CHEMMEDCHEM

CHEMISTRY ENABLING DRUG DISCOVERY

Accepted Article

Title: Consensus Predictive Model for the prediction of Human K562 Cell Growth Inhibition through Enalos Cloud Platform

Authors: Antreas Afantitis, Georgios Leonis, Roberto Gambari, and Georgia Melagraki

To be cited as: *ChemMedChem* 10.1002/cmdc.201700675

Link to VoR: http://dx.doi.org/10.1002/cmdc.201700675

A Journal of ChemPubSoc Europe

WILEY-VCH www.chemmedchem.org

WILEY-VCH

Consensus Predictive Model for Human K562 Cell Growth Inhibition through Enalos Cloud Platform

Antreas Afantitis,*[a] Georgios Leonis,[a] Roberto Gambari,[b] and Georgia Melagraki*[c]

[a] Dr. A. Afantitis, Dr. G. Leonis
NovaMechanics Ltd
Nicosia, Cyprus
E-mail: afantitis@novamechanics.com

[b] Prof. R. Gambari
Department of Life Sciences and Biotechnology
University of Ferrara
Via Fossato di Mortara n.74, 44121 Ferrara, Italy

[c] Dr. G. Melagraki
Department of Military Sciences, Division of Physical Sciences and Applications
Hellenic Army Academy
Vari, Greece
E-mail: georgiamelagraki@gmail.com

Supporting information for this article is given via a link at the end of the document.

Abstract: Beta thalassemia is an inherited hematologic disorder caused by various mutations of the β-globin gene, thus resulting in a significant decrease of adult hemoglobin (HbA) production. The increase of fetal hemoglobin (HbF) levels by drug molecules is considered of great potential in β-thalassemia treatment, being expected to counterbalance the impaired production of HbA. In this work, based on a set of 129 experimentally tested biological inhibitors, we have developed and validated a computational model for the prediction of K562 functional inhibition, possibly associated with HbF induction. To facilitate future advancements in the field, we have incorporated our model into Enalos Cloud Platform that enabled online access to our computational scheme (http://enalos.insilicotox.com/K562) through a user-friendly interface. This web service is offered to the wider community to promote the *in silico* drug discovery through fast and reliable predictions.

Introduction

The occurrence of several mutations of the β-globin gene drastically reduces the production of adult β-globin, causing at the same time an unbalanced accumulation of free α-globin chains in animal cells.[1-2] This results in diminished production of adult hemoglobin (HbA)[3] and causes the development of important hereditary hematologic diseases, which are known as β-thalassemias.[4-7] It is estimated that 1 out of 100,000 individuals is affected worldwide every year,[3] with populations in the Mediterranean region to be most vulnerable to the development of β-thalassemia. Currently, there is no definitive cure of β-thalassemia and clinical practices usually involve continuing blood transfusions along with chelation therapy,[8] and, less frequently, bone marrow transplantation.[9] Alternative therapeutic options are DNA-based treatments of β-thalassemia,[5, 10-12] however, improved therapeutic approaches are eagerly needed since present treatments are accompanied by unwanted side effects and severe limitations.[5]

It has been suggested that efficient treatment of β-thalassemia can be obtained by fetal hemoglobin (HbF) induction.[5, 11, 13-14] This was based on the HbF-inducing ability of hydroxyurea, which significantly improves the clinical parameters of β-thalassemia patients.[11] However, due to important adverse effects associated with hydroxyurea treatment, research efforts have been focused on the discovery of other HbF inducers for the development of new medications against β-thalassemia.[15-18]

Over the years, several proteins have been identified to either directly or indirectly repress the transcription of the γ-globin gene.[19-26] These transcription-repressing proteins can be selectively inhibited by pharmaceutical molecules, which may lead to HbF production through *de novo* activation of γ-globin gene transcription.[27-28] Genetic experiments identified the zinc finger transcription factor B-cell lymphoma/leukemia 11A (BCL11A) as the most important inhibitor of HbF expression, and resulted in the discovery of an HbF-related position on chromosome 2 (found in BCL11A gene).[28] The prevalent isoform of BCL11A in adult erythroid progenitor cells is BCL11A-XL.[21-22] It has been shown that HbF induction occurs from transgenic deactivation of BCL11A in mouse and human cells.[27, 29-31] To progress from the general mechanism of HbF induction to potential therapeutic approaches, controlled and stable HbF induction through BCL11A inhibition based on shRNAs has been achieved with clinically relevant efficiency.[32] Such an approach is of major significance, because it may recognize potential pharmacologic inducers of γ-globin gene repressors. Since low levels of BCL11A-XL are expressed by human erythroleukemia K562 cells, the significance of K562 as an intermediate route to identify pharmaceutical inducers of HbF becomes apparent.

