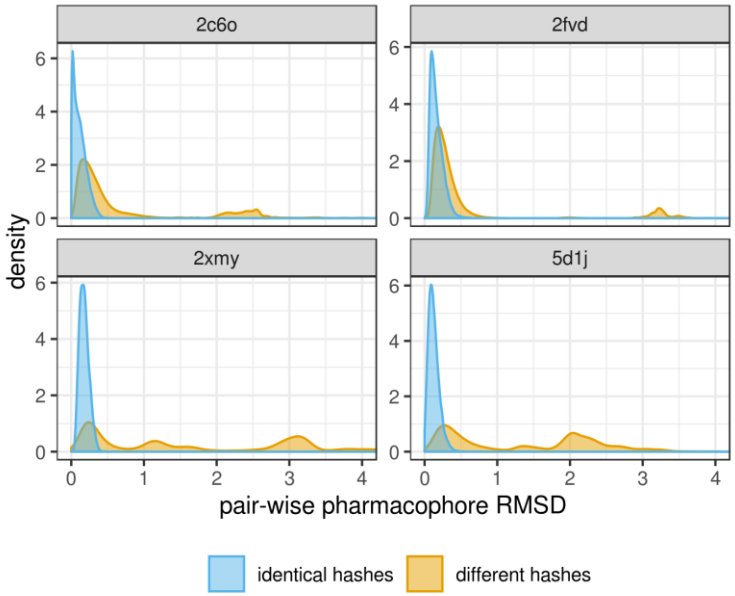


Figure 44. Gaussian kernel density of distribution of root mean squared deviation values for the best fit between pairs of pharmacophores with identical and different hashes.

Elimination of pharmacophores with duplicated hashes substantially reduced the number of pharmacophores for 2C6O, 2FVD and 5D1J targets to 13.5%, 17.6%, and 27.3%, correspondingly. The pharmacophores retrieved for 2XMY target were the most diverse and the number of distinct pharmacophore hashes was high, 80.3%. This can be explained by higher flexibility of the 2XMY ligand and more complex pharmacophore models for 2XMY with a greater number of features than pharmacophores for other complexes. The maximum number of features in a pharmacophore was 10 for 2XMY and 7 for other complexes.

We used ensembles of models having at least 4 pharmacophore features to compare two ranking schemes: CHA and CCA, because simpler models were too promiscuous. Almost in all cases CCA demonstrated higher early enrichment factors than CHA (Figure 45). For example, for ensembles consisting of at least five-feature models enrichment at 0.25% was 6.27 and 10.25 (2C6O), 4.98 and 10.5 (2FVD), 22.7 and 35.0 (2XMY), 4.64 and 4.23 (5D1J) for CHA and CCA, correspondingly.

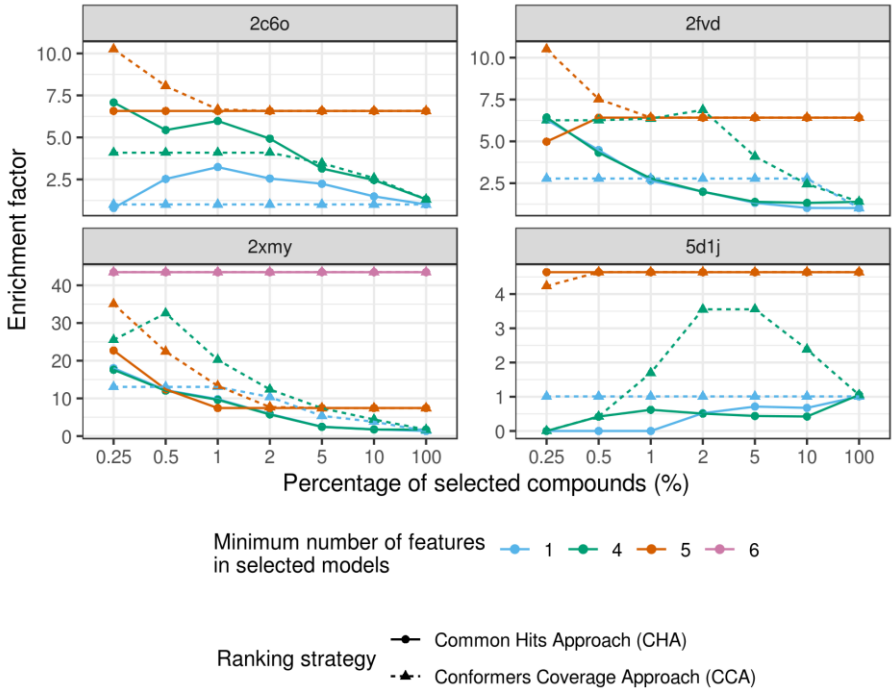


Figure 45. Enrichment factor for two ranking strategies at different complexity of selected models.

Since we had several complexes of the same protein we tested a consensus approach. Compounds were ranked in the descendant order according to their average CCA scores calculated for different protein targets. We used only pharmacophore ensembles including models with at least 4 and 5 features because simpler models resulted in poor performance and more complex models were not available for all studied complexes. The consensus of four complexes demonstrated good performance with $EF_{0.25\%}$ being 24.8 and 22.1 for 4- and 5-point pharmacophores, respectively (Figure 46). Such high performance was mainly determined by high performance of the set of pharmacophore models extracted from the MD trajectory of 2XMY complex and the consensus ranking based on four complexes did not outperform the one for 2XMY complex. However, we consider this as a good result because it demonstrates the power of consensus ranking and encourages the application of consensus ranking whenever possible because this decreases a bias introduced by individual model ensembles and gives more robust output.

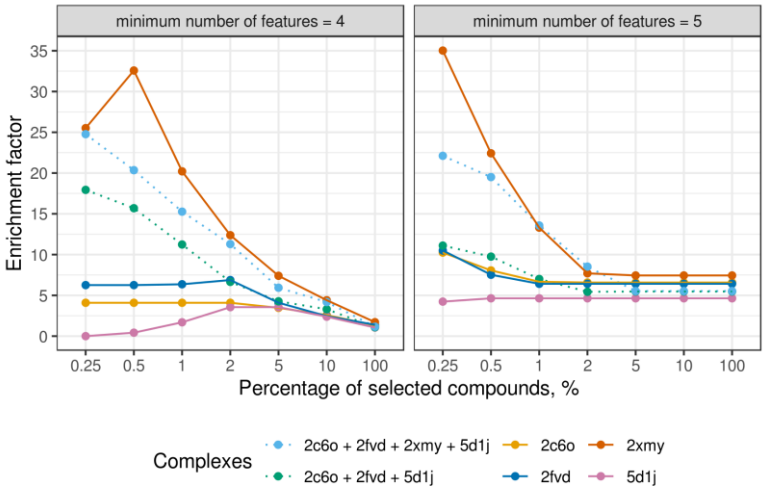


Figure 46. Enrichment factors for single pharmacophore ensembles and for consensus predictions made by averaging of scores of single compounds calculated for individual model ensembles within Conformers Coverage Approach.

The tool *pharmd* is open-source and provided as a Python package - https://github.com/ci-lab-cz/pharmd.

3.4. Summary

We implemented a representation of 3D pharmacophore, which is very flexible and may find applications in different studies. In particular, 3D pharmacophore hashes demonstrated their applicability in ligand-based an MD-based pharmacophore modeling for identification of similar pharmacophores. For ensembles of MD pharmacophores we suggested a new scoring, Conformer Coverage Approach, which showed comparable or better performance than Common Hits Approach. All developments were implemented as open-source software to support further researches in the field of pharmacophore modeling.

Chapter 4. Multi instance modeling as a response to the complexity of modeling chemical systems

Molecules are inherently complex objects. They may exist in different protonation and tautomeric forms, conformational states, and configurations in equilibrium. Which form and a state are the most important and relevant for a particular response is often unclear. Commonly used machine learning methods cannot handle this complexity naturally. They require a single molecular representation converting into a single vector of descriptors which is correlated with a property of a compound. In other words, a compound should be represented by a single instance. This is often achieved by certain simplifications. In 2D modeling structures of molecules are represented as molecular graphs with a fixed tautomeric form and a protonation state which are chosen based on some rules or other models[118]. In 3D modeling usually the lowest energy conformer is used[119] that is not always relevant, because "bioactive" conformation may differ from the lowest energy computed in vacuum. Using of non-relevant conformers for 3D modeling may result in poor predictive performance of obtained models. Therefore, 2D modeling approaches frequently outperform 3D ones[120, 121]. To solve this issue certain efforts were made to represent compounds as ensembles of conformers, so called 4D modeling[122]. Descriptors calculated for individual conformers (instances) were usually averaged to get a single vector of descriptor representing a compound which can be processed by conventional machine learning approaches. Additionally the standard deviation, median values or higher moments, i.e., skewness and kurtosis, can be concatenated to a descriptor vector as additional features. However, such increase in complexity of the molecular representation does not often outperform 2D models[123]. A possible explanation for this is that simple protocols for descriptor aggregation were used. Simple averaging or averaging weighted by Boltzmann distribution in vacuum may be not the optimal strategy, because in the former case all conformers are considered equally important and in the latter case distribution in vacuum and in biological medium may differ.

To solve the issue of multiple representations of a single compound we revisited the multi-instance learning (MIL) approach, which was originally suggested to solve chemical problems with multiple conformers[124]. It did not find wide applicability in chemoinformatics because 4D modeling approaches were actively developed at the same time and attracted more attention due to their simplicity relatively to MIL, which required special algorithms. However, development of MIL was continued and it was successfully applied for other tasks: classification of text documents (information retrieval), classification of images (computer vision), speaker identification (signal processing), bankruptcy prediction (economy), etc[125, 126]. Recently with emerging of neural networks more sophisticated MIL algorithms were introduced which increased accuracy of models and provided additional benefits, such as identification of key instances. In particular, the latter feature of MIL models can be used to identify probable "bioactive" conformers for compounds without prior

knowledge about a structure of a target protein. We adapted the most popular MIL approaches and applied them to model biological properties of compounds[127, 128] and catalyst enantioselectivity[129, 130].

4.1. Multi-instance modeling approaches

The key feature of MIL approach is representation of a molecule by a bag (set) of instances (e.g. conformers) each encoded by its own descriptor vector and where a property label (e.g. bioactivity) is known for a compound but not for individual instances (conformers). The goal is to establish the correlation between property values and bags of descriptor vectors.

All MIL algorithms can be divided on two large groups: instance-based and bag-based methods[131]. Instance-based algorithms consider each conformer as a separate training instance. Bag-based algorithms, on the contrary, represent a molecule by a single vector of descriptors, which is produced from the vectors of descriptors of individual conformers.

The simplest instance-based MIL algorithm is *Instance-Wrapper*, where each training instance of a bag is assigned the same label as for the whole bag. This means, for example, that if a molecule is bioactive, it is assumed that all its conformations are bioactive. As a result, one gets a data set where each conformation is an individual training object and any conventional ML algorithms can be applied to build a model. Given a new molecule, the bioactivity is predicted for each conformation, and predictions are averaged to get the final predicted bioactivity of the molecule (Figure 47a). This approach has an obvious drawback because assigning the same bioactivity to all conformations of a molecule in a training set can bring some noise into the learning process because the fact that a molecule is bioactive does not mean that all its conformations are biologically relevant and responsible for protein-ligand recognition.

In bag-based algorithms there is no need to identify a label for each instance in a bag. Instead, there is an operation that aggregates the instances to get a single vector representing the entire bag. The simplest implementation of the *Bag-Wrapper* algorithm averages descriptor values across all conformations and supply this single vector of descriptors to a conventional single instance machine learning (SIL) method (Figure 47b). The *Bag-Wrapper* algorithm has a similar to *Instance-Wrapper* drawback because while aggregating the descriptor vectors of all conformations the resultant vector may be noised by the contribution of irrelevant conformations. The *Bag-Wrapper* is identical to commonly used averaging of descriptors of conformers in 4D QSAR modeling. Thus, such 4D QSAR approaches can be considered as a special case of MIL approaches.

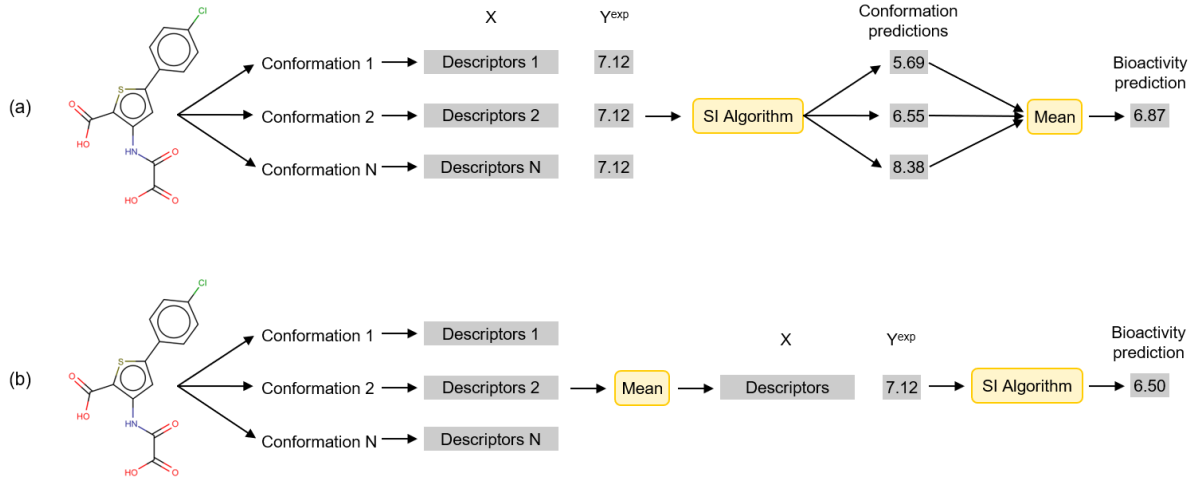


Figure 47. Multi-instance wrapper algorithms: (a) Instance-Wrapper and (b) Bag-Wrapper. SI is a single-instance machine learning algorithm.

Multi-instance neural networks learn in an end-to-end way and take a bag of instances as input and directly output bag prediction. All parameters in MIL networks are optimized via back-propagation. Wang et al.[132] revisited MIL neural networks and proposed a series of novel neural network frameworks. They considered two types of neural networks: mi-Net (hereafter *Instance-Net*) and MI-Net (hereafter *Bag-Net*). In *Instance-Net* (Figure 48a) instances are running through fully-connected layers and an output neuron. Then, instance predictions are averaged in the pooling layer to obtain a bag prediction, its error is calculated and backpropagated to adjust model weights. *Bag-Net* (Figure 48b) consists of fully-connected layers followed by one pooling layer. The pooling layer averages instance representations learned by previous layers into a single embedding vector as a bag representation. The last fully-connected layer takes the embedding vector as input and outputs the bag prediction. Wang et al.[132] examined three typical pooling operators - max pooling, mean pooling, and log-sum-exp pooling and concluded that all of them provided a similar performance on benchmark data sets.

The *Bag-Net* uses a mean pooling function and as was mentioned above the irrelevant conformations can contribute noise to the prediction and reduce model performance. This drawback can be avoided by using more flexible types of pooling, such as weighted averaging pooling, known as attention. This type of pooling was proposed by Ilse et al[133], who used an additional two-layered neural network to obtain weights of instances. In the *Bag-AttentionNet* (Figure 48c), all instances are first fed to the fully-connected layers. Then, the learned instance representations are used by the attention network with a single hidden layer. In the attention network, the number of output neurons is equal to the number of instances. The output layer of attention has the Softmax activation function and predicts instance weights. Finally, the instance weights given by the attention network are used for weighted averaging of instance representations to get the embedding vector that is used to

produce the bag prediction. Using of weighted pooling enables the *Bag-AttentionNet* to automatically identify probable bioactive conformations.

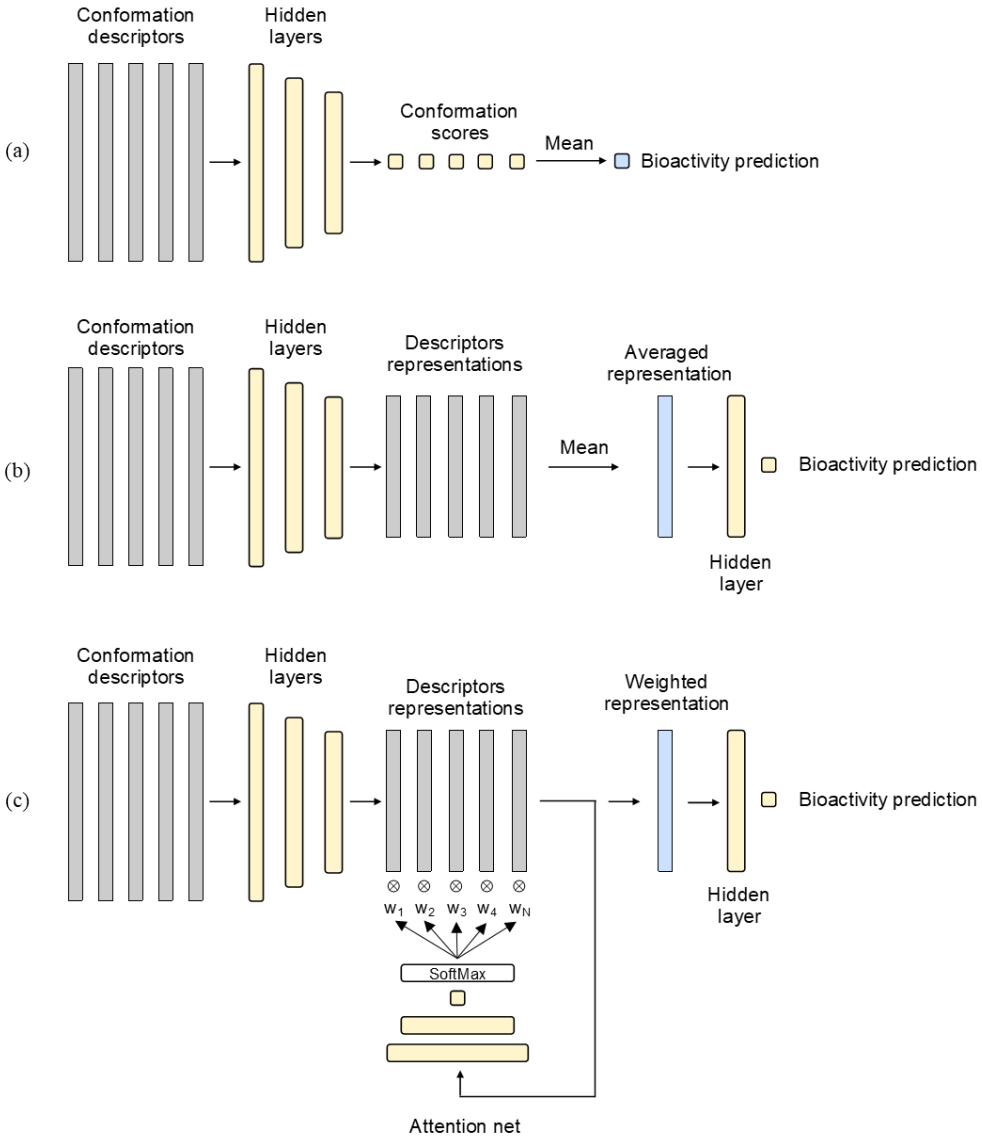


Figure 48. Multi-instance neural networks: (a) *Instance-Net*; (b) *Bag-Net*; (c) *Bag-AttentionNet.*

4.2. Modeling of biological activity of chemical compounds

We applied the implemented approaches to predict bioactivity on a large set of targets. We collected 175 data sets of compounds with measured pK_i or pIC_{50} values extracted from the ChEMBL23 database. The size of the data sets varied from several hundred to several thousand compounds. Molecules with a molecular weight greater than 700 (3 % of the total number of molecules) were discarded. Because the performance of 3D models may depend on the flexibility of studied compounds, the average number of rotatable bonds for molecules in each data set was calculated using RDKit. Most molecules in data sets can be considered as low to moderately flexible with the average number of rotatable bonds within 3-6. Data sets were split on modeling (80%) and

test sets (20%). The modeling sets were split on training (80%) and validation ones (20%) to tune hyperparameters of models.

We generated up to 100 conformers for each molecule using the algorithm implemented in RDKit, which is known to be able to reproduce bioactive conformations observed for ligands in PDB complexes with reasonable accuracy[134]. This ability is important because it may improve the performance of obtained models, make them more reasonable, and in the case of MIL modeling approaches would increase the probability of identifying the most relevant conformations.

Additionally, in the collected data sets, we identified compounds deposited in the Protein Data Bank (PDB) and retrieved their conformations. These PDB conformations were used as references to compare with the conformations predicted by MIL models as the most probable biologically relevant ones.

To encode molecules for 3D modeling we used the previously developed 3D pharmacophore descriptors calculated by *pmapper*[94]. Each conformation was represented by a set of counts of identical 3D pharmacophore quadruplets calculated with the binning step 1Å. Since the descriptor matrix was very sparse we discarded those quadruplets which occurred less than 5% among all conformations of a data set.

As reference 2D models we chose binary Morgan fingerprints (MorganFP) of radius 2 and size 2048 calculated with RDKit because they are widely used and demonstrated high performance in previous benchmarking studies[135]. For comparative purposes we also used 2D physicochemical descriptors (PhysChem) and binary 2D pharmacophore fingerprints (PharmFP) calculated with RDKit. The former included EState indexes, the number of different pharmacophore features, rings systems, functional groups and fragments, etc. To calculate 2D pharmacophore descriptors we used the same definitions of pharmacophore features as in *pmapper* to make comparison more robust. Afterwards, pharmacophore triplets were enumerated using default binning of topological distances (0-2, 2-5, 5-8, 8+).

As a single-instance machine learning algorithm for *Instance-Wrapper* and *Bag-Wrapper* we used a three layer neural network. The same architecture was used for building models based on 2D descriptors.

Comparison of multi-instance approaches

For 45 data sets out of 175, no MI models achieved the required performance of $R^2_{test} > 0.4$. These “non-modellable” datasets were excluded from the further comparison. We performed a pairwise comparison of models using the Wilcoxon-Holm test with a significance level of 5%. Results of pairwise comparison of models were visualized with a critical difference diagram[136] (Figure 49). *Instance-Wrapper* outperformed all other algorithm. It was better than *Bag-Wrapper* indicating that considering of individual instances is better than aggregate them into a single feature

vector. Neural network-based approaches performed similarly to *Bag-Wrapper*. Even the most advanced *Bag-AttentionNet* was not statistically better than *Bag-Wrapper*. Currently we do not have an answer on the question why Instance-Wrapper was so successful and outperformed other models by a large margin. This could be the topic of further studies.

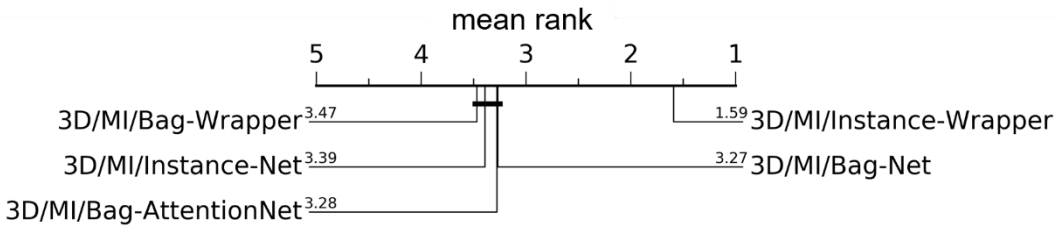


Figure 49. Comparison of MIL algorithms against each other. Groups of models that are not significantly different in performance (at a confidence level of 0.05) are connected by the horizontal line. Axis plots the average ranks of models.