In an early publication, Samid et al. studied the effect of phenylacetate on human K562 cells and they observed that time/dose-dependent erythroid differentiation occurred, followed by HbF accumulation.[33] In 2003, Witt et al. identified a series of histone deacetylase compounds as HbF inducers through their inhibitory actions (including anti-proliferative effects) against human K562 cells.[34] Macari and colleagues discovered three activators of the NRF2-antioxidant response element signaling pathway that induced γ-globin gene expression and HbF production in K562 cells.[35] Also, He et al. identified a methyl transferase inhibitor (Adox) as HbF inducer in K562 cultures.[36] More recently, Ng et al. reviewed the status of current agents that stimulate the production of HbF and concluded that human K562 cells are extensively used as screening platforms for HbF-inducing compounds.[37-39]

The *in vitro* cytotoxicity of new (*E*)-*α*-benzylthio chalcones against K562 cells was evaluated by Reddy et al.[40] Several of the organic molecules profoundly inhibited proliferation of K562 cells and exhibited satisfactory toxicity profiles. K562 cells are also commonly used for the screening of various inhibitors,[41-43] including Bcr-Abl kinase inhibitor-candidates,[44] and antitumor compounds that induce cell differentiation.[45-46] Differentiation treatment is an advanced strategy for neoplasia therapies, based on the identification and use of pharmaceutical agents that act as differentiation promoters.[47]

Augmented expression of embryo-fetal γ-, ε-, and ζ-globin genes is directly related to K562 erythroid differentiation.[48] This feature makes K562 cells a useful model system for the study of compounds that are effective γ-globin inducers, and possibly act against sickle cell disease (SCD) and β–thalassemia.[49] Enhanced γ–globin gene production restricts the polymerization of sickle Hb (HbS) in SCD, and partially substitutes the biological activity of the non-functional β–globin gene in β–thalassemia.[50] It is known that several HbF inducers studied using the K562 cell system inhibit cells proliferation. In addition, it has been indicated that compounds, which act as erythroid-differentiation promoters in K562 cells may activate the expression of γ-globin in primary erythroid cells taken from β–thalassemia or SCD patients.[5]

Several cheminformatics approaches to medicinal chemistry applications have been proposed in the literature.[51] Currently, few molecular modeling studies based on the use of QSAR techniques for the identification of K562 inhibitors have been reported. In a recent article, Vrontaki et al. developed pharmacophore and 3D-QSAR models based on a set of previously synthesized non-ATP-targeting Bcr-Abl inhibitors[40] to design a new strategy against chronic myelogenous leukemia (CML).[52] Bcr-Abl is a kinase, which is directly implicated in cancer development. The authors constructed a robust QSAR model that reliably predicted the cytotoxic effects of the above molecules on K562 cells. Vrontaki and colleagues stressed the appropriateness of the predictive model to assist successful drug design against β-thalassemia based on the fact that Bcr-Abl inhibitors are known to induce erythroid differentiation and γ-globin expression in CML cell lines and primary erythroid cells.

In 2017, Rivera et al. used biological binding assays to assess the activity of 16 quinoxaline analogs against K562 cells.[53] Quinoxalines were selected because they are known to have antitumor action. The authors employed flow cytometry to elucidate the mechanism by which cell death was induced and they developed a QSAR model to evaluate molecular descriptors for the quinoxaline analogs. Therefore, they were able to classify the compounds based on their IC_{50} values and also successfully explained the cell death mechanism as dictated by the action of the most active compound.

In another work, Monga et al. developed predictive 3D-QSAR models based on validated 3D-pharmacophore hypotheses via a selected list of K562 inhibitors.[54] Finally, the cytotoxicity of ent-kauranoids against K562 cells was also modeled through 3D-QSAR CoMFA approaches by Yi et al.[55]

Conclusively, the compensation of the clinically relevant unbalance of α-globin genes production to non-α-globin genes in normal hemoglobin could be obtained by boosting the expression of γ-globin genes.[56] Cytotoxic drugs, growth factors, and erythroid differentiation inducing compounds might increase the generation of fetal hemoglobin in humans.[57] The great diversity of pharmacological inducing compounds (such as, hydroxyurea, 5-azacytidine, erythropoietin, cytosine arabinoside, and sodium butyrate) enables their classification into various groups depending on their chemical structure and putative mechanism(s) of action.[50, 57-59] Hydroxyurea is the most commonly used compound for the production of γ-globin genes, and is considered a first-choice drug for sickle cell anemia. Despite this, it has been demonstrated that long term use of hydroxyurea may cause side effects and also presents significantly diminished beneficial action.[60] Moreover, increased HbF production is obtained through short-chain fatty acids consumption; similar results were observed with 5-azacytidine and thalidomide.[57, 61-63] Examples of erythropoietic growth factors employed for further activation of erythropoiesis include erythropoietin and darvopoietin.[59, 64-65]

Despite the widely-known pharmacological action of HbF inducers against β-thalassemia, current medications present low efficacy and are associated with several unwanted side effects, such as high cytotoxicity, which eventually might lead to cancer development. Therefore, ongoing research efforts necessitate the discovery of novel pharmaceutical molecules that could act as efficient HbF inducers without causing toxic effects.

Since it has been shown that K562 cell regulators may induce therapeutic effects on β–thalassemia patients, we have been motivated to computationally explore a pool of 129 available compounds that have been experimentally evaluated as K562 biological inhibitors. Based on this large dataset, we have combined several cheminformatics techniques to afford the structural features that modulate the cytotoxicity of the 129 compounds in the K562 cell line and to develop and deliver a robust and validated model. On top of that, we have made the model available online through a user-friendly interface for fast and accurate virtual screening of newly proposed structures. Overall, the objective of this work was to develop a cheminformatics strategy for the identification of potential HbF inducers through K562 inhibition, and subsequently to facilitate the use of our developed tool by releasing the model online via the Enalos Cloud Platform.[66-67]

Results and Discussion

A cheminformatics workflow was first designed and then implemented within KNIME for the development of an accurate and robust model for the prediction of human K562 cell growth inhibition. The main steps included in this workflow are briefly described below: data curation and preprocessing, descriptors calculation, variable selection, model development and validation, and domain of applicability determination.

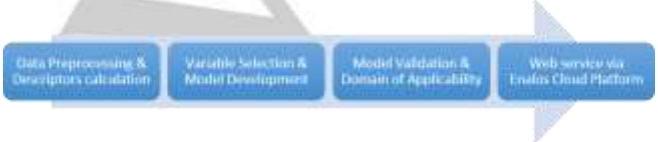

Data Preprocessing & Descriptors calculation
Variable Selection & Model Development
Model Validation & Domain of Applicability
Web service via Enalos Cloud Platform

Figure 1. Cheminformatics workflow for the development of the K562 cell growth inhibition predictive model.

The proposed workflow was built in KNIME by combining existing nodes with NovaMechanics proprietary Enalos KNIME nodes[68] that implement several significant tasks, essential for model development. Based on the procedure described, a validated model was developed and was subsequently incorporated within Enalos Cloud Platform to build a new web service dedicated to the prediction of K562 cell growth inhibition with a user-friendly interface [http://enalos.insilicotox.com/K562/].

The PubChem Enalos+ KNIME node was used to search and retrieve data available in PubChem.[69] This step afforded a set of 129 diverse small molecules, which have been tested as potential K562 inhibitors based on a functional K562 assay that was performed *in vitro* (in biochemical assays) and in cell-based assays using human erythroleukemia cells for each compound. More details on the bioassay used and data collected can be retrieved from PubChem under the AID742260 record (https://pubchem.ncbi.nlm.nih.gov/bioassay/742260). This dataset was used as the starting point to initiate our model development since it satisfies the following requirements: sufficient number of compounds, balance between the active and inactive class, wide range of structural features, and experimental evaluation using the same protocol across all molecules (inhibition of human K-562 cell growth in a cell viability assay). The above are crucial factors for the development of a robust, reliable and accurate predictive model.

Among the tested compounds, those that have exhibited an activity of ≤ 50 μM were reported as active while all remaining compounds were reported as inactive. In total, among available compounds, 67 and 62 compounds were classified as active and inactive, respectively. The compounds were divided into a training and a test set in a ratio of 80:20 to be used for model development and validation, respectively. Among the 129 compounds, 104 were included in the training set and 25 in the test set using the Random Partitioning node in KNIME analytics platform.[70]

The Enalos Mold2 KNIME node[71] was subsequently used for the calculation of a wide range of molecular descriptors as proposed within Mold2 software.[72] For each compound included in the dataset, 777 descriptors were calculated accounting for structural, geometric, and topological features of the compounds. After an initial screening, many descriptors were removed because of their low discrimination power and for this purpose, the Low Variance Filter node was employed.[70]

Throughout the model development process, additional descriptors were eliminated by applying different variable selection techniques to identify the most relevant descriptors for predicting K562 inhibition. Our cheminformatics workflow allowed a fast experimentation with a wide range of variable selection techniques combined with different modeling methodologies. Thus, we have experimented with a great variety of variable selection and modeling techniques to select those that would better describe the relationship between our descriptors and inhibition activity. Among different combinations, we have resulted in three models that were validated as robust and accurate.

These models (**I–III)** included a combination of two different feature selection techniques, namely the Gain Attribute evaluator and the InfoGain Attribute Ratio Feature evaluator, with three modeling methodologies, namely Random Tree, Random Forest and kNN. In particular, Model **I** was developed based on Information Gain Ranking Filter and Random Tree, Model **II** was developed based on Information Gain Ranking Filter and kNN and Model **III** was developed based on Gain Attribute evaluator and Random Forest. The selected descriptors for each different model and their definition within Mold2 are summarized in Table 1.

Table 1. Selected molecular descriptors for each of the three validated predictive models.

Model I: Information Gain Ranking Filter with Random Tree	
Variable	**Description**
D188	Balaban mass weighted index
D189	Balaban van der Waals weighted index
D193	Balaban-type polarizability weighted index
D187	Balaban heteroatoms bonds weighted index
D192	Balaban electronegativity weighted with Allred-Rochow-Scale index
D190	Balaban electronegativity weighted with Pauling-Scale index
Model II: Information Gain Ranking Filter with kNN	
Variable	**Description**
D188	Balaban mass weighted index
D189	Balaban van der Waals weighted index
D193	Balaban-type polarizability weighted index
Model III: Gain Ratio with Random Forest	
Variable	**Description**
D392	sum of topological distance between the vertices F and F
D658	number of group nitriles (aromatic)
D714	number of group CH_3R and CH_4
D596	number of total primary C-sp3
D599	number of total quaternary C-sp3
D189	Balaban van der Waals weighted index
D188	Balaban mass weighted index
D193	Balaban-type polarizability weighted index

More information on the Balaban and Balaban-related molecular descriptors that were predominantly selected, including extended or modified formulas, can be found in two comprehensive books by Todeschini and Consonni.[73-74]

In addition to these three models, a consensus model based on the majority vote approach was also compared with each of the individual models. This approach considers the prediction output from each of the three validated models and makes a final assessment based on the class assigned by the majority of the individual models (Figure 2).