Comparison of 2D and 3D approaches

We compared performance of three types of 2D models, two MIL approaches, *Instance-Wrapper* (so far the best MIL approach) and *Bag-AttentionNet* (the most advanced MIL approach in this study), and a single-instance 3D model built using the most energetically favorable conformations. For the sake of clarity, 29 "non-modellable" data sets for which none of the considered 2D and 3D models had $R^2_{test} > 0.4$ were excluded and the analysis was performed based on the remaining 146 data sets.

Instance-Wrapper demonstrated the best average performance ($R^2_{test} = 0.521$) and outperformed all other models for 71 out of 146 data sets (Table 10). 2D models based on Morgan fingerprints and *Bag-AttentionNet* performed similarly, average R^2_{test} was 0.476 and 0.450, respectively, and the difference between them was statistically insignificant (Figure 50). Other 2D models had even worse performance. 3D single-instance models demonstrated the worst performance. The average R^2_{test} was 0.006.

These results indicate ultimate importance of considering of conformational ensembles for 3D modeling rather than a single conformer due to the difficulty to choose relevant ones. This also shows the promising predictive ability of MIL models which in many cases outperformed the best 2D approaches.

Table 10. Performance comparison of 2D and 3D models. Table reports mean, standard deviations, and median of R^2_{test}. Top-1 is the number of cases where the model was the best. Top-2 is the number of cases where the model was the first- or second-best one.

Model	Mean	Median	Top-1	Top-2
3D/MI/Instance-Wrapper	0.521 ± 0.239	0.530	71	111
2D/MorganFP/Net	0.476 ± 0.186	0.502	41	65
3D/MI/Bag-Attention	0.450 ± 0.401	0.488	14	57
2D/PhysChem/Net	0.436 ± 0.165	0.443	17	41
2D/PharmFP/Net	0.357 ± 0.275	0.383	3	17
3D/SI/Net	0.006 ± 0.936	0.094	0	1

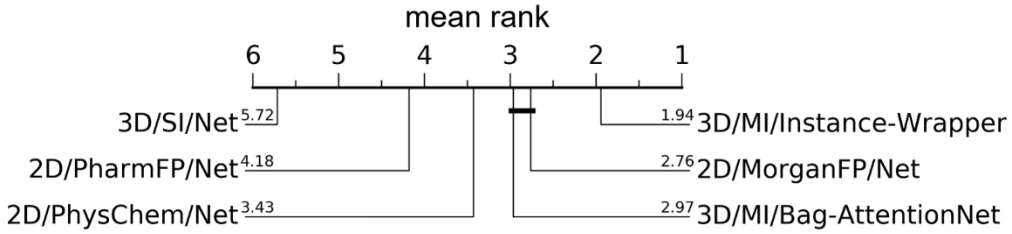


Figure 50. Comparison of 2D and 3D single-instance and 3D multi-instance models against each other. Similarly performed models (at a confidence level of 0.05) are connected by the horizontal line. The numbers correspond to the average ranks of models.

If one looks at the performance of *Instance-Wrapper* and 2D Morgan models it may notice that in some cases *Instance-Wrapper* performs reasonably well while 2D models completely failed (Figure 51). We analyzed physicochemical parameters of data sets where each model performed better than the other to investigate prerequisites of model performance. *Instance-Wrapper* was better than 2D Morgan model for 42 cases, while 2D Morgan was better for 18 data sets. The remaining data sets demonstrated similar performance of both models (difference in $R^2_{test} < 0.1$) and were excluded them from the analysis. We calculated average physicochemical characteristics for both groups of data sets and plot them in Figure 52. The data sets where *Instance-Wrapper* performed better usually had smaller compounds (lower molecular weight, MW). *Instance-Wrapper* also worked better on data sets with less conformationally flexible compounds. This conclusion is supported by observation that the lower average number of rotatable bonds (RTB), the greater number of rings per molecule in a data set and the greater percentage of molecular framework relatively to the size of a whole molecule were more favorable for *Instance-Wrapper* models (Figure 52). This could be explained by the ease to generate more relevant conformers in the case of more rigid molecules. At the same time 3D *Instance-Wrapper* models were better in cases if a data set

contained more molecules with distinct scaffolds. This indicated the better generalizing ability of 3D models and we link this with more abstract nature of 3D pharmacophore quadruplets than circular substructures (Morgan fingerprints).

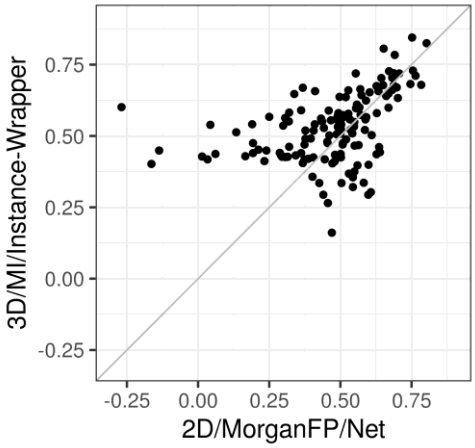


Figure 51. Performance of *Instance-Wrapper* and the best 2D model based on Morgan fingerprints. Only data sets, where at least one of these models had $R^2_{test} > 0.4$, were depicted.

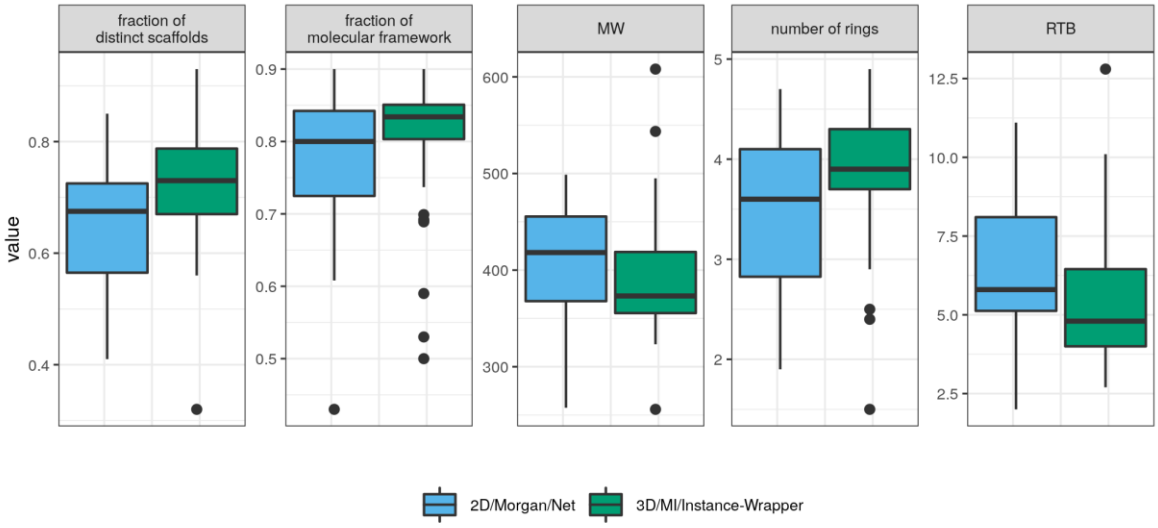


Figure 52. Distribution of average physicochemical parameters for data sets where 2D Morgan or *Instance-Wrapper* models demonstrated better performance with the difference in R^2_{test} greater than 0.1. MW is for molecular weight, RTB – the number of rotatable bonds.

Identification of biologically relevant conformers

The attention mechanism allows *Bag-AttentionNet* models to identify the most relevant conformations during the learning be selecting of conformers with the largest attention weights. We validated this hypothesis by comparing the conformations selected by models with conformations retrieved from PDB. There were only four data sets for which *Bag-AttentionNet* models had $R^2_{test} > 0.4$ and where there were at least 10 compounds in corresponding test sets which had 3D structures in PDB. To measure the accuracy of identification of bioactive conformations we calculated Top-3

success rate as a proportion of compounds for which at least one of three conformations with the highest attention weights fits the experimental structure with RMSD < 2.0 Å. Similarly we calculated the same statistics for the lowest energy conformers.

To compare accuracy of identification of relevant conformers with docking we chose for each protein target a PDB complex with a binding site intersected with most of binding sites of other complexes and used it for docking of the same test set compounds (CHEMBL2820 – 4Y8Y, CHEMBL3048 – 4IMS, CHEMBL335 – 3EAX, CHEMBL4802 – 4KCQ). This cross-docking experiment was more fair then performing re-docking to cognate receptor structures, because in the case of machine learning we do not use information about receptor conformation to select a relevant conformer. Docking was performed using AutoDock Vina[137]. Three top scored poses were taken to calculate top-3 statistics similarly as described above.

Since it was claimed that RDKit conformer generator is able to reproduce bioactive conformations we calculated baseline statistics to estimate top-3 metric if one would randomly choose three conformers for each molecule. We calculated probability of choosing at least one conformer with RMSD below 2 Å for each molecule and averaged these values within each test sets.

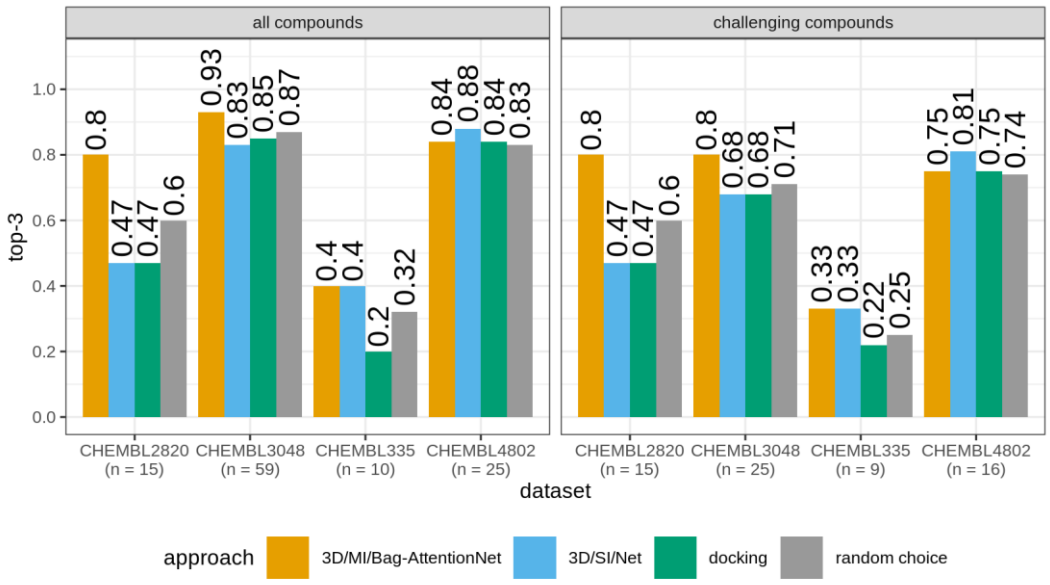


Figure 53. Identification of bioactive conformations within test set compounds for four data sets (n is a number of compounds). *Challenging compounds* is a subset of test set compounds which have mean RMSD of all generated conformers to a bioactive conformation greater than 2Å. R^2_{test} of *3D/MI/Bag-AttentionNet* models was 0.45, 0.55, 0.72 and 0.54 for CHEMBL2820, CHEMBL3048, CHEMBL335 and CHEMBL4802 datasets, respectively.

The calculated baseline statistics was relatively high (Figure 53). This indicates that RDKit conformer generator substantially enriches the set of conformers with those which are close to

experimental ones. This also makes it challenging to improve this baseline performance. *Bag-AttentionNet* models could improve baseline accuracy in identification of bioactive conformers and performed comparably well or better than the random choice. The most remarkable improvement was observed for coagulation factor XI (CHEMBL2820). For two targets, brain and endothelial nitric-oxide synthases (CHEMBL3048 and CHEMBL4802, correspondingly), *Bag-AttentionNet* performed comparable to the baseline. Protein-tyrosine phosphatase 1B (CHEMBL335) was the most difficult target for identification of relevant conformers and all approaches demonstrated low performance. This was caused by the fact that only a part of those compounds bind to the protein, while the remaining part was exposed to water medium and could adopt any conformations with almost no restrictions. Therefore, even docking was not able to identify poses observed in X-ray structures. In general docking performed relatively poor and even worse than the random baseline in cases of CHEMBL2820 and CHEMBL335. Choosing of conformers with the lowest calculated energy resulted in performance comparable to random choice or worse.

Additionally, we considered subsets of "challenging" compounds with mean RMSD to bioactive conformation greater than 2Å. These subsets were enriched by very flexible compounds for which diverse sets of conformations were generated. As expected, the performance of key conformation identification for these compounds was lower (Figure 53), but *Bag-AttentionNet* demonstrated performance comparable or higher than random baseline, supporting an intelligent selection of relevant conformations.

4.3. Enantioselectivity modeling of reactions with chiral catalysts

Enantioselective catalysis is widely used for the synthesis of enantiomerically pure compounds. Design of perspective catalysts is traditionally conducted by iterative modification of the molecular structure aiming to increase the enantioselectivity of a reaction product. Predictive chemoinformatics models may guide chemists toward the most promising catalysts before their synthesis and experimental testing, reducing in such a way both human and material resources[138]. There are several approaches suggested to model catalyst enantioselectivity. One of the state-of-the-art approaches, Average Steric Occupancy (ASO), uses multiple conformer alignment followed by assessment of the occupancy of nodes of a rectangular grid. The ASO descriptors displayed better performance compared to single conformer descriptors[139, 140]. The potential issue of alignment-based 3D method is their applicability to a diverse set of compounds having different scaffolds.

In our pilot study we applied the implemented *Bag-AttentionNet* approach (Figure 48c) on the same data set as used in the work of Zahrt et al[139] to demonstrate its applicability in solving 3D modeling tasks other than prediction of a biological activity. The dataset consists of 1075 reactions of asymmetric addition of thiols to imines catalyzed by phosphoric acid (Figure 54). It was obtained

by systematically performing of 25 transformations in presence of 43 catalysts. The catalyst selectivities were estimated by enantiomeric excess (ee %) ranged from -43 to 99. For the model development, the ee % values were converted into ΔΔG (kcal/mol).

The whole data set was split on training set and three test sets identically as it was done in the study of Zahrt et al. The training set included 384 reactions obtained from 16 transformations in the presence of 24 catalysts. Test sets were selected according to different scenarios: (a) new reactions with known catalysts ("transformation-out"), (b) known reactions with new catalysts ("catalyst-out"), and (c) new reactions with new catalysts ("both-out"). Thus, Test set 1 contained 216 reactions resulted from a combination of 24 catalysts from the training set with 9 new transformations, Test set 2 included 314 instances (19 new catalysts/16 training reactions), and Test set 3 contained 171 instances (19 new catalysts/9 new reactions).

For catalysts we generated up to 50 conformers within 10 kcal/mol by RDKit. Redundant conformers with RMSD < 0.5Å were discarded. Every conformer was encoded by count-based *pmapper* descriptors[94]. Transformations were represented by a condensed graph of reactions (CGR) and encoded by ISIDA fragment descriptors[10]. In this study we used atom-centered subgraphs containing a given atom with the atoms and bonds of its 1 to 4 coordination spheres. Vectors of 2D fragment reaction descriptors and 3D catalyst quadruplets were concatenated to form a combined reaction/catalyst descriptor vector (Figure 55).

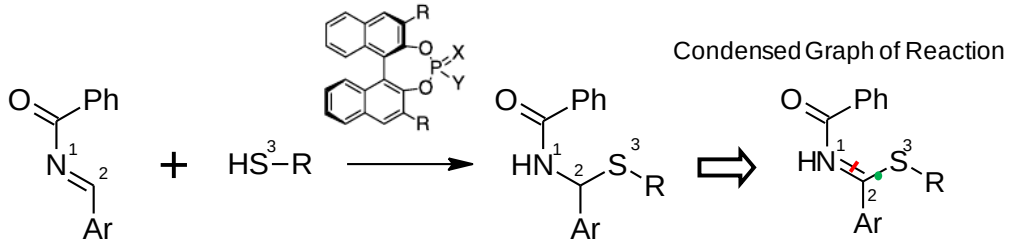


Figure 54. Reaction of N,S-acetal formation and related condensed graph of reaction (CGR). The created bond between the atoms S3 and C2 and double bond transformed to single between the atoms N1 and C2 are highlighted.

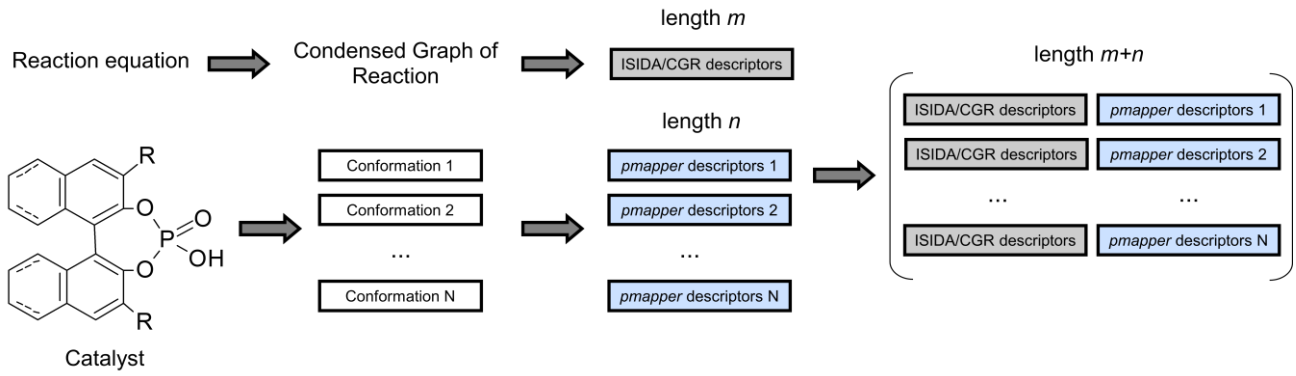


Figure 55. Preparation of descriptors encoding reaction/catalyst combinations. A chemical reaction is encoded by *m* ISIDA/CGR descriptors calculated for the condensed graph of reaction. A catalyst is represented by its *N* conformations, each encoded by *n* pmapper descriptors. Concatenation of reaction and catalyst descriptors results in the vector of (*m*+*n*) size.

For the comparison purpose we built multi-instance as well as a single instance model based on the catalysts with the lowest energy calculated with MMFF. Performances of single-conformation and multi-conformation models (mean absolute error, MAE) in comparison with those of the model by Zahrt et al. are given in Figure 56. One may see that for Test set 1, both single-instance and multi-instance models performed similarly to Zahrt's model, whereas for Test sets 2 and 3, performances of multi-instance model and Zahrt's models were similar whereas the single-instance model performed much worse. This indicates importance of including information about multiple conformations in 3D modeling.

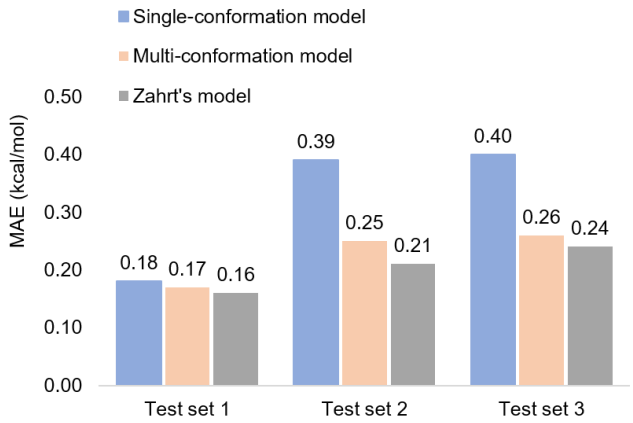


Figure 56. Mean absolute error (MAE, kcal/mol) obtained for three test sets.

We also compared our alignment-free descriptors for 3D modeling with other ones available in RDKit. We built *Bag-AttentionNet* models for the training set and used 25-fold cross-validation to estimate the predictive performance. In each cross-validation fold 1 out of 25 transformations and all 43 associated reactions were moved to a test set. Our *pmapper* descriptors demonstrated the most

robust performance among other available 3D descriptors (Figure 57) confirming the importance of properly chosen descriptors to represent catalysts.

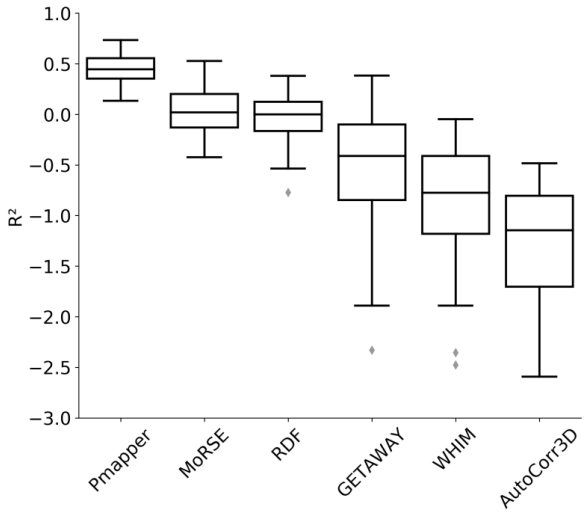


Figure 57. Performance of models based on different classes of 3D descriptors in predicting BINOL-derived catalysts selectivity in 25 reactions. Each box contains a cross-validated determination coefficient R^2 for 25 models (one model per transformation).

In the later study[130] we demonstrated applicability of MIL models to predict enantoselectivity of catalysts on three other datasets. Instance-Wrapper demonstrated the best performance among other MIL approaches and was on par or better than performance of state-of-the-art models.

4.4. Summary

Here, we implemented several multi-instance approaches and demonstrated their applicability in different tasks. Multi-instance learning approaches solve a long-lasting issue of choosing an appropriate conformer for modeling and multi-conformer models systematically outperformed models based on the energetically favorable conformers. The ability of multi-instance models to identify relevant conformers was even higher than for molecular docking. 3D multi-conformer models could even outperform 2D models in a large number of tasks. However, 3D models are more computationally expensive. Therefore, we recommend to use 3D multi-conformer models to predict biological activity of compounds if 2D models fail. In combination with *pmapper* descriptors 3D multi-conformer models do not require pre-alignment of conformers of compounds that enables modeling of data sets with diverse and flexible molecules.

All these studies were pilot ones and there are still many open questions and room for improvement. Many tasks can be formulated and solved within MIL approaches. For example, one may consider compounds as a set of tautomers and protonation states to make more comprehensive

representation of molecules. Modeling of properties of mixtures can be also solved within the MIL framework if individual components of a mixture are considered as instances. A unique feature of MIL approaches is the ability to identify key instances – molecular forms associated with an observed property of a molecule. Despite of success examples in the case of identification of relevant conformer responsible for protein binding it requires more comprehensive and systematic validation. We summarized all these features and issues in the recent review[141].

The implemented models are available in the open-source repository https://github.com/dzankov/3D-MIL-QSAR to stimulate researches in the field of multi-instance learning in chemistry.