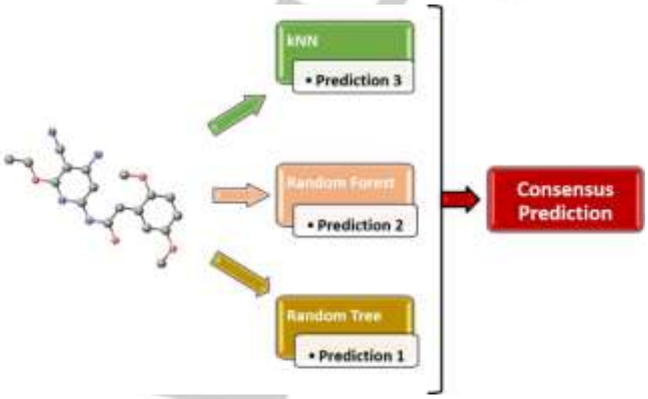


Figure 2. Consensus modeling based on classification majority vote.

The proposed models were fully validated based on OECD principles, following the validation process described in the Experimental Section. The confusion matrix as well as specificity, sensitivity, precision and accuracy for all three proposed models, and the consensus model for the test set are presented below (Tables 2-3):

Table 2. Confusion matrix (test set).

Model I - Random Tree	*Active*	*Inactive*
Active	11	4
Inactive	4	6
Model II - kNN	*Active*	*Inactive*
Active	11	1
Inactive	5	5
Model III - Random Forest	*Active*	*Inactive*
Active	11	4
Inactive	3	7
Consensus Model	*Active*	*Inactive*
Active	14	1
Inactive	3	7

Table 3. Model validation results (test set).

	Specificity	**Sensitivity**	**Precision**	**Accuracy**

10.1002/cmdc.201700675

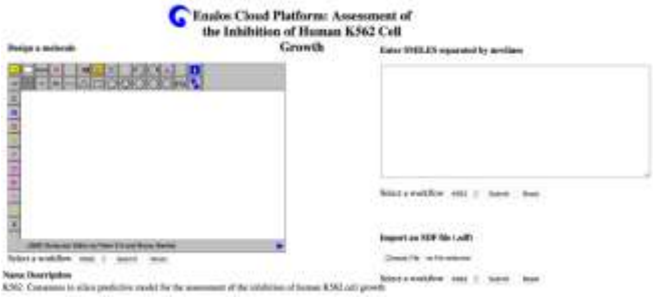
Model I - Random Tree	0.6	0.733	0.733	0.68
Model II - kNN	0.5	0.933	0.737	0.76
Model III - Random Forest	0.7	0.733	0.786	0.72
Consensus Model	0.7	0.933	0.824	0.84

As it can be seen from the validation results, the consensus model based on the majority vote was proven to outperform all others, having the higher accuracy (84%). Thus, this model was finally proposed for the prediction of K562 cell growth inhibition and was used for further exploitation through the Enalos Cloud Platform.

When proposing a validated model, it is very important to simultaneously define its limits so that a well-defined applicability domain indicates those predictions that can be considered reliable.[75-76] When the model is used to screen new compounds, it is important that structures falling out of the domain of applicability of the model are filtered, as the model cannot generate reliable predictions for these structures. Within this context, the domain of applicability of the proposed model was defined based on Euclidean distances using the Enalos Euclidean Domain KNIME node as described in the Experimental Section.[77]

As it has been highlighted, the development of a predictive model could end up being useless, unless it is delivered as a user-friendly tool to ensure sustainability. Based on this, we have designed and implemented our proposed strategy in KNIME that allowed us to easily incorporate the proposed model into Enalos Cloud Platform. Enalos Cloub Platform was developed with the purpose to make our models available to the interested user wishing to generate evidence on potential effects in a decision-making framework. For this work, and based on the final consensus model that was proposed, we have created a web service dedicated to the prediction of K562 cell growth inhibition and is accessible through http://enalos.insilicotox.com/K562/. A simple and user-friendly interface was developed (Figure 3) that allowed the interested user to submit and virtually screen one or several compounds.

Three different options are available for submitting a structure that include: (i) Drawing a structure with the available sketcher;[78] compounds can be easily generated and modified to create a set of structures that can be first visualized and then submitted, (ii) Submitting the SMILES notation for one or many compounds at the form available, (iii) Submitting an .sdf file including a batch of compounds. These different options are indicated with an arrow (Figure 3).