Chapter 5. Structure enumeration and de novo design (chemical space exploration)

The drug-like chemical space is vastly enormous – its size estimates in ~10^{33} compounds[142]. In the nearest future, it will be impossible to enumerate this space or perform any kind of exhaustive search. Therefore, methods and strategies to explore this space effectively attract vivid research interest. One of the popular strategies is de novo design – model-driven generation of new chemical structures with promising predicted properties[143, 144]. The widely used strategy is iterative generation of compounds which is guided by QSAR, docking or pharmacophore models[145-149]. Each iteration new structures are generated and candidates with the most promising predicted properties are selected for the next iteration. The commonly used structure generation approaches can be divided on atom-, fragment- and reaction-based. Atom-based approaches generate new structure by application of basic operations, e.g. add/remove atom/bond, etc[150]. This gives excellent flexibility and the ability to generate every possible molecule. However, there are several issues of this kind of approaches: i) possible combinatorial explosion due to the large number of modification steps required to reach desired molecules; ii) necessity to control chemical validity of generated structures and iii) the main issue is poor synthetic accessibility of designed molecules which hardly can be controlled in this setting. Reaction-based approaches use pre-compiled reaction and reactant libraries which are applied to sequentially modify molecules[148]. They usually use only coupling reactions and therefore applicable only in decoration or expansion of molecules. This can limit the covered chemical space but greatly improves synthetic accessibility of output molecules that was confirmed in many studies[148, 151-153]. Fragment-based structure generators construct molecules by addition/removal/replacement of fragments[146, 149, 154]. They provide greater flexibility in exploration of chemical space than reaction-based approaches but cannot guarantee synthetic feasibility of designed compounds. However, synthetic feasibility should be better than for atom-based approaches and it should be easier to control, e.g. by linking of synthetically feasible fragments.

Another strategy to generate molecules is application of recently emerging deep neural networks[155-161]. They provide great flexibility of structure generation due to different architectures[162-164]. However, they still suffer from the issue of poor synthetic feasibility of designed molecules and this is a major issue for the majority of existing structure-generation approaches[165].

Here, we suggested and implemented a new fragment-based structure generation approach inspired by matched molecular pairs which provides a certain level of control over synthetic feasibility of generated structures and solves to some extent this long lasting issue for fragment-based approaches.

5.1. Chemically reasonable mutations (CReM)

The idea of interchangeable fragments – the core of the developed approach[166] – is directly related to the matched molecular pairs approach considering their local context[167]. Interchangeable fragments are fragments that occur in the same local chemical context in structures of existing compounds (Figure 58). Atoms within a particular radius around attachment points of a fragment represent the local chemical context. We replace one fragment by another if both have identical chemical context that should result in a chemically valid and feasible structure. Thus, by design, the chemical validity of generated structures is guaranteed. Intuitively, it can be also expected that the generated compounds are synthetically feasible.

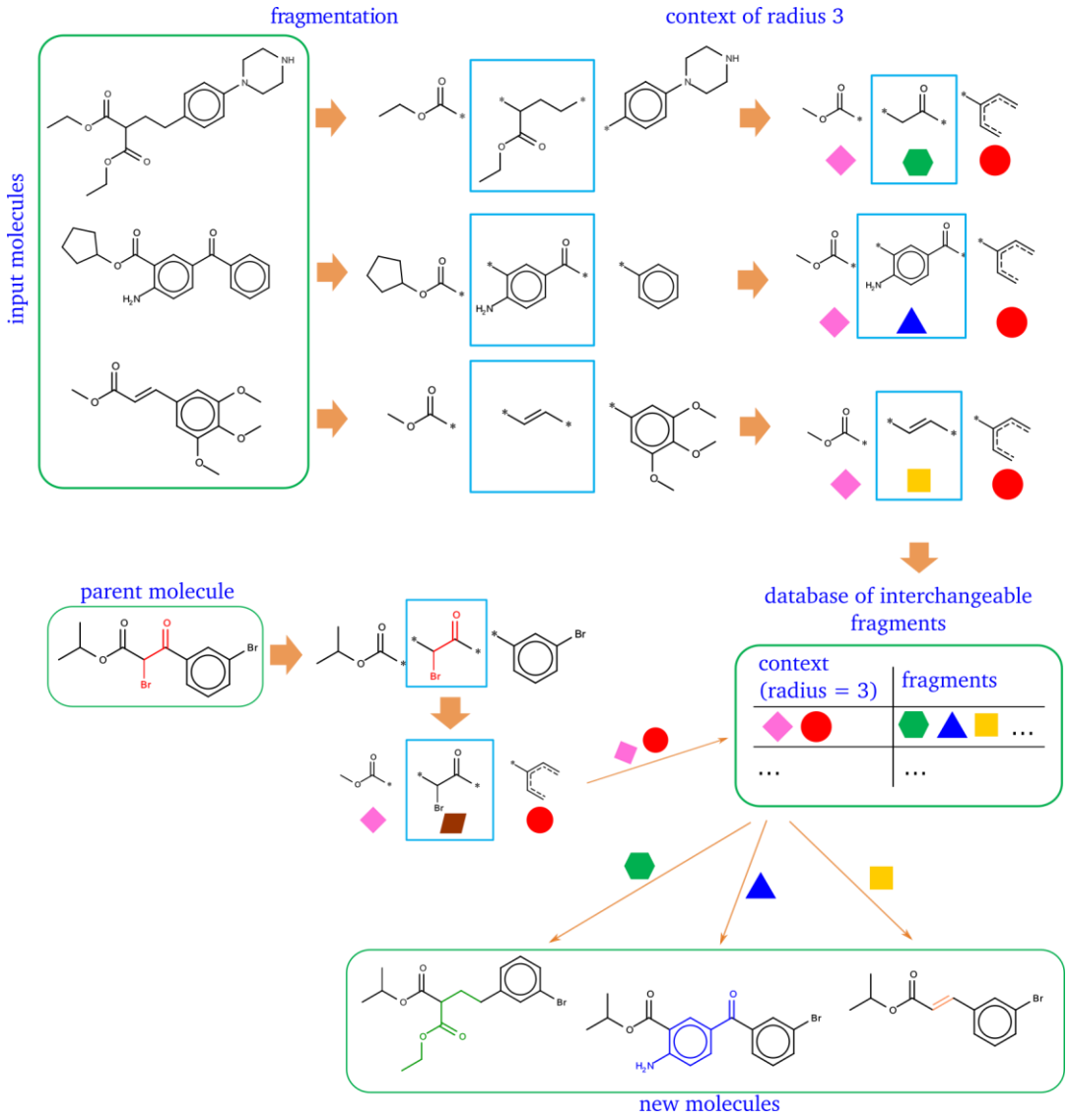


Figure 58. Generation of a database of interchangeable fragments and new molecules.

Generation of a database of interchangeable fragments is a two-step procedure. On the first step, structures of known compounds are exhaustively fragmented by cutting up to 4 non-cyclic

single bonds between two heavy atoms using RDKit implementation of the matched molecular pairs algorithm suggested by Husain and Rea[168]. Hydrogens are cut separately. On the second step, a context of a given radius is determined for attachment points of each fragment and encoded in a SMILES string. This SMILES string is canonicalized to get both a canonical numbering of attachment points and canonical SMILES representation of a context. SMILES representation of a context of a given radius and an associated fragment are stored in a database table as a key-value pair for a subsequent search of interchangeable fragments (values) having an identical context (key) (Figure 58).

To replace a fragment in a molecule its context of a given radius is determined and canonically encoded. The given SMILES string of a context is searched in a fragment database and fragments with the same context are retrieved and used for fragment replacement (Figure 58).

We implemented three modes of structure generation: MUTATE, GROW and LINK (Figure 59). MUTATE is a replacement of an arbitrarily chosen fragment with another one. GROW is a special case of a MUTATE operation – replacement of a hydrogen with another fragment. LINK is a replacement of hydrogen atoms in two molecules to link them by an appropriate fragment.

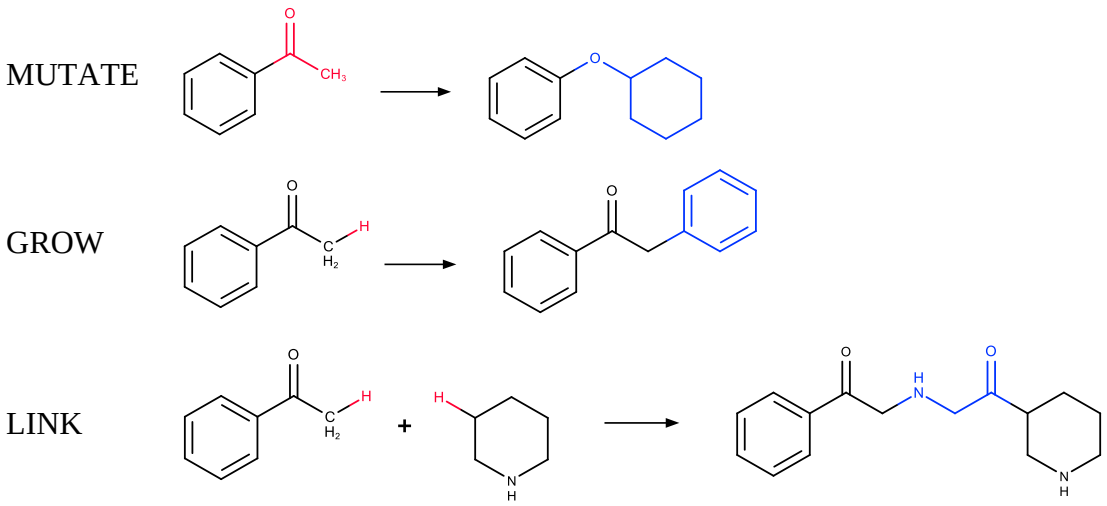


Figure 59. CReM structure generation modes.

Several tuning parameters are available:

1. Structures of the input compounds used to create a database of interchangeable fragments. Management of the content of the input compound database used for fragmentation gives indirect control over enumerated structures and provides additional flexibility. The selection of more synthetically feasible input compounds may improve synthetic feasibility of generated compounds. At the same time pre-selection of compounds for fragment library enumeration may reduce diversity and novelty of generated structures.
2. Radius of a considered molecular context.

Increasing the radius of a considered molecular context will decrease the appearance of new chemotypes in enumerated compounds and makes replacements more conservative. In other words, no new chemotypes smaller than a chosen radius can be created if these chemotypes are absent in the fragment database. This property can be useful if one excludes compounds having undesired patterns (e.g. PAINS or toxicophores) to create a fragment database and chooses a large enough radius to avoid appearance of these undesired fragments in generated molecules. Making more conservative replacements can also improve synthetic accessibility of generated compounds.

3. Frequency of occurrence of interchangeable fragments in the input database.
   Similarly to the synthetic accessibility score suggested by Ertl & Schuffenhauer[169] it can be supposed that replacement with more frequently occurred fragments will lead to more synthetically feasible compounds. This will also reduce the number of replacements and increase search speed.
4. Size of fragments which will replace each other.
   The size of replaceable fragments can control exhaustiveness of chemical space exploration by increasing or decreasing search steps and depends on the goal of a particular study. Lead optimization studies may require small steps to explore local chemical space around a parent compound, whereas lead generation may require large steps in the beginning to quickly and coarsely explore larger chemical space and smaller steps in the end to finely tune generated structures.
5. Maximum number of randomly chosen replacing fragments.
   Limiting the maximum number of replacements can speed up the exploration of a chemical space as generated fragment databases can be very large and making all possible replacements can be costly.
6. Protection of selected atoms from modification or modification of only selected atoms.
   This functionality can be useful for property/activity optimization studies to protect scaffold or pharmacophore features from changes or to modify molecules only at specific positions.

With all these options CReM approach possesses great flexibility and control over generated structures.

Here, we will describe application of CReM to Guacamole tasks. Application of CReM to other use cases is described in the paper[166]. Guacamol is a set of distribution learning and goal-directed benchmarks[170]. Within distribution learning benchmarks a structure generator should reproduce the distribution of the training set molecules. In goal-directed benchmarks it should reproduce structures of known drugs or generate compounds similar to reference ones, enumerate isomers of a particular empirical formula, perform multiobjective optimization of properties of

reference compounds, scaffold hopping and scaffold decoration. There are 20 goal-directed tasks. Scores in each tasks span from 0 to 1 (perfect). So, the maximum total score which can be achieved is 20.

The database of interchangeable fragments was created from compounds of ChEMBL database (version 22). After standardization, removal of duplicates and compounds containing non-organic atoms (organic atoms are C, N, O, S, P, F, Cl, Br, I, B) we got 1.5 million distinct structures. Compounds were fragmented as described above and fragments and their contexts were stored in the database. As expected the number of distinct fragment-context pairs increased with the increase of the context radius: radius 1 had 35 million pairs, 2 – 41M, 3 – 51M, 4 – 62M, 5 – 74M.

Guacamol distribution-learning benchmarks

We implemented a specific approach to test CReM on distribution learning benchmarks. A seed compound having molecular weight less than 350 Da was randomly chosen from the reference ChEMBL database. The MUTATE operation was applied to it to enumerate new structures. The size of replaced fragments was set to a range from 0 to 8 heavy atoms. Different size of replacing fragments relative to replaced ones was chosen: ±2, ±6, ±10 and a non-symmetric one from -10 to 2. The larger difference should result in larger steps and better coverage of chemical space. The maximum number of randomly chosen replacements was set to 2, 5, 10 or 100. A smaller number of replacements should result in more diverse generated compounds and better coverage of chemical space due to the greater number of steps required to generate 10000 distinct structures. Compounds with molecular weight greater than 500 Da were discarded. A random compound from the generated ones on the last iteration was chosen for the next iteration if no compounds with molecular weight less than 500 Da were generated a random compound from an already generated population was picked. Each combination of parameters was tested in three independent runs.

As expected, all generated structures were chemically valid irrespective of the chosen setup (Table 11). Novelty achieved a maximum value in almost all cases. Uniqueness of compounds was also high. KL divergence was somewhat greater in cases where the larger variation of the size of replacing fragments was allowed and where a smaller number of compounds was selected on each iteration. Moderate KL divergence and low Frechet ChemNet Distance scores showed that the implemented iterative search approach could not reproduce the distribution of the reference space well.

Table 11. Results for the distribution learning Guacamol benchmarks.

Case	Min increase	Max increase	Max replacements	Validity	Uniqueness	Novelty	KL divergence	Frechet ChemNet Distance
CReM	-2	2	100	1 ± 0	0.935 ± 0.021	1 ± 0	0.443 ± 0.023	0.021 ± 0.007
CReM	-2	2	10	1 ± 0	0.942 ± 0.008	1 ± 0	0.530 ± 0.061	0.024 ± 0.034
CReM	-2	2	5	1 ± 0	0.941 ± 0.003	1 ± 0	0.572 ± 0.038	0.044 ± 0.053
CReM	-2	2	2	1 ± 0	0.950 ± 0.002	1 ± 0	0.551 ± 0.054	0.019 ± 0.018
CReM	-6	6	100	1 ± 0	0.942 ± 0.023	0.999 ± 0	0.541 ± 0.056	0.018 ± 0.012
CReM	-6	6	10	1 ± 0	0.924 ± 0.010	1 ± 0	0.603 ± 0.019	0.041 ± 0.045
CReM	-6	6	5	1 ± 0	0.921 ± 0.022	1 ± 0	0.584 ± 0.034	0.038 ± 0.040
CReM	-6	6	2	1 ± 0	0.935 ± 0.009	1 ± 0	0.605 ± 0.015	0.053 ± 0.050
CReM	-10	10	100	1 ± 0	0.918 ± 0.019	1 ± 0	0.531 ± 0.058	0.071 ± 0.027
CReM	-10	10	10	1 ± 0	0.907 ± 0.022	0.999 ± 0.001	0.622 ± 0.011	0.030 ± 0.016
CReM	-10	10	5	1 ± 0	0.875 ± 0.025	1 ± 0	0.599 ± 0.035	0.085 ± 0.056
CReM	-10	10	2	1 ± 0	0.850 ± 0.094	1 ± 0	0.590 ± 0.064	0.006 ± 0.005
CReM	-10	2	100	1 ± 0	0.945 ± 0.021	0.999 ± 0	0.550 ± 0.037	0.016 ± 0.012
CReM	-10	2	10	1 ± 0	0.950 ± 0.008	1 ± 0	0.545 ± 0.007	0.045 ± 0.010
CReM	-10	2	5	1 ± 0	0.956 ± 0.001	1 ± 0	0.533 ± 0.073	0.048 ± 0.036
CReM	-10	2	2	1 ± 0	0.962 ± 0.006	1 ± 0	0.577 ± 0.027	0.042 ± 0.037
SMILES LSTM*				0.959	1	0.912	0.991	0.913
Graph MCTS*				1	1	0.994	0.522	0.015
AAE*				0.822	1	0.988	0.886	0.529
ORGAN*				0.379	0.841	0.687	0.267	0
VAE*				0.870	0.999	0.974	0.982	0.963

*results were taken from the Guacamol paper [170]

Guacamol goal-directed benchmarks

To evaluate CReM on goal-directed Guacamol tasks we implemented an iterative search protocol inspired by the genetic algorithm. If the list of the seed structures was empty the seed structures were chosen randomly from the list of SMILES supplied with the Guacamol and represented the whole ChEMBL database. The size of a population selected on each iteration was set to be equal to the size of the output population, but not less than 10 compounds. To make the search adaptive we adjusted the fragment size of replacement according to the current score of the population. If the score was equal or less than 0.3 (far from the goal) the replacing fragment can differ at most on ±10 heavy atoms from the replaced one. If the score was greater than 0.8 (close to the goal) the replacing fragment can differ at most on ±4 heavy atoms from the replaced one. Intermediate fragment sizes (5-9) were chosen if the score was within 0.3 - 0.8 range. This allows to quickly explore chemical space in the beginning and to better tune structures at the end of generation. For each compound in a population up to 1000 randomly chosen mutations were applied. Compounds, which were already used for structure generation, were stored in a separate list and

removed from the list of generated structures. Remaining top-scored compounds were selected for the next iteration.

Since the implemented optimization procedure is local and can get stuck in local optima we implemented three levels of “patience”. At the first level if the best score was not improved after three consecutive iterations the fragment size was increased on ±1 and the number of randomly chosen replacements on 100 irrespectively to the current score. This makes the small stepwise increase in chemical space exploration. If after 10 consecutive iterations no improvement was observed larger changes were applied: the size of replacing fragment was increased on ±10 and the number of replacements on 500. This would enable the rougher exploration of a chemical space around the best candidates. At the third level, if after 33 iterations no improvement was observed new seed compounds were randomly selected to restart the search but the best found candidates were kept. This procedure was not applied if the seed structure was supplied with the task. The list of already visited compounds was cleared after any change of generator parameters whether this was caused by improving the best score or by exceeding one of "patience" levels. The maximum execution time of each task was set to 5 hours or maximum of 1000 iterations were allowed.

The results demonstrated that the implemented search algorithm based on CReM approach compared well with the published reference approaches by achieving the highest score in 16 out of 20 tasks (Table 12). However, the total score was slightly lower than the total score of Graph GA approach, which uses the genetic algorithm on molecular graphs. This is mainly due to the considerable advantage demonstrated by Graph GA approach (0.891) over CReM-based approach (0.763) in the task of generation of molecules, which were structurally dissimilar to sitagliptin but had similar lipophilicity and topological polar surface area. Interestingly, the other reference approaches performed even worse in this task.

Table 12. Results for the Guacamol goal-directed benchmarks.

task	SMILES LSTM*	SMILES GA*	Graph GA*	Graph MCTS*	CReM
Celecoxib rediscovery	1.000	0.732	1.000	0.355	1.000
Troglitazone rediscovery	1.000	0.515	1.000	0.311	1.000
Thiothixene rediscovery	1.000	0.598	1.000	0.311	1.000
Aripiprazole similarity	1.000	0.834	1.000	0.380	1.000
Albuterol similarity	1.000	0.907	1.000	0.749	1.000
Mestranol similarity	1.000	0.79	1.000	0.402	1.000
C11H24	0.993	0.829	0.971	0.410	0.966
C9H10N2O2PF2Cl	0.879	0.889	0.982	0.631	0.940
Median molecules 1	0.438	0.334	0.406	0.225	0.371
Median molecules 2	0.422	0.38	0.432	0.170	0.434
Osimertinib MPO	0.907	0.886	0.953	0.784	0.995
Fexofenadine MPO	0.959	0.931	0.998	0.695	1.000
Ranolazine MPO	0.855	0.881	0.92	0.616	0.969
Perindopril MPO	0.808	0.661	0.792	0.385	0.815
Amlodipine MPO	0.894	0.722	0.894	0.533	0.902
Sitagliptin MPO	0.545	0.689	0.891	0.458	0.763
Zaleplon MPO	0.669	0.413	0.754	0.488	0.770
Valsartan SMARTS	0.978	0.552	0.990	0.04	0.994
Deco Hop	0.996	0.970	1.000	0.590	1.000
Scaffold Hop	0.998	0.885	1.000	0.478	1.000
total score	17.341	14.398	17.983	9.011	17.919

*results were taken from the Guacamol paper [170]

Synthetic accessibility of generated molecules

From our preliminary experiments it was found that synthetic accessibility of generated molecules depends most strongly on the chosen radius and the input set of molecules which are converted into a database of interchangeable fragments. Therefore, we studied this relationship using goal-directed Guacamol benchmarks.

We hypothesized that using more synthetically feasible molecules for fragmentation will improve accessibility of generated compounds. To verify this hypothesis we created additionally to the fragment database created from all ChEMBL compounds two databases where compounds were preliminary filtered according to their synthetic accessibility (SA) score. SA score was suggested by Ertl and Schuffenhauer and it spans from 1 (easy to synthesize) to 10 (hard to synthesize)[169]. The median SA value for compounds from ChEMBL is 3. Therefore, we selected subsets of compounds with SA scores at most 2 and 2.5. This substantially reduced the number of compounds and as a consequence the number of distinct fragment-context pairs in the generated databases (Table 13).

Table 13. Fragment database statistics.

database	label	number of compounds	number of distinct context-fragment pairs in a database				
			radius 1	radius 2	radius 3	radius 4	radius 5
all ChEMBL compounds	all	1 554 260	51 312 712	60 116 865	74 198 048	89 152 239	104 076 972
SAScore ≤ 2.5	SA2.5	572 527	11 260 763	13 858 193	17 478 013	21 628 459	26 032 412
SAScore ≤ 2	SA2	107 806	1 529 346	1 937 044	2 467 559	3 099 537	3 799 202

Since the implemented search algorithm to solve goal-directed Guacamol tasks described above was stochastic we made five independent runs for every combination of the fragment database and radius to get more robust estimates. To estimate performance we summed scores of individual Guacamol tasks which can be maxed to 20. To estimate synthetic feasibility of generated compounds we averaged SA scores among all generated compounds in a particular run to get an overall estimation of synthetic feasibility of generated compounds. As an additional estimation of synthetic feasibility we used the AiZynthFinder retrosynthetic approach[171] which is reimplementation of the seminal study of Segler et al.[172] It tries to reconstruct a sequence of transformations which may result in a desired compound starting from known molecules taken from ZINC. In this case the measure of synthetic feasibility was the percentage of compounds for which the reconstruction of retrosynthetic pathways was successful. As an additional evaluation metric we calculated the average number of steps in successfully reconstructed pathways.

We observed good reproducibility of benchmark results in terms of Guacamol and SA scores across five repeated runs (Figure 60a-b). As expected the Guacamol total score was the highest for the runs based on the full fragment database (Figure 60a). Increasing of the context radius resulted in decreasing of the Guacamol total score. This effect was less pronounced for the full fragment database and more pronounced for fragment databases created from synthetically more feasible compounds. SA scores were also predictably changed if we switched from the full fragment database to those ones created from synthetically more feasible compounds (Figure 60b). Increasing of the context radius also resulted in improvement of SA scores of generated compounds due to making more conservative replacements.

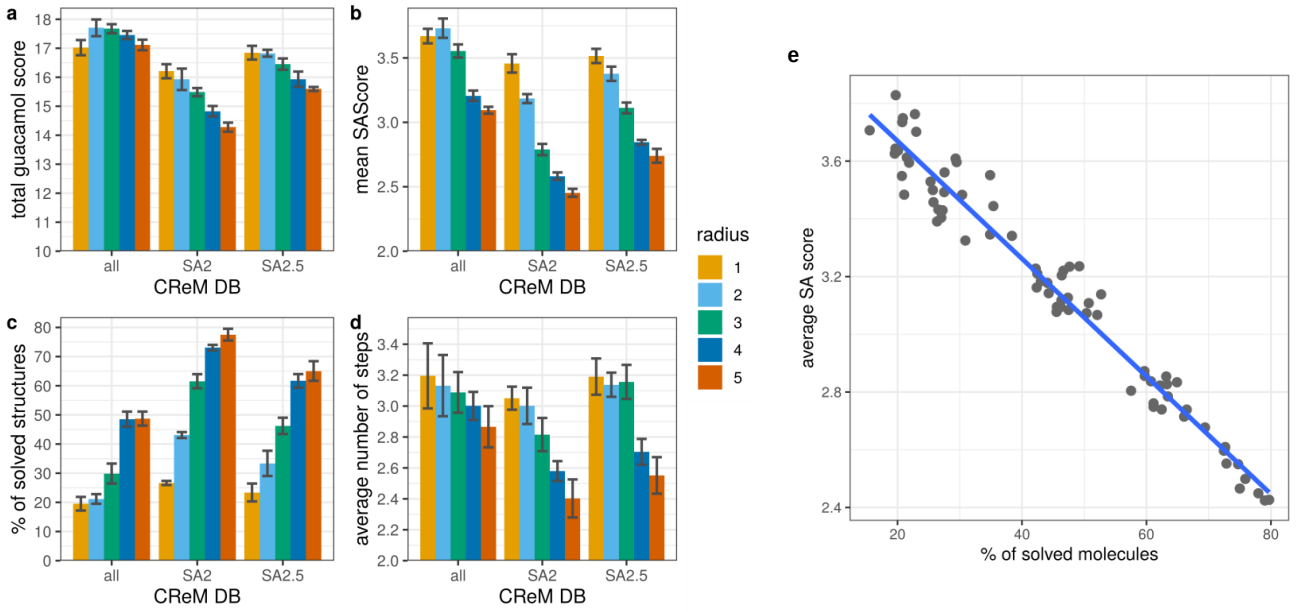


Figure 60. Average guacamole total score (a), average SA score (b), the percentage of structures for which AiZynthFinder successfully reconstructed retrosynthetic trees (c) and the average number of steps in the found pathways leading to solved structures (d). Error bars designate standard deviation from the average among 5 independent runs. (e) Average SA scores vs. the percentage of molecules for which AiZynthFinder solved retrosynthetic pathways for molecules generated in individual runs.

The percentage of molecules for which AiZynthFinder could resolve retrosynthetic pathways strongly correlated with average SA scores ($R_{Pearson}$ = -0.98) (Figure 60e). This suggests that SA scores are relevant for estimation of synthetic accessibility and additionally confirms conclusions made above that increasing of the context radius improves the synthetic accessibility of generated molecules as well as choosing of a more heavily SA-biased fragments database. Pathways for almost 80% of molecules were reconstructed in the case of the SA2 database and context radius 5 (Figure 60c) and in average 2.4 steps were required to get these molecules (Figure 60d).

The trade-off between total Guacamol scores and synthetic feasibility of generated compounds is clearly demonstrated in Figure 61. Increase of the context radius improved synthetic feasibility of compounds but lowered total benchmark scores. Using databases generated from synthetically more feasible compounds resulted in further decrease of Guacamol scores and improvement of synthetic feasibility of generated compounds. There is a noticeable Pareto front created by suboptimal solutions (Figure 61).

Among references approaches only Graph GA[173] and SMILES LSTM[161] were close or lay on the Pareto front. However, Graph GA approach resulted in overly complex molecules with average SA score greater than 4 whereas the average SA score of compounds from ChEMBL was around 3. Other approaches, SMILES GA[174] and Graph MCTS[173] were far from the Pareto front and were not competitive.

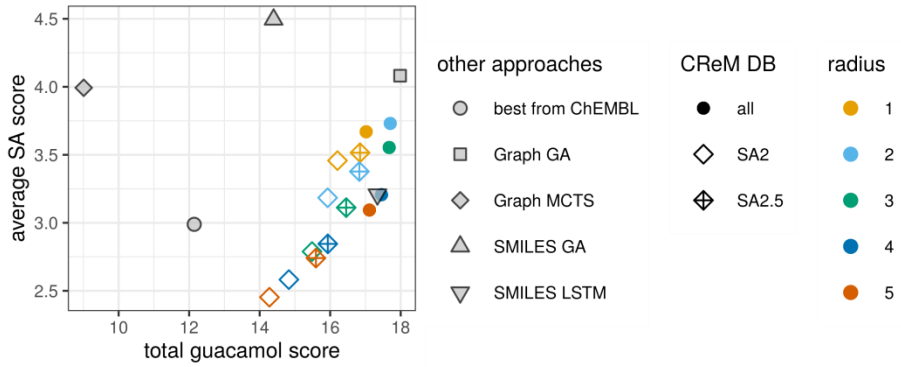


Figure 61. Total Guacamol score and average synthetic accessibility scores for CReM-based and reference approaches which results were taken from the Guacamol paper[170].

Gao and Coley suggested to introduce synthetic feasibility score in Guacamol optimization functions to explicitly bias generated compounds towards more synthetically accessible ones (evaluator bias)[165]. In their study they considered only ten so called hard Guacamol tasks which are related to multiobjective optimization and scaffold hopping/decoration. That was reasonable because for many of the remaining tasks perfect scores were usually achieved. To introduce the evaluator bias the authors implemented desirability functions for SA scores. Compounds having SA scores approximately up to 2.3 were considered favorable, desirability of compounds with SA scores from 2.3 to approximately 4 was decreased from 1 to 0, more complex compounds were considered unfavorable. Thus, it was expected that the majority of solutions biased by SA score would have SA scores below 4.

We took their results and compared with those obtained by CReM for the same tasks (Figure 62). In the case of CReM, average SA scores for compounds generated within 10 hard tasks were comparable or a little bit higher to the corresponding scores of compounds generated in all 20 tasks. At the same time, CReM solutions for hard tasks had high total Guacamol scores. The majority of solutions resulted in the total score from 8 to 9 out of maximum 10 points. The best solutions still formed the noticeable Pareto front (Figure 62).

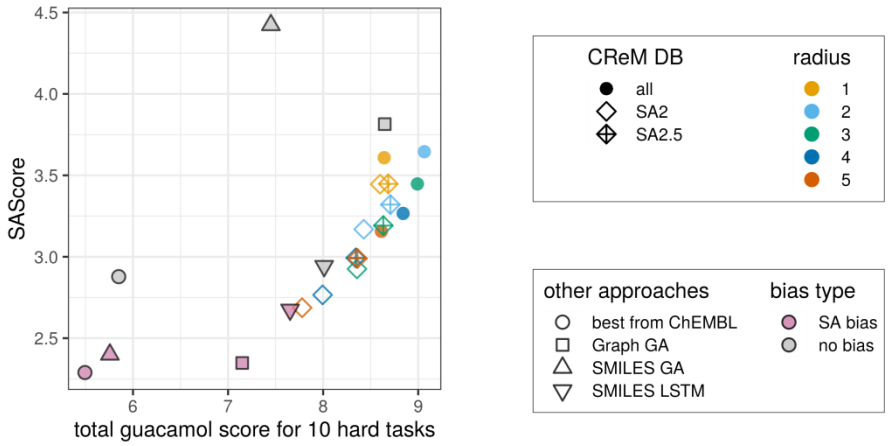


Figure 62. Total Guacamol score and average synthetic feasibility scores for CReM-based and reference approaches which results were taken from the Guacamol paper[170] for 10 hard tasks.

Unbiased reference approaches Graph GA and SMILES LSTM demonstrated relative decrease in total Guacamol scores for hard tasks and they were no longer on the Pareto front but still close to it (Figure 62). Explicit biasing with SA score resulted in solutions having improved synthetic feasibility but lower total Guacamol scores. These reference solutions extended the Pareto front created by the CReM solutions and contributed to the bottom part of that front. Graph GA and SMILES GA approaches were the most sensitive to explicit biasing and generated compounds had substantially improved synthetic feasibility scores whereas compounds generated with SMILES LSTM were less improved in their average synthetic feasibility.

5.2. Summary

The developed CReM approach for fragment-based structure generation solves the main issue of previously available methods. It provides a clear control over synthetic complexity of generated molecules. The Guacamol results demonstrated a competitive nature of CReM relatively to advanced approaches based on neural network generative models. The tool can be combined with any kind of a model to guide the generation process.

CReM is available as open-source software (https://github.com/DrrDom/crem) and as a free web application (https://crem.imtm.cz). It was also included in benchmarking studies by other researchers[175-177] and became an integral part of Distilled Graph Attention Policy Network model[178] developed at the University of California, Berkley (https://github.com/yulun-rayn/DGAPN) and it was included in the developing MolDrug project - https://github.com/ale94mleon/MolDrug. Currently we are working on development of tools which integrate CReM with the most commonly used modeling approaches (molecular docking, 3D pharmacophores, machine learning) to address different aims: i) de novo structure generation; ii) scaffold decoration; iii) expansion of fragment-size hits inside a binding site; iv) multi-objective optimization.

Chapter 6. Computer-aided design of biologically active molecules

I was involved in many drug design and medicinal chemistry projects, which were related to development of anti-platelet[179-182], antiviral[183], anti-parasitic[184, 185] and anti-cancer[186, 187] agents. The most recent examples, where we used some of the developed approaches described above, were related to inhibitors of kinases and ligands of GPCR receptors. We participated in the first CACHE challenge[188] where we also applied our new tools to find hits for the WDR domain of leucine-rich repeat kinase 2 (LRRK2) within the Enamine REAL space containing about 16 billion compounds. However, the recent applications were not ready for publication. Therefore, here, we will describe several selected projects which were already published.

6.1. Antagonists of the open form of integrin $\alpha_{IIb}\beta_3$ as antithrombotic agents

Thrombus formation is the most important pathological mechanism underlying atherothrombotic diseases such as acute coronary syndromes and ischemic stroke/transient ischemic attack, which are responsible for elevated mortality worldwide and which are a platelet-mediated phenomenon[189, 190]. To start to form clots, platelets should be turned from the rested state to the activated one[191]. Rupture of atherosclerotic plaques is supposed to be the main cause of arterial thrombus formation[192, 193]. This exposes such platelet activating proteins as tissue factor, von Willebrand factor, collagen, etc. Activated platelets are able to excrete other agonists of platelet activation such as adenosine diphosphate and thromboxane A2, which promote activation of adjacent platelets[194]. Activated platelets change their shape and expose fibrinogen receptors, integrin $\alpha_{IIb}\beta_3$, which change their conformation from bent conformation to extended conformation with closed headpiece (Figure 63). Then β-subunit moves away from α-subunit and the receptor goes into the high-affinity state with open headpiece in which it binds fibrinogen and von Willebrand factor, resulting in clot formation and clot adherence, respectively[195]. Thus, inhibition of $\alpha_{IIb}\beta_3$ can prevent clot formation regardless of the platelet activation pathway[196-198].

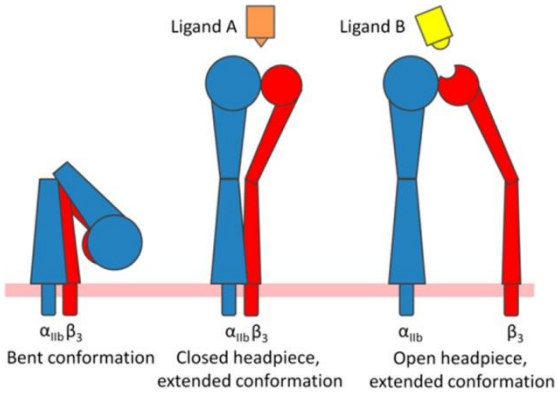


Figure 63. Conformational changes in integrin $\alpha_{IIb}\beta_3$ upon activation and ligand binding.

Most antagonists of $\alpha_{IIb}\beta_3$ represent peptidomimetics, which mimic RGD or KGD sequence of fibrinogen (Figure 64) and bind to the open form of the integrin (Figure 63, Ligand B). There are two marketed drugs Tirofiban[199] and Eptifibatide[200], which are RGD-peptidomimetics, and Abciximab, which is a monoclonal antibody specific for an epitope on β_3 subunit[201]. However, these compounds have some side effects like inducing thrombocytopenia[202, 203], and therefore searching of new antagonists is continuing.

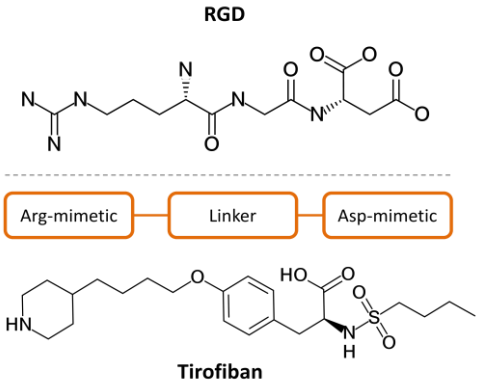


Figure 64. RGD peptide sequence and RGD-peptidomimetic tirofiban, the marked anti-platelet drug.

Due to abundant information about activity of known $\alpha_{IIb}\beta_3$ antagonists and X-ray structures of protein-ligand complexes we were able to implement a comprehensive multi-stage virtual screening pipeline which included QSAR and pharmacophore models and molecular docking[181].

We collected two data sets: (1) 338 compounds with reported affinity values for αIIbβ3 and (2) 453 compounds tested for antiaggregation activity. QSAR models were built using Random Forest method and three types of fragmental descriptors: simplexes[14], ISIDA fragments [20] and fuzzy pH-dependent pharmacophoric triplets[204]. The consensus models demonstrated reasonably high performance estimated by 5-fold cross-validation (Table 14).

Table 14. 5-fold external cross-validation vtatistics for consensus 2D QSAR models of affinity for $\alpha_{IIb}\beta_3$ and antiaggregation activity. AD denotes applicability domain.

	R^2	RMSE	R^2_{AD}	$RMSE_{AD}$	AD coverage
affinity for $\alpha_{IIb}\beta_3$	0.75	0.76	0.76	0.72	0.97
antiaggregation activity	0.52	0.77	0.54	0.74	0.99

Three structure-based pharmacophore models were created using LigandScout from three available complexes of $\alpha_{IIb}\beta_3$ with small molecule antagonists L-739,758, Tirofiban, Eptifibatide (PDB codes 2VC2, 2VDM, and 2VDN, respectively). The models performance has been assessed in virtual screening of a validation set combining the affinity set (338 compounds) with decoys selected from the ChEMBL database (1518 compounds). Due to too large number of pharmacophore features

the models were to specific and did not retrieve any compounds from the validation set. Therefore, some features were manually removed or made optional. These adjustments significantly improved the models performance reaching precision = 0.76–0.86 and recall = 0.13–0.26. The joint application of all three models to the validation set slightly improved overall prediction performance (precision = 0.81 and recall = 0.36), whereas enrichment ratio at 1% and 5% was equal to 5.56 and 6.78, respectively.

In order to improve recall of pharmacophore models we developed ligand-based models based on active compounds ($pIC_{50} \geq 8$) from the affinity data set. We used important information which was retrieved from structure-based pharmacophore models – the distance between key features, centers of negative and positive charges, should be around 16-17Å (Figure 65, top). Therefore, we pre-filtered conformers of training set compounds to satisfy this criterion to get more reasonable models. We obtain seven models which demonstrated reasonable joint predictive performance (precision = 0.67 and recall = 0.93) and higher enrichment ratios than structure-based models: 9.04 and 7.80 at 1% and 5%, respectively.

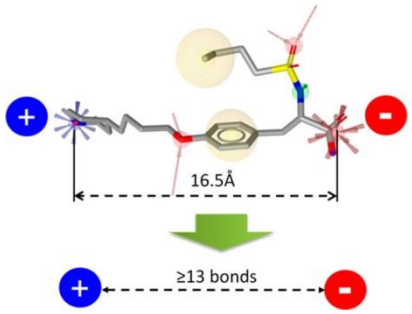


Figure 65. The structure-based pharmacophore model of tirofiban (top) and the simplified 2D pharmacophore suggested for virtual screening (bottom).

To speed up a virtual screening of large databases, a simple topological 2D pharmacophore model has been developed on the basis of structure- based pharmacophores. It consists of only two features, centers of positive and negative charges, separated by 13 bonds (Figure 65, bottom). This roughly corresponds to the distance of 16 Å separating these features in 3D pharmacophores. These features are rarely occurred in typical compounds of chemical database and therefore such pre-filtering should substantially reduce the number of compounds.

We validated three docking tools (FlexX, MoE and PLANTS) on three PDB complexes (2VC2, 2VDM, and 2VDN) for their ability to discriminate actives from inactives in the affinity data set. It was found that MOE has the best performance on 2VDM protein structure (AUC = 0.72), whereas PLANTS and FlexX achieved AUC values 0.59 and 0.49, respectively. Therefore, docking with MOE on 2VDM protein structure was selected for the virtual screening pipeline.

Virtual screening of BioinfoDB[205] contained about three million of commercially available compounds with pharmacophore and QSAR models resulted in no hits. Even after 2D topological

pharmacophore only 210 compounds were remained that is explained by the rarity of occurrence pf positively and negatively charged centers in commercially available compounds. Therefore, the focused virtual compound library has been created using a fragment-based approach. The main requirements for new antagonists of $\alpha_{IIb}\beta_3$ were derived from the pharmacophore models, docking studies, and some experimental observations. They are (i) positively and negatively charged groups should be separated by at least 16 Å; (ii) lipophilic fragment should be attached to the acidic part of a molecule; and (iii) desirable that the above lipophilic fragment is linked to a H-bond acceptor able to bind the Arg214 residue of αIIbβ3. According to these rules various Arg- and Asp-mimetic fragments and different linker groups were proposed (Figure 66). A combinatorial virtual library was generated by in-house computer program. After discarding synthetically irrelevant structures, the remaining 6930 compounds (24066 stereisomers) were used for the screening.

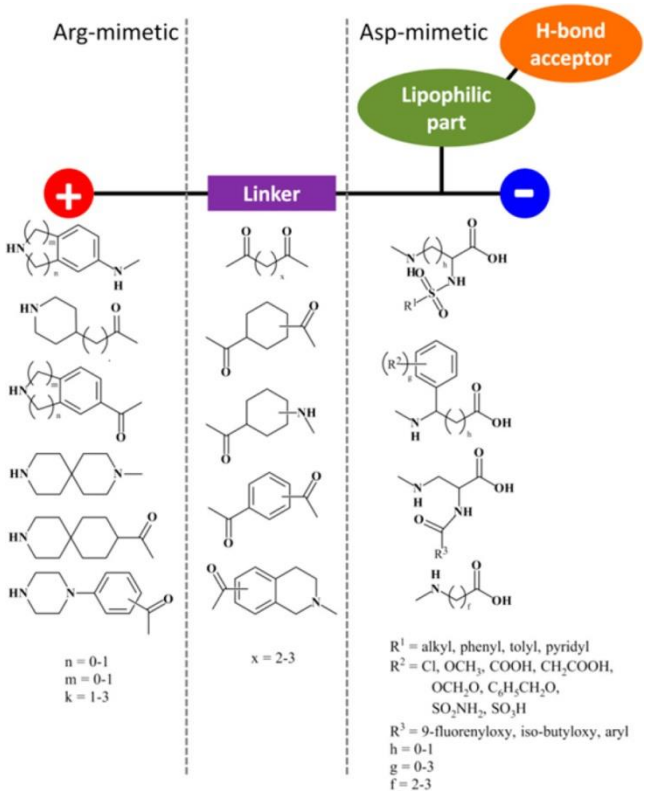


Figure 66. General design principles of the ligands for open form $\alpha_{IIb}\beta_3$ and examples of building blocks used for generation of a virtual combinatorial library.

All compounds were passed through 2D consensus QSAR model and all 3D pharmacophore models that resulted in 93 common hits with predicted affinity $pIC_{50} \geq 8$ and antiaggregation activity $pIC_{50} \geq 7$ (Figure 67). 74 compounds passed the molecular docking stage and for them we predicted water solubility and toxicity. Two compounds represented by two stereoisomers each were finally selected for synthesis.

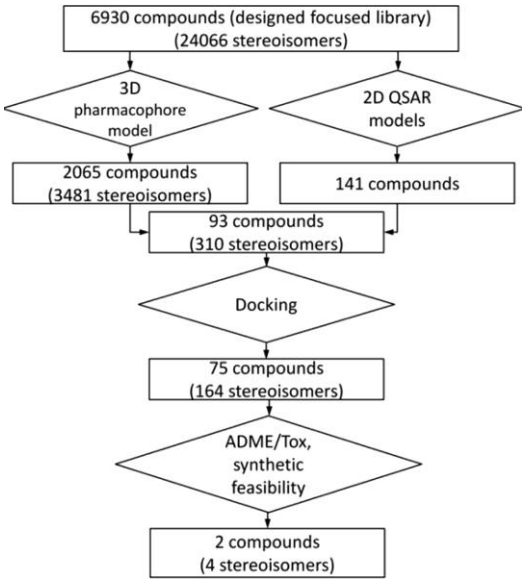
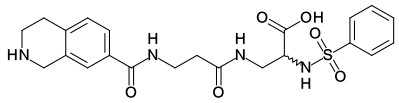
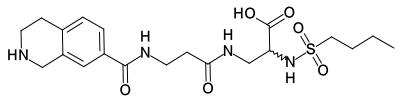
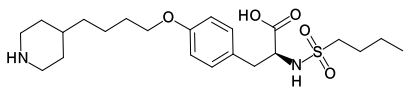


Figure 67. Workflow of the virtual screening of the focused library designed for open form $\alpha_{IIb}\beta_3$.

Both synthesized compounds demonstrated high affinity for $\alpha_{IIb}\beta_3$ and antiaggregation activity (Table 15). Compound 1 was outperformed the reference drug tirofiban. S-isomers of the designed compounds were more active than R-isomers but the difference was small. This corresponded to molecular docking which predicted close docking scores for different stereoisomers and they had very similar binding poses.

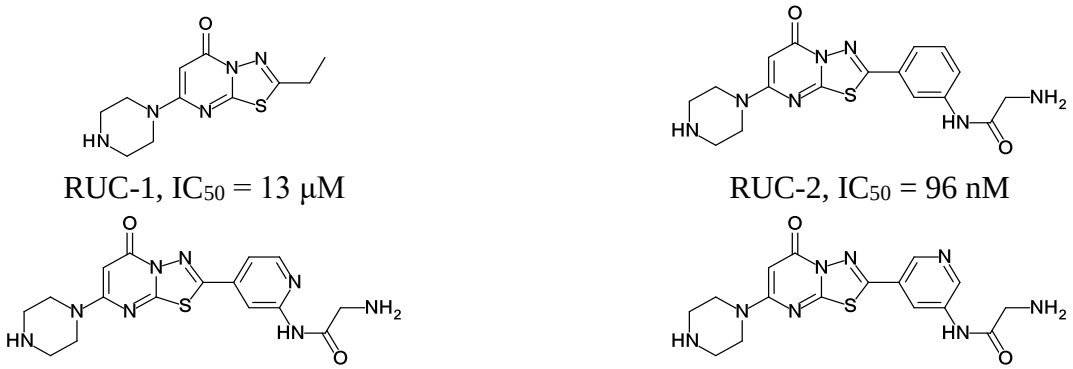
Table 15. Affinity for $\alpha_{IIb}\beta_3$ and antiaggregation activity of the designed antagonists of the open form of $\alpha_{IIb}\beta_3$ and tirofiban.