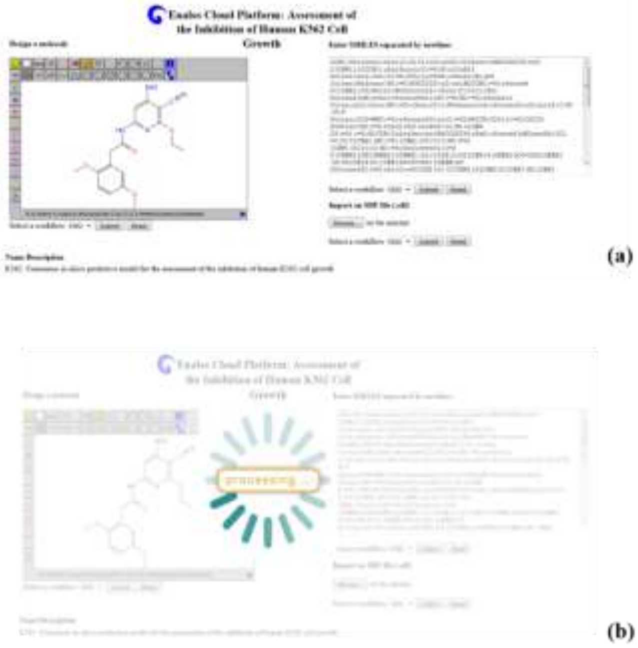
Enalos Cloud Platform: Assessment of the Inhibition of Human K562 Cell Growth

K562: Consensus in silico predictive model for the assessment of the inhibition of human K562 cell growth

Figure 3. A web service dedicated to the prediction of human K562 cell growth inhibition through Enalos Cloud Platform.

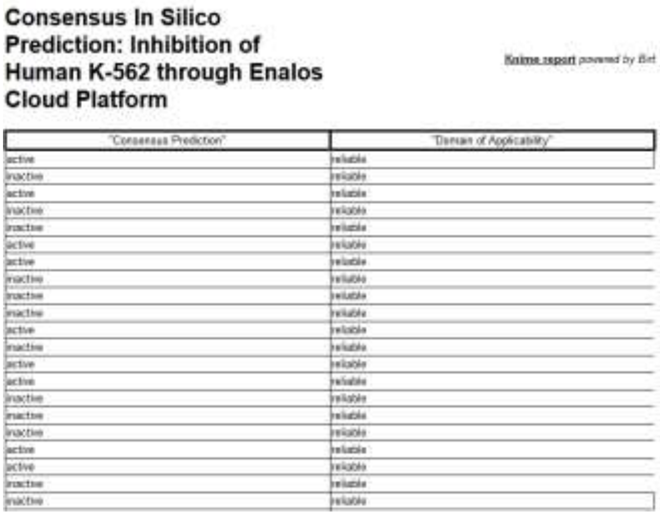
After importing the structures with one of the options described, the workflow of interest must be selected among the workflows available and the submit button must be pressed. This step is indicated in Figure 3 with a dashed arrow. When structures are submitted the results page will appear within seconds. Results page includes a class prediction for each of the structures submitted and an indication of whether this prediction can be considered as reliable or not based on the domain of applicability.

As an example, for a set of structures included in our initial dataset, the input, processing, and results page are shown in Figures 4 and 5.



Enalos Cloud Platform: Assessment of the Inhibition of Human K562 Cell Growth

Figure 4. Compounds included in the initial dataset are **(a)** submitted in the web service, and **(b)** processed.



Consensus In Silico Prediction: Inhibition of Human K-562 through Enalos Cloud Platform

Knime report powered by Birt

"Consensus Prediction"	"Domain of Applicability"
active	reliable
inactive	reliable
active	reliable
inactive	reliable
inactive	reliable
active	reliable
active	reliable
inactive	reliable
inactive	reliable
inactive	reliable
active	reliable
inactive	reliable
active	reliable
active	reliable
inactive	reliable
inactive	reliable
inactive	reliable
active	reliable
active	reliable
inactive	reliable
inactive	reliable

Figure 5. Results page including prediction and domain of applicability assessment for structures included in the dataset.

This web service dedicated to the proposed model can easily facilitate the virtual screening of new structures that fall within its domain of applicability.

Overall, we have computationally explored an extensive dataset of 129 compounds that were experimentally evaluated as human K562 cell growth inhibitors. Within a systematic *in silico* exploitation of this endpoint, we succeeded to afford three robust and validated predictive models. Moreover, a consensus approach based on the majority vote concept was applied to deliver a fourth model that outperformed all others. Our cheminformatics workflow was implemented within KNIME that provided a flexible framework allowing fast and easy experimentation with a wide range of computational techniques. This was significantly facilitated using Enalos and Enalos+ KNIME nodes that performed several critical tasks for data retrieval and exploitation.

The final consensus model was released online as a web service through Enalos Cloud Platform. This facilitated the dissemination of the model to the wider scientific community and allowed fast predictions of K562 cell growth inhibition.[79-80] The web service was designed with a user-friendly interface so that non-experts could easily assess predictions with no prior computational knowledge required. Compounds can be submitted in different formats and results can be obtained very fast and for a large number of compounds. This dedicated web service ensured the model's sustainability and made the high throughput virtual screening of new compounds feasible. Moreover, considering the fact that several erythroid differentiation and HbF inducers also display anti-proliferative effects on the K562 cells system, this approach might be considered as a first screening of potential molecules for further validation using as final experimental system erythroid precursor cells isolated from β–thalassemia patients.