		Affinity for fibrinogen receptors, IC_{50} (nM)	Anti-aggregation activity, IC_{50} (nM)
 1	R	0.22 ± 0.01	6.2 ± 0.9
	S	0.96 ± 0.07	25.0 ± 5
 2	R	62.0 ± 9.0	320 ± 50
	S	79.0 ± 12.0	670 ± 100
 Tirofiban	S	2.4 ± 0.4	32 ± 4

6.2. Antagonists of the closed form of integrin $\alpha_{IIb}\beta_3$ as antithrombotic agents

It was suggested that thrombocytopenia caused by existing anti-platelet drugs bound to the open form of $\alpha_{IIb}\beta_3$ was an immunological response of an organism on the conformational changes in integrin αIIbβ3 upon binding with RGD-peptidomimetics[206, 207]. As a response to this, RUC-1

(Figure 68) was discovered by high-throughput screening of about 33000 small molecules[208]. It had weak antiaggregation activity (IC_{50} = 13 μM), however, according to mutagenesis studies, it binds only to the α_{IIb} subunit of the integrin. As it was shown in gel filtration and dynamic light scattering experiments, it did not induce transformations leading to open headpiece form (Figure 63, Ligand A). Later on, this was confirmed by X-ray study of the complex of RUC-1 and $\alpha_{IIb}\beta_3$ [209]. In order to explore the RUC-1 binding pocket and to obtain additional information concerning binding mechanism and induction of conformational changes in the receptor, a series of derivatives of RUC-1 have been synthesized[210]. One of them, named RUC-2 (Figure 68), was found some 100 times more potent in inhibiting ADP-induced platelet aggregation than RUC-1 (IC_{50} = 96 nM). At the same time RUC-2 does not induce any conformation changes in the $\alpha_{IIb}\beta_3$ headpiece, which may reduce adverse effects. Later, RUC-3 and RUC-4 ligands with improved antiaggregation activity were designed (Figure 68) [211].



RUC-1, IC_{50} = 13 μM

RUC-2, IC_{50} = 96 nM

RUC-3, IC_{50} = 45 nM

RUC-4, IC_{50} = 33 nM

Figure 68. Structures of $\alpha_{IIb}\beta_3$ antagonists bound to the closed headpiece of the integrin and their ADP-induced antiaggregation actiivty

Protein–ligand binding patterns in $\alpha_{IIb}\beta_3$ open and closed forms differ. Thus, Tirofiban binds with Asp224 residue of the α_{IIb} subunit and with Mg2+ ion of metal ion-dependent adhesion site at the β_3 subunit in the open form of integrin (Figure 69). RUC-2 binds to Asp224 residue of the α_{IIb} subunit but it displaces Mg^{2+} ion directly binds to Glu220 residue of the β_3 subunit. These differences are key factors determining ligands effects on the conformational state of the receptor.

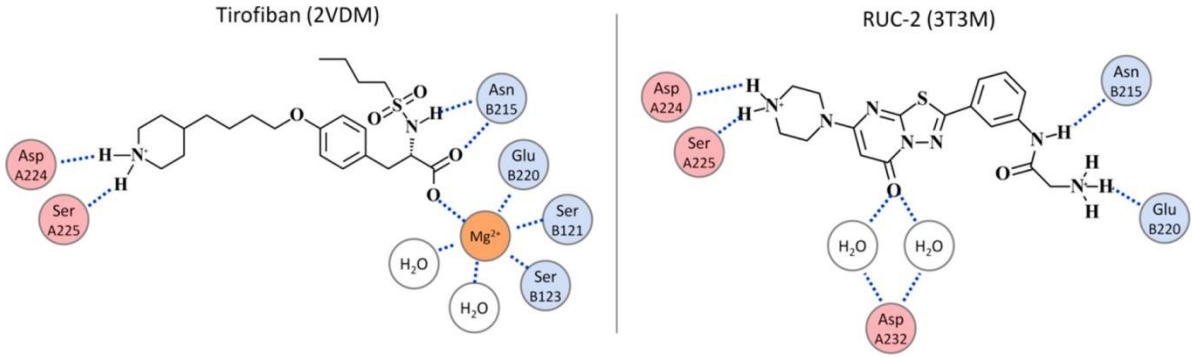


Figure 69. Interaction patterns of Tirofiban and RUC-2 compounds with integrin $\alpha_{IIb}\beta_3$ in its open (left) and closed (right) forms

Since very few experimental data on ligands for closed form were available, only structure-based pharmacophore and docking methods were used to search for new antagonists of $\alpha_{IIb}\beta_3$. A structure-based pharmacophore model (Figure 70) has been generated with LigandScout using the structure of the RUC-2-$\alpha_{IIb}\beta_3$ complex (PDB code 3T3M). This model contains: (i) two positively charged centers separated by 15.8 Å, (ii) five H-bond donors associated with positive centers, (iii) three H-bond acceptors associated with the carbonyl group of the ligand, which binds to αIIbAsp232 residue via two water molecules (see also Figure 69), (iv) H-bond donor bounded with β3Asn215, and (v) one H-bond acceptor and hydrophobic feature shifted toward one of the positive centers. This 3D pharmacophore model was translated into the additional 2D pharmacophore model which consisted of two positively charged centers separated by at least 12 bonds. This is a minimal number of bonds required to cover the distance 15.8 Å between these centers in 3D space.

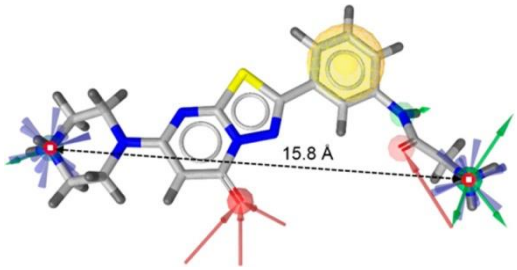


Figure 70. Pharmacophore model derived from the RUC-2-$\alpha_{IIb}\beta_3$ complex. The following labels for pharmacophore features were used: red stars, centers of negative charge; blue stars, centers of positive charge; red arrows – H-bond acceptors; green arrows – H-bond donors; yellow spheres, hydrophobic parts. Exclusion volumes are not shown for clarity.

To select more optimal docking settings we performed re-docking studies of RUC-2 ligand (3T3M) using MOE and FlexX. FlexX demonstrated better performance and could reproduce the pose of the ligand with high accuracy (RMSD = 0.78Å), whereas MOE failed to bind the ligand to Glu220 residue, which seems to be crucial for ligand-protein recognition, and resulted in higher RMSD, 2.2Å. Therefore, FlexX was selected for virtual screening.

The developed 2D and 3D pharmacophore models were used to screen several large databases of commercially available compounds: (i) advanced and HTS Enamine databases, containing 1.5 million structurally diverse compounds; (ii) REAL Enamine database, containing ~17 million synthetically feasible compounds; and (iii) ZINC database, which ensembles collections of compounds from different vendors with overall more than 17 million compounds. This resulted in 50 compounds, which have been docked with FlexX. Only two high score compounds have been selected. One of them is the known drug Nafamostat, a serine protease inhibitor[212]. This compound was also identified by Negri et al. in their structure-based virtual screening of possible $\alpha_{IIb}\beta_3$ antagonists[213]. Nafamostat has some clear drawbacks: it does not possess high antiaggregation

activity (IC_{50} = 12.5 μM), and it may have some side effects because of its ability to bind different proteins, such as thrombin, urokinase, trypsin, plasmin, etc [214-216]. It should, however, be noted that Nafamostat was introduced as an alternative anticoagulant in continuous renal replacement therapy (CRRT) in 1990, but its usage is mainly limited to Japan[217, 218]. Another selected compound was not available for purchasing at that moment.

Since no compounds were selected from the commercial databases, a small focused virtual library of RUC-2 analogues has been designed. According to 3D structure analysis, a ligand for the integrin closed form should possess: (i) a positively charged part (preferably pyperazine residue) able to interact with the Asp224 residue; (ii) a heterocyclic moiety interacting with the Tyr190 residue; (iii) an acceptor group (preferably carbonyl) interacting with the Asp232 residue, and (iv) positively charged part (amino group) displacing Mg^{2+} ion and, in such a way, providing with interactions with Glu220 residue of the β_3 subunit. Potentially, a molecule combining 6-amino-2-(piperazin-1-yl)-3H-quinazolin-4-one scaffold connected to amino-group, as it is shown in Figure 71, may fulfill these conditions. This scaffold was chosen because substituted quinazolinediones and quinazolinones derivatives were known as platelet aggregation inhibitors and fibrinogen receptor antagonists[219].

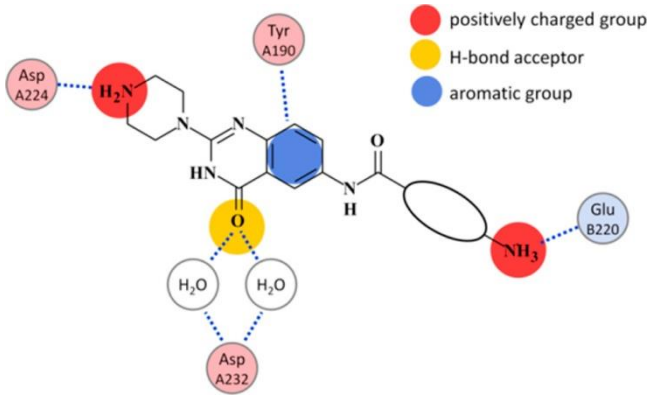
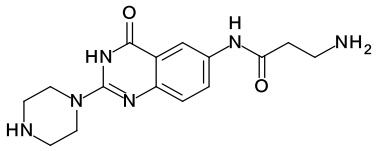
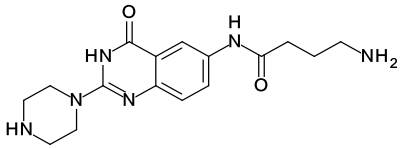
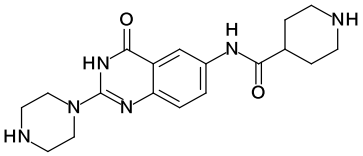
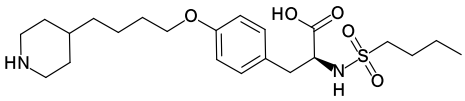


Figure 71. Schematic representation of ligands for closed form of $\alpha_{IIb}\beta_3$ used for generation of virtual focused library.

29 compounds (41 stereoisomers) were designed and screened against 3D pharmacophore models, followed by docking with FlexX and application of solubility and toxicity filters. This resulted in 20 hits, three of which were selected for the synthesis and biological tests (Table 16) [181]. All compounds had pronounced affinity for $\alpha_{IIb}\beta_3$ and antiaggregation activity. Compound 4 demonstrated antiaggregation activity even better than tirofiban and other RUC analogs: RUC-3 (IC_{50} = 45 nM) and RUC-4 (IC_{50} = 33 nM) (Figure 68) [211].

Table 16. Affinity for $\alpha_{IIb}\beta_3$ and antiaggregation activity of the designed antagonists of the closed form of $\alpha_{IIb}\beta_3$ and tirofiban.

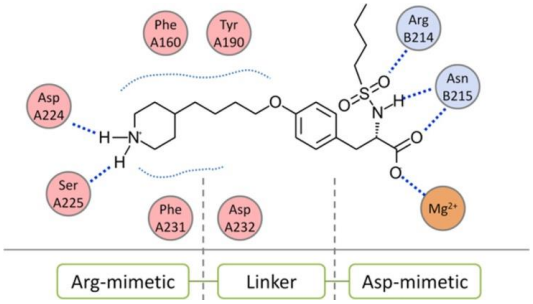
Compound	Affinity for fibrinogen receptors, IC_{50} (nM)	Anti-aggregation activity, IC_{50} (nM)
 3	5.0 ± 0.8	150 ± 25
 4	2.2 ± 0.3	11 ± 1
 5	3.8 ± 0.4	100 ± 15
 Tirofiban	2.4 ± 0.4	32 ± 4

In the separate study we further studied the structure-activity relationship by modification and replacement of the γ-aminobutyric group in the compound 4, but all designed compounds did not outperform the parent compound 4. However, we established that further elongation of the aminoacyl moiety caused drop in activity in two orders of magnitude. Replacement of γ-aminobytyric moiety in the compound 4 with δ-aminovaleric or ε-aminocaproic residues resulted in moderate antiaggregation activity, IC_{50} 1.4 μM and 1.3 μM, respectively[182]. This probably happen due to unfavorable binding entropy because these compounds comprising a long flexible poylemthylene chain lose many conformational degrees of freedom upon binding.

6.3. Global interpretation of QSAR models on the example of $\alpha_{IIb}\beta_3$ antagonist data set

Interpretation of QSAR models predicting a biological activity introduces additional complexity for finding of a general structure-activity relationship (SAR) trend (global interpretation). In such a case one cannot simply aggregate contributions of identical fragments to reveal a global SAR, because the same fragments in different parts of a molecule will bind to different amino acids and may have different contributions to the activity. Therefore, the analysis of fragment contributions should take into account information about fragment binding inside a protein binding site.

The dataset of 338 compounds with measured affinity $\alpha_{IIb}\beta_3$ collected from ChEMBL in the study described above was used to validate the ability of QSAR models to retrieve relevant patterns and establish global SAR trends. All $\alpha_{IIb}\beta_3$ antagonists could be represented by the common pattern (Figure 72). Compounds consisted of Arg- and Asp-mimetic moieties which were mainly represented by positively charged amines and negatively charged derivatives of aliphatic carboxylic acids, respectively. The linkers were mainly represented by linear or constrained cyclic aliphatic moieties. Arg-mimetics interact mainly with Asp224 on the α_{IIb} chain, where as Asp-mimetics bind to Mg^{2+} ion bound to β_3 chain of the integrin. All compounds corresponded to this pattern. Therefore, we expected the same or similar binding modes for the majority compounds.



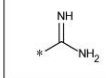
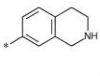
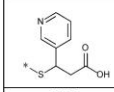
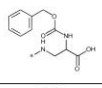
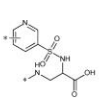
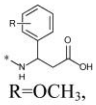
Arg-mimetic						
						
R1	R2	R3	R4	R5	R6	R7
Linker						
						
L1	L2	L3	L4	L5	L6	L7
Asp-mimetic						
						
D1	D2	D3	D4	D5	D6	D7
	 R=OCH_3, OCH_2O					
D8	D9					

Figure 72. Binding pattern of tirofiban, a commercial $\alpha_{IIb}\beta_3$ antagonist (PDB code 2VDM) (top). Most frequently occurring fragments in the data set of $\alpha_{IIb}\beta_3$ antagonists used to built QSAR models (bottom).

We built Random Forest (RF), Support Vector Machine (SVM), Gradient boosting (GBM) and partial least square (PLS) models using simplex representation of molecules structure (SiRMS) (Figure 2). Atoms in simplexes were labeled according to their partial charges (to capture electrostatic interactions with a protein), lipophilicity (hydrophobic interactions), molecular refraction (dispersive interactions) and H-bonding. All models had reasonably high accuracy estimated by 5-fold cross-validation while the consensus model has the highest accuracy (Table 17).

Table 17. 5-fold cross-validation performance of models predicting affinity for $\alpha_{IIb}\beta_3$.

	RF	GBM	SVM	PLS	consensus
R^2_{CV}	0.72	0.68	0.70	0.67	0.73
$RMSE_{CV}$	0.81	0.86	0.82	0.88	0.79

To interpret these models we applied the developed universal interpretation approach described in the Chapter 2 (Figure 15). Calculated fragments contribution from individual models had high mutual correlation ($R_{Pearson}$ = 0.89-0.98), therefore we analyzed interpretation outputs of the consensus model only. We grouped contributions of identical fragments for each part of antagonists (Asp and Arg-mimetics and linkers) to retrieve the global SAR trend captured by the model (Figure 73). The two-sided Wilcoxon rank test was applied to test the statistical significance of the contributions. However, the calculated contributions are affected by the accuracy and predictive performance of the models. For this reason, it is reasonable to compare contributions relative to a modeling error (RMSE). Contributions that are within 1 unit of RMSE may be considered insignificant, and their analysis should be done with care.

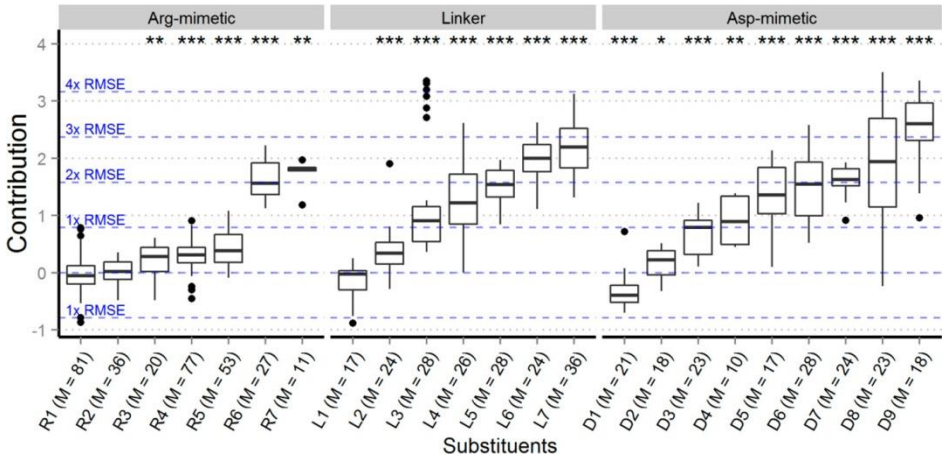


Figure 73. Distribution of fragment contributions of $\alpha_{IIb}\beta_3$ antagonists calculated from the consensus QSAR model (global interpretation). M is the number of compounds containing a given fragment. Asterisks refer to statistical significance calculated by the two-sided Wilcoxon rank test (p value): ***, $p < 0.001$; **, $p < 0.01$, *, $p < 0.05$.

Asp-mimetics were the most diverse part of the $\alpha_{IIb}\beta_3$ antagonists. Fragment D8, which occurred in many compounds, has a very large range of contribution values due to the substantial influence of molecular context. At the same time, fragment D6, which is also present in different molecular contexts, has a smaller range of contributions. In general, the variance of contribution values of Arg-mimetics is substantially smaller than those of linkers and Asp-mimetics. This indicates that the nature of the Arg-mimetic may be more important for binding to the integrin than those of the linker and Asp mimetic, whose contributions are highly context-dependent.

The global physicochemical interpretation revealed the large contribution of the electrostatic term (Figure 74), suggesting that this is the main driving force of the ligand–receptor interaction. This assumption can be supported by the following considerations: (1) ligands have at least one positively group and one negatively charged group, which is essential for ligand–receptor recognition (Figure 75); (2) there are commonly one or two charged H-bonds in the ligand–receptor complexes according to previous molecular docking studies[181, 182] (Figure 75); (3) the desolvation effect of ligands, which depends on the distribution of partial atomic charges, can also play an important role in ligand binding and cannot be estimated directly. The less significant effects of H-bonding relative to the electrostatic term may be explained by the charged nature of the H-bonds formed between the ligands and Asp224 and Arg214 of the fibrinogen receptor. There are few hydrophobic residues in the binding pocket, and correspondingly, relatively small contributions of hydrophobic effects of fragments are observed in the consensus QSAR model. The contributions of dispersive interactions are smallest, as these forces are usually very weak and do not substantially influence affinity values.

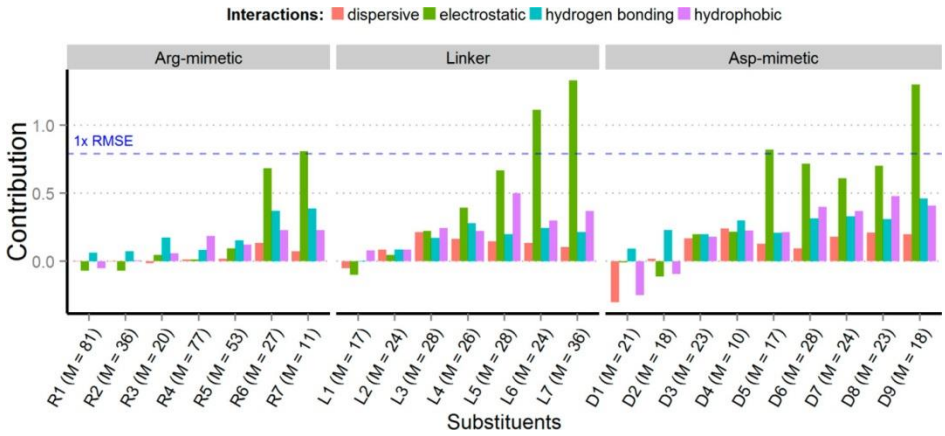


Figure 74. Median contributions of physicochemical terms in affinity for $\alpha_{IIb}\beta_3$ calculated from the consensus QSAR model (global interpretation). M is the number of compounds containing a given fragment.

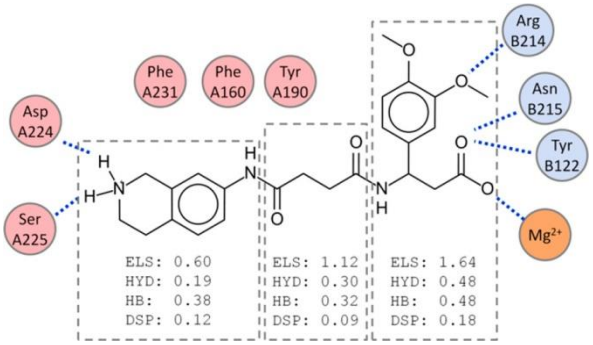


Figure 75. Interaction map of a selected ligand with the integrin $\alpha_{IIb}\beta_3$ and calculated contributions of physicochemical terms for individual fragments from the consensus model (ELS, electrostatic; HYD, hydrophobic; HB, hydrogen bonding; DSP, dispersive) (local interprettion).

These findings are in a good agreement with experimentally observed structure-activity relationships and suggest reasonable explanations of driving forces of ligand-receptor recognition without explicit knowledge about protein binding site, which is not available for QSAR models. It also emphasizes importance to consider identical fragments separately if it is expected that they bind to different parts of a binding site. Otherwise it would be impossible to get a reasonable picture of structure-activity relationships.

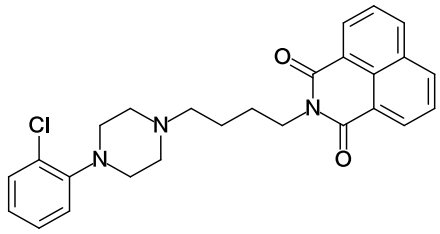
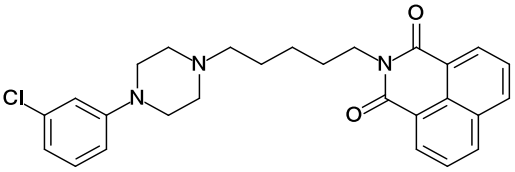
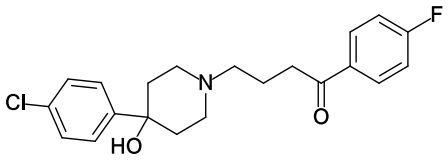
6.4. Development of compounds with anti-leichmanial activity

Leishmaniases are a group of important zoonotic diseases that are caused by various parasitic kinetoplastid species from the genus Leishmania[220]. Nineteen different Leishmania species are associated with cutaneous and mucocutaneous diseases, but only two (*L. donovani* and *L. major*) are strongly associated with the deadly visceral forms of the diseases[221]. Leishmania parasites are transmitted to humans and other mammals by an insect vector, the phlebotomine sandfly[222]. They reside as promastigotes in the insect gut[221], while in humans, they exist as non-motile intracellular amastigotes in infected macrophages[223]. Leishmaniases are currently endemic in 88 tropical and subtropical countries of Asia, Africa, southern Europe and Americas[221, 224]. They have a significant global socioeconomic impact because of their high overall prevalence, co-occurrence with HIV infection and spread in non-endemic regions. Worldwide incidence is around 12 million cases per year, and mortality is about 50,000, mainly due to visceral forms of the disease[221, 224].

Currently no effective vaccine exists and the disease can be managed only through chemotherapy using a limited set of drugs, including pentavalent antimonials, meglumine antimoniate, sodium stibogluconate, miltefosine, and amphotericin B[225, 226]. These drugs have become less effective in areas with high disease prevalence and elsewhere, and their use is also complicated by high toxicity and side effects[227, 228].

Repurposing of known biologically active compounds is one of possible strategies to identify primary hit molecules. We performed a high-throughput screening of compounds from LOPAC@1280 library. This is a well-known library consisting of compounds with different mechanisms of action. The primary screening identified 57 compounds which at concentration 50 µM inhibited the growth of *L. major* FV1-RFP promastigotes at a rate of ≥90% [185]. Twenty five compounds were ignored due to their previously published activity. From the remaining compounds 7 compounds demonstrated IC_{50} below 10 µM. However, 6 out of 7 compounds had high cytotoxicity to J774 cells used in the study and they were discarded. Only one compound, haloperidol, which is a known dopamine 2 receptor antagonist used as antipsychotic, was selected as a final hit (Table 18). However, haloperidol was inactive against intracellular amastigotes of *L major* FV1-RFP and *L. mexicana*, suggesting that its prospects as an antileishmanial drug are negligible.

Table 18. Anti-leishmanial activity of reference compounds and compounds discovered in the study.

Compound	IC_{50}, μM			
	Promastigotes		Amastigotes	
	L. major	*L. mexicana*	*L. major*	*L. mexicana*
 6	16.98 ± 0.75	18.26 ± 1.83	not tested	not tested
 7	3.42 ± 0.80	2.94 ± 0.06	9.41 ± 1.79	13.29 ± 1.99
 Haloperidol	8.45 ± 0.83	9.18 ± 1.83	inactive	inactive
Amphotericin B	0.40 ± 0.04	0.39 ± 0.08	0.71 ± 0.14	0.60 ± 0.10

To find analogs of haloperidol we performed similarity search in our proprietary library using 2D pharmacophore fingerprints implemented in RDKit and identified 11 compounds having similarity greater than 0.6[185]. Eight compounds were 1-aryl-4-(phthalimidoalkyl)piperazines[229] and three 1-aryl-4-(naphthalimidoalkyl)piperazines[230]. Two of those compounds, 6 and 7, demonstrated activity against promastigotes (Table 18) and both of them were not cytotoxic. Experiments against intracellular amastigotes of *L major* FV1-RFP and *L. mexicana* in infected J774 cells showed that compound 7 inhibited their growth in dose-dependent manner with IC_{50} 9.41 ± 1.79 μM and 13.29 ± 1.99 μM, respectively. In additional experiments it was found that compound 7 did not affect plasma membrane integrity, induced collapse the mitochondrial electrochemical potential and caused increase in the intracellular Ca^{2+} concentration in promastigotes of *L. major* and *L. mexicana*. This suggested that activity of compound 11 could be directly associated with the depolarization of the mitochondrial membrane and not to detergent-like effects of membrane acting drugs associated with rapid disruption of the plasma membrane. While for the compound 7 activity against serotonin 1a receptor was reported, no anxiolytic activity was observed.[229] All these make compound 7 promising for further studies to identify its molecular targets and perform structural optimization.

6.5. Summary

Here we demonstrated how comprehensive virtual screening pipelines in combination with carefully designed enumerated virtual libraries could help to identify potent biologically active compounds. Within our studies we found highly active $\alpha_{IIb}\beta_3$ antagonists which may bind to closed or open form of the $\alpha_{IIb}\beta_3$ headpiece. The latter may help to avoid side effects of existing drugs, in particular thrombocytopenia. Some of the found compounds outperformed the reference drug tirofiban as well as other published analogs.

In a separate study we retrospectively estimated the ability of QSAR model to reveal general trends in structure-activity relationships for $\alpha_{IIb}\beta_3$ antagonists. We demonstrated that the previously suggested structural and physicochemical interpretation approach could identify relevant structural motifs and explain the main reason of their activity, electrostatic interactions, that is in a good agreement with experimentally observed structure-activity relationships.

In the last exemplar study we showed that in the era of deep learning a simple similarity search can be a powerful tool to find new promising compounds. Within the study of searching for new anti-leishmania agents we identified that haloperidol was active against extracellular forms but inactive against intracellular forms of *Leishmania*. The similarity search could identify a promising analog of haloperidol with improved biological activity.

General conclusions

This work presents a comprehensive overview of research activities conducted over the past eleven years in which the author actively participated. The primary focus of these investigations revolved around the advancement of methodological techniques within the domain of machine learning applied to chemoinformatics. This contribution encompasses several key facets.

The first among these achievements is the refinement of the simplex representation of molecular structures for encoding mixtures of organic molecules with arbitrary compositions. This enhancement facilitated the prediction of properties of pure chemicals which the "quasi"-mixture approach as well as rate constants of chemical reactions.

A second significant accomplishment was the development of various approaches for the interpretation of QSAR/QSPR models. Notably, the author introduced the universal interpretation approach, enabling the direct estimation of the contributions of atoms and fragments from a model. This innovation introduced a novel paradigm for QSAR interpretation and permitted structural interpretation of any model, irrespective of the employed machine learning methods and descriptors. Analogous approaches, founded on atom or fragment "masking", have gained popularity, especially with the emergence of deep learning models that were commonly regarded as black boxes. Additionally, the author initiated the development of the first benchmark for validating interpretation approaches, which has been already adopted by other researchers.

The author's contributions to machine learning in chemoinformatics also encompass the exploration of various multi-instance learning approaches, a field that had long been neglected. It was demonstrated that, in multiple cases, multi-conformer models outperformed conventional 2D models in predicting biological activity of compounds and enantioselectivity of catalysts. The multi-instance paradigm, coupled with modern deep learning techniques, promises to advance predictive modeling in chemoinformatics.

A separate research domain in which the author made notable contributions is pharmacophore representation and modeling. The implementation of a representation based on stereosensitive quadruplets enabled the development of a ligand-based pharmacophore modeling approach suitable for large datasets and pharmacophore modeling based on molecular dynamics trajectories of ligand-protein complexes. This representation approach was particularly beneficial in the context of multi-conformer modeling using multi-instance techniques.

The most recent contribution field pertains to structure generation and de novo design. The author's fragment-based approach, known as CReM, offers the generation of chemically valid structures while allowing control over their synthetic accessibility, a distinct advantage over prior fragment-based methodologies. Comparative studies confirmed the competitiveness of this approach with modern structure generators relying on neural networks. Furthermore, the CReM framework has

been integrated into several third-party tools for de novo design, and ongoing work focuses on its integration with QSAR, pharmacophore, and docking approaches to address diverse drug design objectives such as scaffold decoration, hit expansion, lead optimization, etc.

All the methodologies and tools developed as part of this work have been disseminated as open-source Python libraries (Annex I), contributing to open science and facilitating further advancements in the field of chemoinformatics and drug design. Recent statistics indicate a substantial uptake, with more than 4,000 monthly downloads of these libraries. Certain tools are accessible as web applications.

The author's personal commitment extends to the prospective validation of these tools in real-world projects, aiming to enhance their practical utility. This endeavor has led to successful applications of comprehensive virtual screening pipelines as well as straightforward similarity searches in the discovery of potent biologically active compounds. The author anticipates that the application of the newly developed tools will further enhance drug design pipelines.

Inevitably, this work raises pertinent questions for future investigations. What is the potential utility of contemporary QSAR interpretation methodologies? How can we formulate more effective criteria for evaluating the interpretability of models? In what ways can we enhance current interpretation techniques? To what extent can multi-instance learning be applied effectively within the field of chemoinformatics, and how can we consistently identify key instances in this context? How large is the synthetically accessible drug-like chemical space, and to what extent can it be encompassed through de novo generation methods? Answers to these and other questions await further exploration in subsequent research endeavors.

Annex I. The list of developed software and repositories

Machine learning	
sirms – 2D descriptors for single compounds, “quasi”-mixtures, mixtures and reactions	https://github.com/DrrDom/sirms
spci – automatic QSAR model building and interpretation with GUI	https://github.com/DrrDom/spci
rspci – R package to analyze fragment contributions from spci output	https://github.com/DrrDom/rspci
ibenchmark – benchmark interpretability of machine learning models	https://github.com/ci-lab-cz/ibenchmark
3D pharmacophore modeling	
pmapper – 3D pharmacophore processing, signatures and fingerprints	https://github.com/DrrDom/pmapper
psearch – automated 3D ligand-based modeling and screening	https://github.com/meddwl/psearch
pharmd – retrieve 3D pharmacophores from MD trajectories and screening	https://github.com/ci-lab-cz/pharmd
De novo design	
CReM - Python module for structure generation	https://github.com/DrrDom/crem
Automated pipelines	
easydock – Python module to run automatic molecular docking using vina, smina and gnina on distributed systems	https://github.com/ci-lab-cz/easydock
StreaMD – automated pipeline for high-throughput MD simulations	https://github.com/ci-lab-cz/md-scripts
Auxiliary RDKit repositories	
Various RDKit scripts	https://github.com/DrrDom/rdkit-scripts
Scripts to create local databases for similarity and substructure search using RDKit and Chemicalite	https://github.com/DrrDom/chemicalite-scripts

References

(1) Fujita, T. Recent Success Stories Leading to Commercializable Bioactive Compounds with the Aid of Traditional QSAR Procedures. *Quantitative Structure-Activity Relationships* **1997**, *16* (2), 107-112.
(2) Muratov, E. N.; Bajorath, J.; Sheridan, R. P.; Tetko, I. V.; Filimonov, D.; Poroikov, V.; Oprea, T. I.; Baskin, I. I.; Varnek, A.; Roitberg, A.; et al. QSAR without borders. *Chem. Soc. Rev.* **2020**.
(3) Verhaar, H. J. M.; Busser, F. J. M.; Hermens, J. L. M. Surrogate Parameter for the Baseline Toxicity Content of Contaminated Water: Simulating the Bioconcentration of Mixtures of Pollutants and Counting Molecules. *Environmental Science & Technology* **1995**, *29* (3), 726-734.
(4) Wei, D. B.; Zhai, L. H.; Hu, H. Y. QSAR-based toxicity classification and prediction for single and mixed aromatic compounds. *SAR QSAR Environ. Res.* **2004**, *15* (3), 207-216.
(5) Wang, C.; Lu, G.; Tang, Z.; Guo, X. Quantitative structure-activity relationships for joint toxicity of substituted phenols and anilines to Scenedesmus obliquus. *Journal of Environmental Sciences* **2008**, *20* (1), 115-119.
(6) Ajmani, S.; Rogers, S. C.; Barley, M. H.; Livingstone, D. J. Application of QSPR to Mixtures. *J. Chem. Inf. Model.* **2006**, *46* (5), 2043-2055.
(7) Zhang, L.; Zhou, P.-j.; Yang, F.; Wang, Z.-d. Computer-based QSARs for predicting mixture toxicity of benzene and its derivatives. *Chemosphere* **2007**, *67* (2), 396-401.
(8) Ajmani, S.; Rogers, S. C.; Barley, M. H.; Burgess, A. N.; Livingstone, D. J. Characterization of Mixtures Part 1: Prediction of Infinite-Dilution Activity Coefficients Using Neural Network-Based QSPR Models. *QSAR & Combinatorial Science* **2008**, *27* (11-12), 1346-1361.
(9) Ajmani , S.; Rogers, S. C.; Barley, M. H.; Burgess, A. N.; Livingstone , D. J. Characterization of Mixtures. Part 2: QSPR Models for Prediction of Excess Molar Volume and Liquid Density Using Neural Networks. *Mol. Inf.* **2010**, *29* (8-9), 645-653.
(10) Varnek, A.; Fourches, D.; Hoonakker, F.; Solov'ev, V. P. Substructural fragments: an universal language to encode reactions, molecular and supramolecular structures. *J. Comput.-Aided Mol. Des.* **2005**, *19* (9), 693-703.
(11) Hoonakker, F.; Lachiche, N.; Varnek, A.; Wagner, A. A representation to apply usual data mining techniques to chemical reactions - illustration on the rate constant of SN2 reactions in water. *International Journal on Artificial Intelligence Tools* **2011**, *20* (02), 253-270.
(12) Schneider, N.; Lowe, D. M.; Sayle, R. A.; Landrum, G. A. Development of a Novel Fingerprint for Chemical Reactions and Its Application to Large-Scale Reaction Classification and Similarity. *J. Chem. Inf. Model.* **2015**, *55* (1), 39-53.
(13) Kravtsov, A. A.; Karpov, P. V.; Baskin, I. I.; Palyulin, V. A.; Zefirov, N. S. Prediction of rate constants of S_N2 reactions by the multicomponent QSPR method. *Doklady Chemistry* **2011**, *440* (2), 299-301.
(14) Kuz'min, V. E.; Artemenko, A. G.; Polischuk, P. G.; Muratov, E. N.; Khromov, A. I.; Liahovskiy, A. V.; Andronati, S. A.; Makan, S. Y. Hierarchic System of QSAR Models (1D-4D) on the Base of Simplex Representation of Molecular Structure *J. Mol. Modeling* **2005**, *11*, 457-467.
(15) Kuz'min, V. E.; Artemenko, A. G.; Muratov, E. N. Hierarchical QSAR technology based on the Simplex representation of molecular structure. *J. Comput.-Aided Mol. Des.* **2008**, *22* (6), 403-421.
(16) Oprisiu, I.; Varlamova, E.; Muratov, E.; Artemenko, A.; Marcou, G.; Polishchuk, P.; Kuz'min, V.; Varnek, A. QSPR Approach to Predict Nonadditive Properties of Mixtures. Application to Bubble Point Temperatures of Binary Mixtures of Liquids. *Mol. Inf.* **2012**, *31* (6-7), 491-502.
(17) Kang, J. W.; Yoo, K. P.; Kim, H. Y.; Lee, H.; Yang, D. R.; Lee, C. S. Development and Current Status of the Korea Thermophysical Properties Databank (KDB). *International Journal of Thermophysics* **2001**, *22* (2), 487-494.
(18) Breiman, L. Random Forests. *Machine Learning* **2001**, *45* (1), 5-32.
(19) Palm, V. *Tables of rate and equilibrium constants of heterolytic organic reactions*; 1974-1978.
(20) Varnek, A.; Fourches, D.; Horvath, D.; Klimchuk, O.; Gaudin, C.; Vayer, P.; Solov'ev, V.; Hoonakker, F.; Tetko, V. I.; Marcou, G. ISIDA - Platform for Virtual Screening Based on Fragment and Pharmacophoric Descriptors. *Current Computer-Aided Drug Design* **2008**, *4* (3), 191-198.
(21) *RDKit: Open-Source Cheminformatics Software*; 2021. http://rdkit.org/ (accessed.
(22) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom pairs as molecular features in structure - activity studies. Definition and application. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 64-73.

(23) Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **2010**, *50*.
(24) Nilakantan, R.; Bauman, N.; Dixon, J. S.; Venkataraghavan, R. Topological torsion: a new molecular descriptor for SAR applications. Comparison with other descriptors. *J. Chem. Inf. Comput. Sci.* **1987**, *27* (2), 82-85.
(25) Hanser, T.; Jauffret, P.; Gruber, E.; Marchaland, J. F.; Ellermann, L.; Fang, T.; Kaufmann, G. CGS: Condensed graph of synthesis, a new computer representation for molecules, reactions and syntheses. In *AIP Conference Proceedings*, 1995; Vol. 330.
(26) Polishchuk, P.; Madzhidov, T.; Gimadiev, T.; Bodrov, A.; Nugmanov, R.; Varnek, A. Structure–reactivity modeling using mixture-based representation of chemical reactions. *J. Comput.-Aided Mol. Des.* **2017**, *31* (9), 829-839.
(27) Yang, S.; Lu, W.; Chen, N.; Hu, Q. Support vector regression based QSPR for the prediction of some physicochemical properties of alkyl benzenes. *Journal of Molecular Structure: THEOCHEM* **2005**, *719* (1), 119-127.
(28) Kazakov, A.; Muzny, C. D.; Diky, V.; Chirico, R. D.; Frenkel, M. Predictive correlations based on large experimental datasets: Critical constants for pure compounds. *Fluid Phase Equilibria* **2010**, *298* (1), 131-142.
(29) Turner, B. E.; Costello, C. L.; Jurs, P. C. Prediction of Critical Temperatures and Pressures of Industrially Important Organic Compounds from Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1998**, *38* (4), 639-645.
(30) Katritzky, A. R.; Maran, U.; Lobanov, V. S.; Karelson, M. Structurally Diverse Quantitative Structure−Property Relationship Correlations of Technologically Relevant Physical Properties. *J. Chem. Inf. Comput. Sci.* **2000**, *40* (1), 1-18.
(31) Mokshyna, E.; Nedostup, V. I.; Polischuk, P. G.; Kuzmin, V. E. 'Quasi-Mixture' Descriptors for QSPR Analysis of Molecular Macroscopic Properties. The Critical Properties of Organic Compounds. *Mol. Inf.* **2014**, *33* (10), 647-654.
(32) Reid, R. C.; Prausnitz, J. M.; Poling, B. E. *The Properties of Gases and Liquids*; McGraw-Hill, 1987.
(33) Chatterjee, M.; Roy, K. Application of cross-validation strategies to avoid overestimation of performance of 2D-QSAR models for the prediction of aquatic toxicity of chemical mixtures. *SAR QSAR Environ. Res.* **2022**, *33* (6), 463-484.
(34) Oprisiu, I.; Novotarskyi, S.; Tetko, I. V. Modeling of non-additive mixture properties using the Online CHEmical database and Modeling environment (OCHEM). *J. Cheminf.* **2013**, *5* (1), 4.
(35) Rakhimbekova, A.; Akhmetshin, T. N.; Minibaeva, G. I.; Nugmanov, R. I.; Gimadiev, T. R.; Madzhidov, T. I.; Baskin, I. I.; Varnek, A. Cross-validation strategies in QSPR modelling of chemical reactions. *SAR QSAR Environ. Res.* **2021**, *32* (3), 207-219.
(36) Varlamova, E. V.; Mileyko, S.; Tropsha, A.; Artemenko, A. G.; Fourches, D.; Muratov, E. N.; Muratov, N. N.; Kuzmin, V. E. "Everything Out" Validation Approach for Qsar Models of Chemical Mixtures *JSM Clinical Pharmaceutics* **2014**, *1* (1), 1005.
(37) Polischuk, P. G.; Muratov, E. N.; Artemenko, A. G.; Kolumbin, O. G.; Muratov, N. N.; Kuz'min, V. E. Application of Random Forest method to QSAR prediction of aquatic toxicity. *J. Chem. Inf. Model.* **2009**, *49*, 2481-2488.
(38) Kuz'min, V. E.; Polishchuk, P. G.; Artemenko, A. G.; Andronati, S. A. Interpretation of QSAR models based on Random Forest method. *Mol. Inf.* **2011**, *30*, 593-603.
(39) Wold, S.; Ruhe, A.; Wold, H.; III, W. J. D. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM J. Sci. Stat. Comput.* **1984**, *5*, 735-743.
(40) Gaillard, P.; Carrupt, P.-A.; Testa, B.; Schambel, P. Binding of Arylpiperazines, (Aryloxy)propanolamines, and Tetrahydropyridylindoles to the 5-HT1A Receptor: Contribution of the Molecular Lipophilicity Potential to Three-Dimensional Quantitative Structure−Affinity Relationship Models. *J. Med. Chem.* **1996**, *39* (1), 126-134.
(41) Nowak, M.; Kołaczkowski, M.; Pawłowski, M.; Bojarski, A. J. Homology Modeling of the Serotonin 5-HT1A Receptor Using Automated Docking of Bioactive Compounds with Defined Geometry. *J. Med. Chem.* **2006**, *49* (1), 205-214.
(42) López-Rodríguez, M. L.; Rosado, M. L.; Benhamú, B.; Morcillo, M. J.; Fernández, E.; Schaper, K.-J. Synthesis and Structure−Activity Relationships of a New Model of Arylpiperazines. 2. Three-Dimensional Quantitative Structure−Activity Relationships of Hydantoin−Phenylpiperazine Derivatives with Affinity for 5-HT1A and α1 Receptors. A Comparison of CoMFA Models. *J. Med. Chem.* **1997**, *40* (11), 1648-1656.
(43) López-Rodríguez, M. L.; Morcillo, M. J.; Fernández, E.; Rosado, M. L.; Pardo, L.; Schaper, K.-J. Synthesis and Structure−Activity Relationships of a New Model of Arylpiperazines. 6. Study of the 5-HT1A/α1-