Conclusions

Within this work, we have developed a robust model for the prediction of K562 cell growth inhibition that has been implicated to β-thalassemia disease. First, we have collected and curated a comprehensive dataset of 129 compounds available through PubChem and then we have built a cheminformatics workflow based on KNIME to deliver a fully validated predictive model. To achieve that, we have employed our in house Enalos and Enalos+ KNIME nodes that perform several important tasks essential for data retrieval and exploitation. Our final model was based on a consensus approach considering the prediction results from three individual predictive models. The proposed consensus model was subsequently made available online through Enalos Cloud Platform. For this purpose, a dedicated web service with a user-friendly interface was built to ensure the model's sustainability and to allow the high throughput virtual screening of any given set of untested compounds that fall within its domain of applicability.

Experimental Section

Cheminformatics Workflow—Enalos KNIME nodes

Within this work, a cheminformatics workflow has been proposed to develop a predictive *in silico* model that would provide structural insights of molecules that inhibit the K562 cell growth, allow accurate classification of compounds among the active and inactive class and facilitate the virtual screening of any given set of structures. Our modelling strategy included the following steps: data preprocessing, descriptors calculation, variable selection, model development and validation and domain of applicability determination.

To deliver a robust and validated model that would be easily expandable as a user-friendly tool for end users interested in K-562 cell growth evaluation, we have chosen to combine all various components essential for model development within KNIME platform.[70] Apart from the nodes already available in KNIME, we have also employed our Enalos KNIME nodes that offer a great variety of significant additional functionalities essential for model development.[68]

The Enalos family of KNIME nodes can facilitate several important tasks that are crucial within the cheminformatics framework. Enalos and Enalos+ nodes significantly contribute in data search and retrieval from various databases (among which the PubChem database), and also in predictive model development and validation. Enalos KNIME nodes are freely accessible through either NovaMechanics website or the KNIME platform,[81] and can be used for a wide range of applications including: Calculation of Mold2 molecular descriptors (Enalos Mold2 node), 2) Validation of the Quality-of-Fit and estimation of the predictive ability of a model (Enalos Model Acceptability Criteria node), and 3) Definition of the model's domain of applicability (Enalos Domain–Similarity/Leverages nodes).

Enalos+ KNIME nodes[68] offer additional functionalities including access and retrieval of data (i.e., within PubChem and UniChem) as well as more functionalities for model development and validation (i.e., separation of training and test set, and validation through Y-randomization). More information on the Enalos+ nodes can be found at http://enalosplus.novamechanics.com.

Dataset

A dataset of 129 diverse small molecules, which have been tested as potential K562 inhibitors was collected from PubChem (Table S1) to initiate model development. Data retrieved from PubChem database included the evaluation results of a functional K562 assay that was performed *in vitro* and in human erythroleukemia cells for each of the 129 small molecules (https://pubchem.ncbi.nlm.nih.gov/bioassay/742260).
Compounds that were experimentally evaluated with an activity of ≤ 50 μM were reported as active and all others were reported as inactive. Therefore, among compounds available, 67 and 62 compounds were classified as active and inactive, respectively.

Descriptors Calculation – Enalos Mold2 KNIME node

Appropriate descriptors that determine the structural features of compounds are necessary components for a successful QSAR development. For this purpose, we employed the Mold2 software (National Center for Toxicological Research, FDA), which has been particularly successful in several applications.[72] Mold2 uses 2D structural information to calculate molecular descriptors very fast. Within our KNIME workflow, we have added the Enalos Mold2 KNIME node,[82] which enables the calculation of 777 descriptors that account for the structural, geometric, and topological features of the compounds.[83] The available data were curated and preprocessed and the initial data set was divided into training and test set. Data preprocessing included normalization as well as refinement of variables based on their discrimination power. After an initial screening, many descriptors were removed because of their low discrimination power; for this purpose, the 'Low Variance Filter' node was employed.[84]

Among the descriptors selected from the different variable selection methodologies used in this work, the Balaban index and its variations were predominant and are briefly described below. One of the most widely used molecular descriptor is the Balaban distance connectivity index (J), which is practically invariant with small differences in molecule size. J is defined by the following equation:

$$J = \frac{A}{B+1}\sum_{i=1}^{C-1}\sum_{j=i+1}^{C} c_{ij}\left(\sigma_i\sigma_j\right)^{-\frac{1}{2}} \qquad (1)$$

where A is the number of graph edges, B is the number of rings in the molecule, (B+1 contributes to the normalization against the number of the rings), C is the number of graph vertices, σ_i and σ_j are the vertex distance degrees (row sums of the distance matrix) of vertices y_i and y_j, respectively, and c_{ij} are the elements of the "adjacency matrix" (c_{ij}=1 for adjacent vertices pairs, and c_{ij}=0 for all other cases).

Balaban-like molecular descriptors can be calculated in a similar way as distance connectivity indices (J), by substituting σ_i elements with row summations of "graph-theoretical matrices" (Vs_i) or with "local vertex invariants" (L_i). A Balaban-like index is then expressed:[73]

$$J(\boldsymbol{N};w) = \frac{A}{B+1}\sum_{i=1}^{C-1}\sum_{j=i+1}^{C} c_{ij}\left(Vs_i(\boldsymbol{N};w)\,Vs_j(\boldsymbol{N};w)\right)^{-\frac{1}{2}} \qquad (2)$$

where ***N*** is a graph-theoretical matrix, *w* is the "weighting scheme", and *Vsᵢ* is the "vertex sum operator", which is applied on ***N***.