Adrenergic Receptor Affinity by Classical Hansch Analysis, Artificial Neural Networks, and Computational Simulation of Ligand Recognition. *J. Med. Chem.* **2001**, *44* (2), 198-207.
(44) van Steen, B. J.; van Wijngaarden, I.; Tulp, M. T. M.; Soudijn, W. Structure-Affinity Relationship Studies on 5-HT1A receptor Ligands. 2. Heterobicyclic Phenylpiperazines with N4-Aralkyl Substituents. *J. Med. Chem.* **1994**, *37* (17), 2761-2773.
(45) Hibert, M. F.; Gittos, M. W.; Middlemiss, D. N.; Mir, A. K.; Fozard, J. R. Graphics computer-aided receptor mapping as a predictive tool for drug design: development of potent, selective, and stereospecific ligands for the 5-HT1A receptor. *J. Med. Chem.* **1988**, *31* (6), 1087-1093.
(46) Mellin, C.; Vallgaarda, J.; Nelson, D. L.; Bjoerk, L.; Yu, H.; Anden, N. E.; Csoeregh, I.; Arvidsson, L. E.; Hacksell, U. A 3 dimensional model for 5-HT1A-receptor agonists based on stereoselective methyl-substituted and conformationally restricted analogs of 8-hydroxy-2-(dipropylamino)tetralin. *J. Med. Chem.* **1991**, *34* (2), 497-510.
(47) Chilmonczyk, Z.; Szelejewska-Wozniakowska, A.; Cybulski, J.; Cybulski, M.; Koziol, A. E.; Gdaniec, M. Conformational Flexibility of Serotonin1A Receptor Ligands from Crystallographic Data. Updated Model of the Receptor Pharmacophore. *Archiv der Pharmazie* **1997**, *330* (5), 146-160.
(48) Palczewska, A.; Palczewski, J.; Marchese Robinson, R.; Neagu, D. Interpreting Random Forest Classification Models Using a Feature Contribution Method. In *Integration of Reusable Systems*, Bouabana-Tebibel, T., Rubin, S. H. Eds.; Springer International Publishing, 2014; pp 193-218.
(49) Polishchuk, P. Interpretation of Quantitative Structure–Activity Relationship Models: Past, Present, and Future. *J. Chem. Inf. Model.* **2017**, *57* (11), 2618-2639.
(50) Polishchuk, P. G.; Kuz'min, V. E.; Artemenko, A. G.; Muratov, E. N. Universal Approach for Structural Interpretation of QSAR/QSPR Models. *Mol. Inf.* **2013**, *32* (9-10), 843-853.
(51) Polishchuk, P.; Tinkov, O.; Khristova, T.; Ognichenko, L.; Kosinskaya, A.; Varnek, A.; Kuz'min, V. Structural and Physico-Chemical Interpretation (SPCI) of QSAR Models and Its Comparison with Matched Molecular Pair Analysis. *J. Chem. Inf. Model.* **2016**, *56* (8), 1455-1469.
(52) Free, S. M.; Wilson, J. W. A Mathematical Contribution to Structure-Activity Studies. *J. Med. Chem.* **1964**, *7* (4), 396-399.
(53) Hitchcock, S. A.; Pennington, L. D. Structure−Brain Exposure Relationships. *J. Med. Chem.* **2006**, *49* (26), 7559-7583.
(54) Wager, T. T.; Chandrasekaran, R. Y.; Hou, X.; Troutman, M. D.; Verhoest, P. R.; Villalobos, A.; Will, Y. Defining Desirable Central Nervous System Drug Space through the Alignment of Molecular Properties, in Vitro ADME, and Safety Attributes. *ACS Chemical Neuroscience* **2010**, *1* (6), 420-434.
(55) Wager, T. T.; Hou, X.; Verhoest, P. R.; Villalobos, A. Moving beyond Rules: The Development of a Central Nervous System Multiparameter Optimization (CNS MPO) Approach To Enable Alignment of Druglike Properties. *ACS Chemical Neuroscience* **2010**, *1* (6), 435-449.
(56) Ghose, A. K.; Herbertz, T.; Hudkins, R. L.; Dorsey, B. D.; Mallamo, J. P. Knowledge-Based, Central Nervous System (CNS) Lead Selection and Lead Optimization for CNS Drug Discovery. *ACS Chemical Neuroscience* **2012**, *3* (1), 50-68.
(57) Matveieva, M.; Cronin, M. T. D.; Polishchuk, P. Interpretation of QSAR Models: Mining Structural Patterns Taking into Account Molecular Context. *Mol. Inf.* **2018**, *38*, 1800084.
(58) Bietz, S.; Schomburg, K. T.; Hilbig, M.; Rarey, M. Discriminative Chemical Patterns: Automatic and Interactive Design. *J. Chem. Inf. Model.* **2015**, *55* (8), 1535-1546.
(59) Schultz, T. W.; Cronin, M. T. D.; Netzeva, T. I.; Aptula, A. O. Structure−Toxicity Relationships for Aliphatic Chemicals Evaluated with Tetrahymena pyriformis. *Chemical Research in Toxicology* **2002**, *15* (12), 1602-1609.
(60) Enoch, S. J.; Ellison, C. M.; Schultz, T. W.; Cronin, M. T. D. A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity. *Critical Reviews in Toxicology* **2011**, *41* (9), 783-802.
(61) Jiménez-Luna, J.; Grisoni, F.; Schneider, G. Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence* **2020**, *2* (10), 573-584.
(62) Bach, S.; Binder, A.; Montavon, G.; Klauschen, F.; Müller, K.-R.; Samek, W. On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLOS ONE* **2015**, *10* (7), e0130140.
(63) Shrikumar, A.; Greenside, P.; Kundaje, A. Learning important features through propagating activation differences. In Proceedings of the 34th International Conference on Machine Learning - Volume 70, Sydney, NSW, Australia; 2017.

(64) Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning Deep Features for Discriminative Localization. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 27-30 June 2016, 2016; pp 2921-2929.
(65) Selvaraju, R. R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*, 22-29 Oct. 2017, 2017; pp 618-626.
(66) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California, USA; 2017.
(67) McCloskey, K.; Taly, A.; Monti, F.; Brenner, M. P.; Colwell, L. J. Using attribution to decode binding mechanism in neural network models for chemistry. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (24), 11624-11629.
(68) Štrumbelj, E.; Kononenko, I. Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems* **2014**, *41* (3), 647-665.
(69) Rodríguez-Pérez, R.; Bajorath, J. Interpretation of Compound Activity Predictions from Complex Machine Learning Models Using Local Approximations and Shapley Values. *J. Med. Chem.* **2020**, *63* (16), 8761-8777.
(70) Webel, H. E.; Kimber, T. B.; Radetzki, S.; Neuenschwander, M.; Nazaré, M.; Volkamer, A. Revealing cytotoxic substructures in molecules using deep learning. *J. Comput.-Aided Mol. Des.* **2020**, *34* (7), 731-746.
(71) Tang, B.; Kramer, S. T.; Fang, M.; Qiu, Y.; Wu, Z.; Xu, D. A self-attention based message passing neural network for predicting molecular lipophilicity and aqueous solubility. *J. Cheminf.* **2020**, *12* (1), 15.
(72) Sheridan, R. P. Interpretation of QSAR Models by Coloring Atoms According to Changes in Predicted Activity: How Robust Is It? *J. Chem. Inf. Model.* **2019**.
(73) Riniker, S.; Landrum, G. A. Similarity maps - a visualization strategy for molecular fingerprints and machine-learning methods. *J. Cheminf.* **2013**, *5* (1), 43.
(74) Matveieva, M.; Polishchuk, P. Benchmarks for interpretation of QSAR models. *J. Cheminf.* **2021**, *13* (1), 41.
(75) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29* (3), 163-172.
(76) Cheng, T.; Zhao, Y.; Li, X.; Lin, F.; Xu, Y.; Zhang, X.; Li, Y.; Wang, R.; Lai, L. Computation of Octanol–Water Partition Coefficients by Guiding an Additive Model with Knowledge. *J. Chem. Inf. Model.* **2007**, *47* (6), 2140-2148.
(77) *Deepchem*. https://github.com/deepchem/deepchem (accessed.
(78) Robinson, M. C.; Glen, R. C.; Lee, A. A. Validating the validation: reanalyzing a large-scale comparison of deep learning and machine learning models for bioactivity prediction. *J. Comput.-Aided Mol. Des.* **2020**, *34* (7), 717-730.
(79) Chen, Q.; Zhang, Y.; Gao, P.; Zhang, J. An Interpretable Graph Representation Learning Model for Accurate Predictions of Drugs Aqueous Solubility. *Artificial Intelligence Chemistry* **2023**, 100010.
(80) Fujita, T.; Winkler, D. A. Understanding the Roles of the "Two QSARs". *J. Chem. Inf. Model.* **2016**, *56* (2), 269-274.
(81) Langer, T.; Wolber, G. Pharmacophore definition and 3D searches. *Drug Discovery Today: Technologies* **2004**, *1* (3), 203-207.
(82) Chen, Z.; Li, H.-l.; Zhang, Q.-j.; Bao, X.-g.; Yu, K.-q.; Luo, X.-m.; Zhu, W.-l.; Jiang, H.-l. Pharmacophore-based virtual screening versus docking-based virtual screening: a benchmark comparison against eight targets. *Acta Pharmacologica Sinica* **2009**, *30*, 1694.
(83) Hessler, G.; Baringhaus, K.-H. The scaffold hopping potential of pharmacophores. *Drug Discovery Today: Technologies* **2010**, *7* (4), e263-e269.
(84) Schuster, D. 3D pharmacophores as tools for activity profiling. *Drug Discovery Today: Technologies* **2010**, *7* (4), e205-e211.
(85) Schaller, D.; Šribar, D.; Noonan, T.; Deng, L.; Nguyen, T. N.; Pach, S.; Machalz, D.; Bermudez, M.; Wolber, G. Next generation 3D pharmacophore modeling. *WIREs Computational Molecular Science* **2020**, *10* (4), e1468.
(86) Vuorinen, A.; Schuster, D. Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling. *Methods* **2015**, *71*, 113-134.

(87) Koes, D. R.; Camacho, C. J. Pharmer: Efficient and Exact Pharmacophore Search. *J. Chem. Inf. Model.* **2011**, *51* (6), 1307-1314.
(88) Taminau, J.; Thijs, G.; De Winter, H. Pharao: Pharmacophore alignment and optimization. *J. Mol. Graphics Modell.* **2008**, *27* (2), 161-169.
(89) Sunseri, J.; Koes, D. R. Pharmit: interactive exploration of chemical space. *Nucleic Acids Research* **2016**, *44* (W1), W442-W448.
(90) Schneidman-Duhovny, D.; Dror, O.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. PharmaGist: a webserver for ligand-based pharmacophore detection. *Nucleic Acids Research* **2008**, *36* (suppl_2), W223-W228.
(91) Wang, X.; Shen, Y.; Wang, S.; Li, S.; Zhang, W.; Liu, X.; Lai, L.; Pei, J.; Li, H. PharmMapper 2017 update: a web server for potential drug target identification with a comprehensive target pharmacophore database. *Nucleic Acids Research* **2017**, *45* (W1), W356-W360.
(92) Schaller, D.; Pach, S.; Wolber, G. PyRod: Tracing Water Molecules in Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2019**, *59* (6), 2818-2829.
(93) Lee, J. Y.; Krieger, J. M.; Li, H.; Bahar, I. Pharmmaker: Pharmacophore modeling and hit identification based on druggability simulations. *Protein Science* **2020**, *29* (1), 76-86.
(94) Kutlushina, A.; Khakimova, A.; Madzhidov, T.; Polishchuk, P. Ligand-Based Pharmacophore Modeling Using Novel 3D Pharmacophore Signatures. *Molecules* **2018**, *23* (12), 3094.
(95) Mason, J. S.; Morize, I.; Menard, P. R.; Cheney, D. L.; Hulme, C.; Labaudiniere, R. F. New 4-Point Pharmacophore Method for Molecular Similarity and Diversity Applications: Overview of the Method and Applications, Including a Novel Approach to the Design of Combinatorial Libraries Containing Privileged Substructures. *J. Med. Chem.* **1999**, *42* (17), 3251-3264.
(96) Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *Journal of Chemical Documentation* **1965**, *5* (2), 107-113.
(97) Butina, D. Unsupervised data base clustering based on daylight's fingerprint and tanimoto similarity: a fast and automated way to cluster small and large data sets. *J Chem Inf Comput Sci* **1999**, *39*.
(98) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17* (5-6), 490-519.
(99) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; et al. Chembl: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res* **2012**, *40*.
(100) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99* (6), 1129-1143.
(101) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59* (9), 4035-4061.
(102) Liu, X.; Shi, D.; Zhou, S.; Liu, H.; Liu, H.; Yao, X. Molecular dynamics simulations and novel drug discovery. *Expert Opinion on Drug Discovery* **2018**, *13* (1), 23-37.
(103) Hu, B.; Lill, M. A. Protein Pharmacophore Selection Using Hydration-Site Analysis. *J. Chem. Inf. Model.* **2012**, *52* (4), 1046-1060.
(104) Yu, W.; Lakkaraju, S. K.; Raman, E. P.; MacKerell, A. D. Site-Identification by Ligand Competitive Saturation (SILCS) assisted pharmacophore modeling. *J. Comput.-Aided Mol. Des.* **2014**, *28* (5), 491-507.
(105) Jung, S. W.; Kim, M.; Ramsey, S.; Kurtzman, T.; Cho, A. E. Water Pharmacophore: Designing Ligands using Molecular Dynamics Simulations with Water. *Scientific Reports* **2018**, *8* (1), 10400.
(106) Wieder, M.; Garon, A.; Perricone, U.; Boresch, S.; Seidel, T.; Almerico, A. M.; Langer, T. Common Hits Approach: Combining Pharmacophore Modeling and Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2017**, *57* (2), 365-385.
(107) Polishchuk, P.; Kutlushina, A.; Bashirova, D.; Mokshyna, O.; Madzhidov, T. Virtual Screening Using Pharmacophore Models Retrieved from Molecular Dynamic Simulations. *International Journal of Molecular Sciences* **2019**, *20* (23), 5834.
(108) McGibbon, Robert T.; Beauchamp, Kyle A.; Harrigan, Matthew P.; Klein, C.; Swails, Jason M.; Hernández, Carlos X.; Schwantes, Christian R.; Wang, L.-P.; Lane, Thomas J.; Pande, Vijay S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **2015**, *109* (8), 1528-1532.
(109) Salentin, S.; Schreiber, S.; Haupt, V. J.; Adasme, M. F.; Schroeder, M. PLIP: fully automated protein–ligand interaction profiler. *Nucleic Acids Research* **2015**, *43* (W1), W443-W447.

(110) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55* (14), 6582-6594.
(111) Sayle, K. L.; Bentley, J.; Boyle, F. T.; Calvert, A. H.; Cheng, Y.; Curtin, N. J.; Endicott, J. A.; Golding, B. T.; Hardcastle, I. R.; Jewsbury, P.; et al. Structure-Based design of 2-Arylamino-4-cyclohexylmethyl-5-nitroso-6-aminopyrimidine inhibitors of cyclin-Dependent kinases 1 and 2. *Bioorganic & Medicinal Chemistry Letters* **2003**, *13* (18), 3079-3082.
(112) Pratt, D. J.; Bentley, J.; Jewsbury, P.; Boyle, F. T.; Endicott, J. A.; Noble, M. E. M. Dissecting the Determinants of Cyclin-Dependent Kinase 2 and Cyclin-Dependent Kinase 4 Inhibitor Selectivity. *J. Med. Chem.* **2006**, *49* (18), 5470-5477.
(113) Coxon, C. R.; Anscombe, E.; Harnor, S. J.; Martin, M. P.; Carbain, B.; Golding, B. T.; Hardcastle, I. R.; Harlow, L. K.; Korolchuk, S.; Matheson, C. J.; et al. Cyclin-Dependent Kinase (CDK) Inhibitors: Structure–Activity Relationships and Insights into the CDK-2 Selectivity of 6-Substituted 2-Arylaminopurines. *J. Med. Chem.* **2017**, *60* (5), 1746-1767.
(114) Choong, I. C.; Serafimova, I.; Fan, J.; Stockett, D.; Chan, E.; Cheeti, S.; Lu, Y.; Fahr, B.; Pham, P.; Arkin, M. R.; et al. A diaminocyclohexyl analog of SNS-032 with improved permeability and bioavailability properties. *Bioorganic & Medicinal Chemistry Letters* **2008**, *18* (21), 5763-5765.
(115) Fan, J.; Fahr, B.; Stockett, D.; Chan, E.; Cheeti, S.; Serafimova, I.; Lu, Y.; Pham, P.; Walker, D. H.; Hoch, U.; et al. Modifications of the isonipecotic acid fragment of SNS-032: Analogs with improved permeability and lower efflux ratio. *Bioorganic & Medicinal Chemistry Letters* **2008**, *18* (23), 6236-6239.
(116) Wang, S.; Griffiths, G.; Midgley, C. A.; Barnett, A. L.; Cooper, M.; Grabarek, J.; Ingram, L.; Jackson, W.; Kontopidis, G.; McClue, S. J.; et al. Discovery and Characterization of 2-Anilino-4- (Thiazol-5-yl)Pyrimidine Transcriptional CDK Inhibitors as Anticancer Agents. *Chemistry & Biology* **2010**, *17* (10), 1111-1121.
(117) Chu, X.-J.; DePinto, W.; Bartkovitz, D.; So, S.-S.; Vu, B. T.; Packman, K.; Lukacs, C.; Ding, Q.; Jiang, N.; Wang, K.; et al. Discovery of [4-Amino-2-(1-methanesulfonylpiperidin-4-ylamino)pyrimidin-5-yl](2,3-difluoro-6- methoxyphenyl)methanone (R547), A Potent and Selective Cyclin-Dependent Kinase Inhibitor with Significant in Vivo Antitumor Activity. *J. Med. Chem.* **2006**, *49* (22), 6549-6560.
(118) Fourches, D.; Muratov, E.; Tropsha, A. Trust, But Verify: On the Importance of Chemical Structure Curation in Cheminformatics and QSAR Modeling Research. *J. Chem. Inf. Model.* **2010**, *50* (7), 1189-1204.
(119) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959-5967.
(120) Bahia, M. S.; Kaspi, O.; Touitou, M.; Binayev, I.; Dhail, S.; Spiegel, J.; Khazanov, N.; Yosipof, A.; Senderowitz, H. A comparison between 2D and 3D descriptors in QSAR modeling based on bio-active conformations. *Mol. Inf.* **2023**, *42* (4), 2200186.
(121) Daré, J. K.; Freitas, M. P. Is conformation relevant for QSAR purposes? 2D Chemical representation in a 3D-QSAR perspective. *J. Comput. Chem.* **2022**, *43* (13), 917-922.
(122) Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. Construction of 3D-QSAR Models Using the 4D-QSAR Analysis Formalism. *J. Am. Chem. Soc.* **1997**, *119* (43), 10509-10524.
(123) Kyaw Zin, P. P.; Borrel, A.; Fourches, D. Benchmarking 2D/3D/MD-QSAR Models for Imatinib Derivatives: How Far Can We Predict? *J. Chem. Inf. Model.* **2020**, *60* (7), 3342-3360.
(124) Dietterich, T. G.; Lathrop, R. H.; Lozano-Pérez, T. Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence* **1997**, *89* (1), 31-71.
(125) Carbonneau, M.-A.; Cheplygina, V.; Granger, E.; Gagnon, G. Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognition* **2018**, *77*, 329-353.
(126) Herrera, F.; Ventura, S.; Bello, R.; Cornelis, C.; Zafra, A.; Sánchez-Tarragó, D.; Vluymans, S. *Multiple Instance Learning Foundations and Algorithms*; Springer, 2016.
(127) Nikonenko, A.; Zankov, D.; Baskin, I.; Madzhidov, T.; Polishchuk, P. Multiple Conformer Descriptors for QSAR Modeling. *Mol. Inf.* **2021**, *40* (11), 2060030.
(128) Zankov, D. V.; Matveieva, M.; Nikonenko, A. V.; Nugmanov, R. I.; Baskin, I. I.; Varnek, A.; Polishchuk, P.; Madzhidov, T. I. QSAR Modeling Based on Conformation Ensembles Using a Multi-Instance Learning Approach. *J. Chem. Inf. Model.* **2021**, *61* (10), 4913-4923.
(129) Zankov, D.; Polishchuk, P.; Madzhidov, T.; Varnek, A. Multi-Instance Learning Approach to Predictive Modeling of Catalysts Enantioselectivity. *Synlett* **2021**, *32* (18), 1833-1836.

(130) Zankov, D.; Madzhidov, T.; Polishchuk, P.; Sidorov, P.; Varnek, A. Multi-Instance Learning Approach to the Modeling of Enantioselectivity of Conformationally Flexible Organic Catalysts. *J. Chem. Inf. Model.* **2023**, *63* (21), 6629-6641.
(131) Amores, J. Multiple instance classification: Review, taxonomy and comparative study. *Artificial Intelligence* **2013**, *201*, 81-105.
(132) Wang, X.; Yan, Y.; Tang, P.; Bai, X.; Liu, W. Revisiting multiple instance neural networks. *Pattern Recognition* **2018**, *74*, 15-24.
(133) Ilse, M.; Tomczak, J.; Welling, M. Attention-based Deep Multiple Instance Learning. In Proceedings of the 35th International Conference on Machine Learning, Proceedings of Machine Learning Research; 2018.
(134) Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *J. Chem. Inf. Model.* **2015**, *55* (12), 2562-2574.
(135) Mayr, A.; Klambauer, G.; Unterthiner, T.; Steijaert, M.; Wegner, J. K.; Ceulemans, H.; Clevert, D.-A.; Hochreiter, S. Large-scale comparison of machine learning methods for drug target prediction on ChEMBL. *Chem. Sci.* **2018**, *9* (24), 5441-5451.
(136) Demšar, J. Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research* **2006**, *7* (1), 1-30.
(137) Trott, O.; Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2009**, *31* (2), 455-461.
(138) Yang, W.; Fidelis, T. T.; Sun, W.-H. Machine Learning in Catalysis, From Proposal to Practicing. *ACS Omega* **2020**, *5* (1), 83-88.
(139) Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning. *Science* **2019**, *363* (6424), eaau5631.
(140) Henle, J. J.; Zahrt, A. F.; Rose, B. T.; Darrow, W. T.; Wang, Y.; Denmark, S. E. Development of a Computer-Guided Workflow for Catalyst Optimization. Descriptor Validation, Subset Selection, and Training Set Analysis. *J. Am. Chem. Soc.* **2020**, *142* (26), 11578-11592.
(141) Zankov, D.; Madzhidov, T.; Varnek, A.; Polishchuk, P. Chemical complexity challenge: Is multi-instance machine learning a solution? *WIREs Computational Molecular Science* **2024**, *14* (1), e1698.
(142) Polishchuk, P. G.; Madzhidov, T. I.; Varnek, A. Estimation of the size of drug-like chemical space based on GDB-17 data. *J. Comput.-Aided Mol. Des.* **2013**, *27* (8), 675-679.
(143) Schneider, P.; Schneider, G. De Novo Design at the Edge of Chaos. *J. Med. Chem.* **2016**, *59* (9), 4077-4086.
(144) Schneider, G. Automating drug discovery. *Nature Reviews Drug Discovery* **2017**, *17*, 97.
(145) Böhm, H.-J. The computer program LUDI: A new method for the de novo design of enzyme inhibitors. *J. Comput.-Aided Mol. Des.* **1992**, *6* (1), 61-78.
(146) Wang, R.; Gao, Y.; Lai, L. LigBuilder: A Multi-Purpose Program for Structure-Based Drug Design. *Molecular modeling annual* **2000**, *6* (7), 498-516.
(147) Brown, N.; McKay, B.; Gasteiger, J. A novel workflow for the inverse QSPR problem using multiobjective optimization. *J. Comput.-Aided Mol. Des.* **2006**, *20* (5), 333-341.
(148) Hartenfeller, M.; Zettl, H.; Walter, M.; Rupp, M.; Reisen, F.; Proschak, E.; Weggen, S.; Stark, H.; Schneider, G. DOGS: Reaction-Driven de novo Design of Bioactive Compounds. *PLoS Comput. Biol.* **2012**, *8* (2), e1002380.
(149) Chéron, N.; Jasty, N.; Shakhnovich, E. I. OpenGrowth: An Automated and Rational Algorithm for Finding New Protein Ligands. *J. Med. Chem.* **2016**, *59* (9), 4171-4188.
(150) Hoksza, D.; Škoda, P.; Voršilák, M.; Svozil, D. Molpher: a software framework for systematic chemical space exploration. *J. Cheminf.* **2014**, *6* (1), 7.
(151) Szymkuć, S.; Gajewska, E. P.; Klucznik, T.; Molga, K.; Dittwald, P.; Startek, M.; Bajczyk, M.; Grzybowski, B. A. Computer-Assisted Synthetic Planning: The End of the Beginning. *Angew. Chem., Int. Ed.* **2016**, *55* (20), 5904-5937.
(152) Batiste, L.; Unzue, A.; Dolbois, A.; Hassler, F.; Wang, X.; Deerain, N.; Zhu, J.; Spiliotopoulos, D.; Nevado, C.; Caflisch, A. Chemical Space Expansion of Bromodomain Ligands Guided by in Silico Virtual Couplings (AutoCouple). *ACS Cent. Sci.* **2018**, *4* (2), 180-188.
(153) Merk, D.; Grisoni, F.; Friedrich, L.; Gelzinyte, E.; Schneider, G. Computer-Assisted Discovery of Retinoid X Receptor Modulating Natural Products and Isofunctional Mimetics. *J. Med. Chem.* **2018**, *61* (12), 5442-5447.

(154) Kutchukian, P. S.; Lou, D.; Shakhnovich, E. I. FOG: Fragment Optimized Growth Algorithm for the de Novo Generation of Molecules Occupying Druglike Chemical Space. *J. Chem. Inf. Model.* **2009**, *49* (7), 1630-1642.
(155) Olivecrona, M.; Blaschke, T.; Engkvist, O.; Chen, H. Molecular de-novo design through deep reinforcement learning. *J. Cheminf.* **2017**, *9* (1), 48.
(156) Popova, M.; Isayev, O.; Tropsha, A. Deep Reinforcement Learning for De-Novo Drug Design. **2017**.
(157) Yuan, W.; Jiang, D.; Nambiar, D. K.; Liew, L. P.; Hay, M. P.; Bloomstein, J.; Lu, P.; Turner, B.; Le, Q.-T.; Tibshirani, R.; et al. Chemical Space Mimicry for Drug Discovery. *J. Chem. Inf. Model.* **2017**, *57* (4), 875-882.
(158) Li, Y.; Zhang, L.; Liu, Z. Multi-Objective De Novo Drug Design with Conditional Graph Generative Model. **2018**.
(159) Polykovskiy, D.; Zhebrak, A.; Vetrov, D.; Ivanenkov, Y.; Aladinskiy, V.; Mamoshina, P.; Bozdaganyan, M.; Aliper, A.; Zhavoronkov, A.; Kadurin, A. Entangled Conditional Adversarial Autoencoder for de Novo Drug Discovery. *Molecular Pharmaceutics* **2018**, *15* (10), 4398-4405.
(160) Putin, E.; Asadulaev, A.; Ivanenkov, Y.; Aladinskiy, V.; Sanchez-Lengeling, B.; Aspuru-Guzik, A.; Zhavoronkov, A. Reinforced Adversarial Neural Computer for de Novo Molecular Design. *J. Chem. Inf. Model.* **2018**.
(161) Segler, M. H. S.; Kogej, T.; Tyrchan, C.; Waller, M. P. Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks. *ACS Cent. Sci.* **2018**, *4* (1), 120-131.
(162) Mouchlis, V. D.; Afantitis, A.; Serra, A.; Fratello, M.; Papadiamantis, A. G.; Aidinis, V.; Lynch, I.; Greco, D.; Melagraki, G. Advances in De Novo Drug Design: From Conventional to Machine Learning Methods. *International Journal of Molecular Sciences* **2021**, *22* (4), 1676.
(163) Tong, X.; Liu, X.; Tan, X.; Li, X.; Jiang, J.; Xiong, Z.; Xu, T.; Jiang, H.; Qiao, N.; Zheng, M. Generative Models for De Novo Drug Design. *J. Med. Chem.* **2021**.
(164) Bilodeau, C.; Jin, W.; Jaakkola, T.; Barzilay, R.; Jensen, K. F. Generative models for molecular discovery: Recent advances and challenges. *WIREs Computational Molecular Science* **2022**, *n/a* (n/a), e1608.
(165) Gao, W.; Coley, C. W. The Synthesizability of Molecules Proposed by Generative Models. *J. Chem. Inf. Model.* **2020**, *60* (12), 5714-5723.
(166) Polishchuk, P. CReM: chemically reasonable mutations framework for structure generation. *J. Cheminf.* **2020**, *12* (1), 28.
(167) Dalke, A.; Hert, J.; Kramer, C. mmpdb: An Open-Source Matched Molecular Pair Platform for Large Multiproperty Data Sets. *J. Chem. Inf. Model.* **2018**, *58* (5), 902-910.
(168) Hussain, J.; Rea, C. Computationally Efficient Algorithm to Identify Matched Molecular Pairs (MMPs) in Large Data Sets. *J. Chem. Inf. Model.* **2010**, *50* (3), 339-348.
(169) Ertl, P.; Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminf.* **2009**, *1* (1), 8.
(170) Brown, N.; Fiscato, M.; Segler, M. H. S.; Vaucher, A. C. GuacaMol: Benchmarking Models for de Novo Molecular Design. *J. Chem. Inf. Model.* **2019**.
(171) Genheden, S.; Thakkar, A.; Chadimová, V.; Reymond, J.-L.; Engkvist, O.; Bjerrum, E. AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning. *J. Cheminf.* **2020**, *12* (1), 70.
(172) Segler, M. H. S.; Preuss, M.; Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **2018**, *555*, 604.
(173) Jensen, J. H. A graph-based genetic algorithm and generative model/Monte Carlo tree search for the exploration of chemical space. *Chem. Sci.* **2019**, *10* (12), 3567-3572.
(174) Yoshikawa, N.; Terayama, K.; Sumita, M.; Homma, T.; Oono, K.; Tsuda, K. Population-based De Novo Molecule Generation, Using Grammatical Evolution. *Chem. Lett.* **2018**, *47* (11), 1431-1434.
(175) Reeves, S.; DiFrancesco, B.; Shahani, V.; MacKinnon, S.; Windemuth, A.; Brereton, A. E. Assessing methods and obstacles in chemical space exploration. *Applied AI Letters* **2020**, *1* (2), e17.
(176) Hadfield, T. E.; Imrie, F.; Merritt, A.; Birchall, K.; Deane, C. M. Incorporating Target-Specific Pharmacophoric Information into Deep Generative Models for Fragment Elaboration. *J. Chem. Inf. Model.* **2022**.
(177) Ishitani, R.; Kataoka, T.; Rikimaru, K. Molecular Design Method Using a Reversible Tree Representation of Chemical Compounds and Deep Reinforcement Learning. *J. Chem. Inf. Model.* **2022**.
(178) Wu, Y.; Cashman, M.; Choma, N.; Prates, É. T.; Vergara, V. G. M.; Shah, M.; Chen, A.; Clyde, A.; Brettin, T. S.; Jong, W. A. d.; et al. Spatial Graph Attention and Curiosity-driven Policy for Antiviral Drug Discovery. *arxiv* **2022**.

(179) Krysko, A. A.; Samoylenko, G. V.; Polishchuk, P. G.; Andronati, S. A.; Kabanova, T. A.; Khristova, T. M.; Kuz'min, V. E.; Kabanov, V. M.; Krysko, O. L.; Varnek, A. A.; et al. RGD mimetics containing phthalimidine fragment, novel ligands of fibrinogen receptor. *Bioorganic & Medicinal Chemistry Letters* **2011**, *21*, 5971–5974.
(180) Krysko, A. A.; Samoylenko, G. V.; Polishchuk, P. G.; Fonari, M. S.; Kravtsov, V. C.; Andronati, S. A.; Kabanova, T. A.; Lipkowski, J.; Khristova, T. M.; Kuz'min, V. E.; et al. Synthesis, biological evaluation, X-ray molecular structure and molecular docking studies of RGD mimetics containing 6-amino-2,3-dihydroisoindolin-1-one fragment as ligands of integrin αIIbβ3. *Bioorganic & Medicinal Chemistry* **2013**, *21* (15), 4646-4661.
(181) Polishchuk, P. G.; Samoylenko, G. V.; Khristova, T. M.; Krysko, O. L.; Kabanova, T. A.; Kabanov, V. M.; Kornylov, A. Y.; Klimchuk, O.; Langer, T.; Andronati, S. A.; et al. Design, Virtual Screening, and Synthesis of Antagonists of αIIbβ3 as Antiplatelet Agents. *J. Med. Chem.* **2015**, *58* (19), 7681-7694.
(182) Krysko, A. A.; Kornylov, A. Y.; Polishchuk, P. G.; Samoylenko, G. V.; Krysko, O. L.; Kabanova, T. A.; Kravtsov, V. C.; Kabanov, V. M.; Wicher, B.; Andronati, S. A. Synthesis, biological evaluation and molecular docking studies of 2-piperazin-1-yl-quinazolines as platelet aggregation inhibitors and ligands of integrin αIIbβ3. *Bioorganic & Medicinal Chemistry Letters* **2016**, *26* (7), 1839-1843.
(183) Klimenko, K.; Lyakhov, S.; Shibinskaya, M.; Karpenko, A.; Marcou, G.; Horvath, D.; Zenkova, M.; Goncharova, E.; Amirkhanov, R.; Krysko, A.; et al. Virtual screening, synthesis and biological evaluation of DNA intercalating antiviral agents. *Bioorganic & Medicinal Chemistry Letters* **2017**, *27* (16), 3915-3919.
(184) Schadich, E.; Kryshchyshyn-Dylevych, A.; Holota, S.; Polishchuk, P.; Džubak, P.; Gurska, S.; Hajduch, M.; Lesyk, R. Assessing different thiazolidine and thiazole based compounds as antileishmanial scaffolds. *Bioorganic & Medicinal Chemistry Letters* **2020**, *30* (23), 127616.
(185) Schadich, E.; Nylen, S.; Gurska, S.; Kotulova, J.; Andronati, S.; Pavlovsky, V.; Soboleva, S.; Polishchuk, P.; Hajduch, M.; Dzubak, P. Activity of 1-aryl-4-(naphthalimidoalkyl) piperazine derivatives against Leishmania major and Leishmania mexicana. *Parasitology International* **2022**, 102647.
(186) Rehulka, J.; Subtelna, I.; Kryshchyshyn-Dylevych, A.; Cherniienko, A.; Ivanova, A.; Matveieva, M.; Polishchuk, P.; Gurska, S.; Hajduch, M.; Zagrijtschuk, O.; et al. Anticancer 5-arylidene-2-(4-hydroxyphenyl)aminothiazol-4(5H)-ones as tubulin inhibitors. *Archiv der Pharmazie* **2022**, e2200419.
(187) Jurášek, M.; Řehulka, J.; Hrubá, L.; Ivanová, A.; Gurská, S.; Mokshyna, O.; Trousil, P.; Huml, L.; Polishchuk, P.; Hajdúch, M.; et al. Triazole-based estradiol dimers prepared via CuAAC from 17α-ethinyl estradiol with five-atom linkers causing G2/M arrest and tubulin inhibition. *Bioorganic Chemistry* **2023**, *131*, 106334.
(188) Ackloo, S.; Al-awar, R.; Amaro, R. E.; Arrowsmith, C. H.; Azevedo, H.; Batey, R. A.; Bengio, Y.; Betz, U. A. K.; Bologa, C. G.; Chodera, J. D.; et al. CACHE (Critical Assessment of Computational Hit-finding Experiments): A public–private partnership benchmarking initiative to enable the development of computational methods for hit-finding. *Nature Reviews Chemistry* **2022**, *6*, 287-295.
(189) Libby, P.; Theroux, P. Pathophysiology of Coronary Artery Disease. *Circulation* **2005**, *111* (25), 3481-3488.
(190) Davì, G.; Patrono, C. Platelet Activation and Atherothrombosis. *New England Journal of Medicine* **2007**, *357* (24), 2482-2494.
(191) Ruf, A.; Patscheke, H. Flow Cytometric Detection of Activated Platelets: Comparison of Determining Shape Change, Fibrinogen Binding, and P-Selectin Expression. *Semin Thromb Hemost* **1995**, *21* (02), 146-151.
(192) Massberg , S.; Brand , K.; Grüner , S.; Page , S.; Müller , E.; Müller , I.; Bergmeier , W.; Richter , T.; Lorenz , M.; Konrad , I.; et al. A Critical Role of Platelet Adhesion in the Initiation of Atherosclerotic Lesion Formation. *Journal of Experimental Medicine* **2002**, *196* (7), 887-896.
(193) Furie, B.; Furie, B. C. Mechanisms of Thrombus Formation. *New England Journal of Medicine* **2008**, *359* (9), 938-949.
(194) Offermanns, S. Activation of Platelet Function Through G Protein–Coupled Receptors. *Circulation Research* **2006**, *99* (12), 1293-1304.
(195) Wang, X.; Dorsam, R. T.; Lauver, A.; Wang, H.; Barbera, F. A.; Gibbs, S.; Varon, D.; Savion, N.; Friedman, S. M.; Feuerstein, G. Z. Comparative Analysis of Various Platelet Glycoprotein IIb/IIIa Antagonists on Shear-Induced Platelet Activation and Adhesion. *Journal of Pharmacology and Experimental Therapeutics* **2002**, *303* (3), 1114-1120.
(196) Clappers, N.; Brouwer, M. A.; Verheugt, F. W. A. Antiplatelet treatment for coronary heart disease. *Heart* **2007**, *93* (2), 258-265.

(197) Freson, K.; Thys, C.; Wittevrongel, C.; Geet, V. C. Mechanisms of Action and Targets for Actual and Future Antiplatelet Drugs. *Mini-Reviews in Medicinal Chemistry* **2006**, *6* (6), 719-726.
(198) Andronati, S. A.; Karaseva, T. L.; Krysko, A. A. Peptidomimetics - Antagonists of the Fibrinogen Receptors: Molecular Design, Structures, Properties and Therapeutic Applications. *Current Medicinal Chemistry* **2004**, *11* (9), 1183-1211.
(199) Hartman, G. D.; Egbertson, M. S.; Halczenko, W.; Laswell, W. L.; Duggan, M. E.; Smith, R. L.; Naylor, A. M.; Manno, P. D.; Lynch, R. J. Non-peptide fibrinogen receptor antagonists. 1. Discovery and design of exosite inhibitors. *J. Med. Chem.* **1992**, *35* (24), 4640-4642.
(200) Scarborough, R. M.; Naughton, M. A.; Teng, W.; Rose, J. W.; Phillips, D. R.; Nannizzi, L.; Arfsten, A.; Campbell, A. M.; Charo, I. F. Design of potent and specific integrin antagonists. Peptide antagonists with high specificity for glycoprotein IIb-IIIa. *Journal of Biological Chemistry* **1993**, *268* (2), 1066-1073.
(201) Coller, B. S. A new murine monoclonal antibody reports an activation-dependent change in the conformation and/or microenvironment of the platelet glycoprotein IIb/IIIa complex. *The Journal of Clinical Investigation* **1985**, *76* (1), 101-108.
(202) Berkowitz, S. D.; Sane, D. C.; Sigmon, K. N.; Shavender, J. H.; Harrington, R. A.; Tcheng, J. E.; Topol, E. J.; Califf, R. M. Occurrence and clinical significance of thrombocytopenia in a population undergoing high-risk percutaneous coronary revascularization. *Journal of the American College of Cardiology* **1998**, *32* (2), 311-319.
(203) McClure, M. W.; Berkowitz, S. D.; Sparapani, R.; Tuttle, R.; Kleiman, N. S.; Berdan, L. G.; Lincoff, A. M.; Deckers, J.; Diaz, R.; Karsch, K. R.; et al. Clinical Significance of Thrombocytopenia During a Non–ST-Elevation Acute Coronary Syndrome. *Circulation* **1999**, *99* (22), 2892-2900.
(204) Bonachéra, F.; Parent, B.; Barbosa, F.; Froloff, N.; Horvath, D. Fuzzy Tricentric Pharmacophore Fingerprints. 1. Topological Fuzzy Pharmacophore Triplets and Adapted Molecular Similarity Scoring Schemes. *J. Chem. Inf. Model.* **2006**, *46* (6), 2457-2477.
(205) Rognan, D. BioinfoDB: un inventaire de molécules commercialement disponibles à des fins de criblage biologique. *La Gazette du CINES* **2005**, 1-4.
(206) Bougie, D. W.; Wilker, P. R.; Wuitschick, E. D.; Curtis, B. R.; Malik, M.; Levine, S.; Lind, R. N.; Pereira, J.; Aster, R. H. Acute thrombocytopenia after treatment with tirofiban or eptifibatide is associated with antibodies specific for ligand-occupied GPIIb/IIIa. *Blood* **2002**, *100* (6), 2071-2076.
(207) Bougie, D. W.; Rasmussen, M.; Zhu, J.; Aster, R. H. Antibodies causing thrombocytopenia in patients treated with RGD-mimetic platelet inhibitors recognize ligand-specific conformers of αIIb/β3 integrin. *Blood* **2012**, *119* (26), 6317-6325.
(208) Blue, R.; Murcia, M.; Karan, C.; Jiroušková, M.; Coller, B. S. Application of high-throughput screening to identify a novel αIIb-specific small- molecule inhibitor of αIIbβ3-mediated platelet interaction with fibrinogen. *Blood* **2008**, *111* (3), 1248-1256.
(209) Zhu, J.; Zhu, J.; Negri, A.; Provasi, D.; Filizola, M.; Coller, B. S.; Springer, T. A. Closed headpiece of integrin αIIbβ3 and its complex with an αIIbβ3-specific antagonist that does not induce opening. *Blood* **2010**, *116* (23), 5050-5059.
(210) Zhu, J.; Choi, W.-S.; McCoy, J. G.; Negri, A.; Zhu, J.; Naini, S.; Li, J.; Shen, M.; Huang, W.; Bougie, D.; et al. Structure-Guided Design of a High-Affinity Platelet Integrin α_{IIb}β₃ Receptor Antagonist That Disrupts Mg²⁺ Binding to the MIDAS. *Science Translational Medicine* **2012**, *4* (125), 125ra132-125ra132.
(211) Jiang, J.-k.; McCoy, J. G.; Shen, M.; LeClair, C. A.; Huang, W.; Negri, A.; Li, J.; Blue, R.; Harrington, A. W.; Naini, S.; et al. A novel class of ion displacement ligands as antagonists of the αIIbβ3 receptor that limit conformational reorganization of the receptor. *Bioorganic & Medicinal Chemistry Letters* **2014**, *24* (4), 1148-1153.
(212) Aoyama, T.; Ino, Y.; Ozeki, M.; Oda, M.; Sato, T.; Koshiyama, Y.; Suzuki, S.; Fujita, M. Pharmacological Studies of FUT-175, Nafamstat Mesilate I. Inhibition of Protease Activity in in Vitro and in Vivo Experiments. *Japanese Journal of Pharmacology* **1984**, *35* (3), 203-227.
(213) Negri, A.; Li, J.; Naini, S.; Coller, B. S.; Filizola, M. Structure-based virtual screening of small-molecule antagonists of platelet integrin αIIbβ3 that do not prime the receptor to bind ligand. *J. Comput.-Aided Mol. Des.* **2012**, *26* (9), 1005-1015.
(214) Pâques, E.-P.; Römisch, J. Comparative study on the in vitro effectiveness of antithrombotic agents. *Thrombosis Research* **1991**, *64* (1), 11-21.

(215) Tsuda, Y.; Nakahara, T.; Ueda, K.; Mori, A.; Sakamoto, K.; Ishii, K. Effect of Nafamostat on *N*-Methyl-D-aspartate-Induced Retinal Neuronal and Capillary Degeneration in Rats. *Biological and Pharmaceutical Bulletin* **2012**, *35* (12), 2209-2213.
(216) Inman, R. D.; Chiu, B. Nafamostat mesylate, a serine protease inhibitor, demonstrates novel antimicrobial properties and effectiveness in Chlamydia-induced arthritis. *Arthritis Research & Therapy* **2012**, *14* (3), R150.
(217) Ohtake, Y.; Hirasawa, H.; Sugai, T.; Oda, S.; Shiga, H.; Matsuda, K.; Kitamura, N. Nafamostat Mesylate as Anticoagulant in Continuous Hemofiltration and Continuous Hemodiafiltration. In *Continuous Hemofiltration: 2nd International Conferenc, Baden/Austria, September 1990*, Sieberth, H. G., Mann, H., Stummvoll, H. K. Eds.; Vol. 93; S.Karger AG, 1991; p 0.
(218) Nakae, H.; Tajimi, K. Pharmacokinetics of Nafamostat Mesilate During Continuous Hemodiafiltration with a Polyacrylonitrile Membrane. *Therapeutic Apheresis and Dialysis* **2003**, *7* (5), 483-485.
(219) Liverton, N. J.; Armstrong, D. J.; Claremon, D. A.; Remy, D. C.; Baldvin, J. J.; Lynch, R. J.; Zhang, G.; Gould, R. J. Nonpeptide glycoprotein IIb/IIIa inhibitors: Substituted quinazolinediones and quinazolinones as potent fibrinogen receptor antagonists. *Bioorganic & Medicinal Chemistry Letters* **1998**, *8* (5), 483-486.
(220) Ashford, R. W. The leishmaniases as emerging and reemerging zoonoses. *International Journal for Parasitology* **2000**, *30* (12), 1269-1281.
(221) Desjeux, P. Leishmaniasis: current situation and new perspectives. *Comparative Immunology, Microbiology and Infectious Diseases* **2004**, *27* (5), 305-318.
(222) Palatnik-de-Sousa, C. B.; Day, M. J. One Health: The global challenge of epidemic and endemic leishmaniasis. *Parasites & Vectors* **2011**, *4* (1), 197.
(223) Burchmore, R. J. S.; Barrett, M. P. Life in vacuoles – nutrient acquisition by Leishmania amastigotes. *International Journal for Parasitology* **2001**, *31* (12), 1311-1320.
(224) Pigott, D. M.; Bhatt, S.; Golding, N.; Duda, K. A.; Battle, K. E.; Brady, O. J.; Messina, J. P.; Balard, Y.; Bastien, P.; Pratlong, F.; et al. Global distribution maps of the leishmaniases. *eLife* **2014**, *3*, e02851.
(225) Berman, J. Clinical status of agents being developed for leishmaniasis. *Expert Opinion on Investigational Drugs* **2005**, *14* (11), 1337-1346.
(226) Moore, E. M.; Lockwood, D. N. Treatment of Visceral Leishmaniasis. *Journal of Global Infectious Diseases* **2010**, *2* (2), 151-158.
(227) Croft, S. L.; Olliaro, P. Leishmaniasis chemotherapy—challenges and opportunities. *Clinical Microbiology and Infection* **2011**, *17* (10), 1478-1483.
(228) Reguera, R. M.; Calvo-Álvarez, E.; Álvarez-Velilla, R.; Balaña-Fouce, R. Target-based vs. phenotypic screenings in Leishmania drug discovery: A marriage of convenience or a dialogue of the deaf? *International Journal for Parasitology: Drugs and Drug Resistance* **2014**, *4* (3), 355-357.
(229) Andronati, S. A.; Soboleva, S. G.; Makan, S. Y.; Sava, V. M.; Kolodeev, G. E.; Voronina, T. A.; Molodavkin, G. M. Anxiolytic Properties of 1-Aryl-4-(phthalimidoalkyl)piperazines and Their Affinity to 5-HT1A Serotonin and D2 Dopamine Receptors. *Pharmaceutical Chemistry Journal* **2003**, *37* (1), 15-18.
(230) Andronati, S. A.; Soboleva, S. G.; Golturenko, A. V.; Kostenko, E. A.; Karaseva, T. L. Synthesis and Pharmacological Properties of 1-Aryl-4-[4-(naphthalimidoalkyl)]piperazines. *Pharmaceutical Chemistry Journal* **2002**, *36* (11), 588-590.