Balaban-like indices can be more generally expressed according to local vertex invariants as follows:

$$J(L) = \frac{A}{B+1}\sum_{i=1}^{C-1}\sum_{j=i+1}^{C} c_{ij}\left(L_iL_j\right)^{-\frac{1}{2}} \qquad (3)$$

Variable Selection and Model Development

Initially, an attribute selection method based on the training data was employed to select the most important variables among the set of originally determined descriptors. The InfoGainAttributeEval and GainAttributeEval evaluators were used in combination with the Ranker search method (Ranker is a method that produces a ranked list of attributes for attribute evaluators). More details on these methodologies can be found in the literature.[85-86]

Subsequently, different modeling methodologies were used. Among the variety of combinations with the available attribute selection methods, three machine learning modeling methodologies were highlighted as most appropriate for the specific dataset, namely kNN, Random Forest, and Random Tree. All three methodologies have been incorporated into our KNIME workflow.

The kNN approach is part of the instance-based (or lazy) learning, where objects are classified according to the closest training examples in the feature space. An object is categorized by the majority vote of its neighbors, and being assigned to the most common class amongst its k nearest neighbors (k is a positive integer, usually small). Here, we have considered an optimal k value and Euclidean distance, with all descriptors and contributions of neighbors weighted by the inverse of distance.

The discrimination among different classes was also achieved with the Random Tree (RT) classification technique as implemented in WEKA66 program. A supervised approach to classification may be represented by decision trees, where a simple structure is composed of root, nodes, ranches and leaves. The first node corresponds to a root. Typically, a decision tree is constructed starting from the root downward. Nonterminal nodes stand for tests on attributes, and each node relates to a specific feature. Nodes are interconnected through two or more branches projecting from each node; each branch represents a range of values that divide the set of values of the selected feature. Terminal nodes are named leaves and denote decision outcomes. A Random Tree is generated from a set of possible trees using different characteristics at each node. The WEKA implementation of the random decision tree algorithm yields a decision RT without pruning and taking into account only the $\log_2(N)$ at each node, where N is the number of available descriptors. Random Tree generation is an efficient process and the accumulation of large RT sets usually results in accurate models.[87]

While trees constructed by Random Tree are taken from a set containing a group of random features at each node, Random Forests can be generated through bagged ensembles containing Random Trees.[86] Random Forest machine learning methodology combines the results provided by several individual decision trees that are grown based on samples from the initial dataset.[88] For each tree a random number of attributes, that form the nodes and leafs, are chosen. Random Forest was developed based on four variables for model development and predictions were made by averaging predicted activities over all trees in the final forest. In this work, the implementation of the random forest algorithm available in WEKA [25] was used.

In addition to the available models developed, a consensus approach was also used based on the “majority vote” among the results produced by each individual model. When a class, active or inactive, was predicted for a given compound from the majority (two or three) of the available models, then this class was considered as the final assessment from the consensus model.

Model validation

The predictive model was validated according to the criteria proposed by the Organization for Economic Cooperation and Development (OECD).[89] Specifically, goodness-of-fit, robustness and predictivity were considered to (internally and externally) validate the model. As described above, the dataset was split into training and validation sets. During model development, we considered compounds included in the validation set as a “blind set”.[90]

To validate the performance of the model, the following criteria[91-92] were calculated:

$$\text{Precision} = TP / (TP+FP) \tag{4}$$

$$\text{Sensitivity} = TP / (TP+FN) \tag{5}$$

$$\text{Specificity} = TN / (TN+FP) \tag{6}$$

$$\text{Accuracy} = (TP+TN) / (TP+FP+FN+TN) \tag{7}$$

where: TP = True Positive, FP = False Positive, TN = True Negative, FN = False Negative.

Furthermore, a Y-randomization test was performed to validate the robustness and the statistical significance of the generated models using the Enalos+ KNIME nodes.[68] During this test, the dependent variable vector (rather than the independent matrix) is randomly shuffled to yield a new model. The procedure is repeated many times and the accuracy and other statistical parameters of the produced model are assessed and are expected to be diminished compared with those of the primary model. By applying this approach, the statistical significance of the model can be confirmed. If the model does not pass Y-randomization test, this is an indication of poor statistical significance, and a robust predictive model cannot be generated by the particular modeling method and training set.

Definition of the applicability domain with the Enalos Domain–Similarity node

Reliable predictions of a model fall within its domain of applicability, which is constructed using similarity measurements based on the Euclidean distances among all training molecules. The distance between a test molecule and its closest neighbor in the training set is compared to a predefined threshold (APD) and the prediction is considered reliable if the distance is lower than this value. The calculation of APD was based on the following relationship:

$$APD = \langle d \rangle + Z\sigma \quad (8)$$

The values of <d> and σ were estimated as follows: initially, the average of Euclidean distances between all pairs of training molecules was calculated. Next, the set of distances, which were below the average distance was formed. Finally, <d> and σ were calculated as the average and standard deviation, respectively, of all distances included in this set. Z is an empirical cutoff and its value was set to 0.5.[92-95] The Enalos+ Domain–Similarity node is included in our workflow and was used to assess the domain of applicability of the proposed model.[92-95]

Enalos cloud platform

The Enalos Cloud Platform[96-97] is a service that embraces a number of predictive models for drug discovery and risk assessment.[79] Above, we described the development of a predictive consensus model for K562 inhibition. To our knowledge, this is the first ligand-based model constructed from an extensive dataset of K562 inhibitors, and therefore we decided to release it as a free web service to facilitate the virtual screening and design of new effective small-molecule inhibitors of K562, by providing prompt access to the model's results. The model can be accessed through http://enalos.insilicotox.com/K562. One can apply the methodology through a user-friendly graphical interface following a minimum-step procedure. A structure can be submitted by using one of the following ways: (i) manually draw the structure of a molecule using the sketcher provided in the platform,[78] (ii) submit a SMILES file of a molecule, or (iii) upload a structure-data file (sdf). It is noted that more than one molecules can be simultaneously submitted. Next, the K562 workflow is selected from the menu, and a prediction is generated. The outcomes offer the predicted classification of selected molecules and an indication on whether this prediction is reliable or not based on the domain of applicability. Screenshots of the Enalos online tool for K562 inhibition prediction are presented in the Results and Discussion Section.

Acknowledgements

This work is supported by funding under the Seven Research Framework Programme of the European Union. Project THALAMOSS (HEALTH.2012.1.2-1 Grant agreement no: 306201).

.

Keywords: β-thalassemia • HbF inducers • Human K562 cell line • Predictive modeling • Cheminformatics

References:

[1] B. Giardine, J. Borg, E. Viennas, C. Pavlidis, K. Moradkhani, P. Joly, M. Bartsakoulia, C. Riemer, W. Miller, G. Tzimas, H. Wajcman, R. C. Hardison, G. P. Patrinos, *Nucleic Acids Research* **2014**, *42*, D1063-D1069.

[2] J. M. Old, *Blood Reviews*, *17*, 43-53.

[3] R. Colah, A. Gorakshakar, A. Nadkarni, *Expert Review of Hematology* **2010**, *3*, 103-117.

[4] L. Quek, S. L. Thein, *Brit J Haematol* **2007**, *136*, 353-365.

[5] R. Gambari, E. Fibach, *Current Medicinal Chemistry* **2007**, *14*, 199-212.

[6] D. J. Weatherall, *Baillieres Clin Haematol* **1998**, *11*, 127-146.

[7] D. J. Weatherall, *Nature reviews. Genetics* **2001**, *2*, 245-255.

[8] C. Goss, P. Giardina, D. Degtyaryova, D. Kleinert, S. Sheth, M. Cushing, *Transfusion* **2014**, *54*, 1773-1781.

[9] J. G. Michlitsch, M. C. Walters, *Current molecular medicine* **2008**, *8*, 675-689.

[10] A. Finotti, L. Breda, C. W. Lederer, N. Bianchi, C. Zuccato, M. Kleanthous, S. Rivella, R. Gambari, *Journal of Blood Medicine* **2015**, *6*, 69-85.

[11] A. Finotti, R. Gambari, *Expert Opinion on Biological Therapy* **2014**, *14*, 1443-1454.

[12] R. Gambari, *Expert Opinion on Biological Therapy* **2012**, *12*, 443-462.

[13] S. L. Thein, *Blood Reviews*, *26*, S35-S39.

[14] A. El-Beshlawy, M. Hamdy, M. El Ghamrawy, *Hemoglobin* **2009**, *33*, S197-S203.

[15] S. P. Perrine, B. S. Pace, D. V. Faller, *Hematol Oncol Clin North Am* **2014**, *28*, 233-248.

[16] G. Atweh, H. Fathallah, *Hematol Oncol Clin North Am* **2010**, *24*, 1131-1144.

[17] E. Fibach, E. Prus, N. Bianchi, C. Zuccato, G. Breveglieri, F. Salvatori, A. Finotti, M. Lipucci di Paola, E. Brognara, I. Lampronti, M. Borgatti, R. Gambari, *Int J Mol Med* **2012**, *29*, 974-982.

[18] E. Fibach, N. Bianchi, M. Borgatti, C. Zuccato, A. Finotti, I. Lampronti, E. Prus, C. Mischiati, R. Gambari, *European Journal of Haematology* **2006**, *77*, 437-441.

[19] J. Ju, Y. Wang, R. Liu, Y. Zhang, Z. Xu, Y. Wang, Y. Wu, M. Liu, L. Cerruti, F. Zou, C. Ma, M. Fang, R. Tan, S. M. Jane, Q. Zhao, *Nucleic Acids Research* **2014**, *42*, 9740-9752.

[20] W. Deng, J. W. Rupon, I. Krivega, L. Breda, I. Motta, K. S. Jahn, A. Reik, P. D. Gregory, S. Rivella, A. Dean, G. A. Blobel, *Cell*, *158*, 849-860.

[21] V. G. Sankaran, T. F. Menne, J. Xu, T. E. Akie, G. Lettre, B. Van Handel, H. K. A. Mikkola, J. N. Hirschhorn, A. B. Cantor, S. H. Orkin, *Science* **2008**, *322*, 1839-1842.

[22] V. G. Sankaran, J. Xu, T. Ragoczy, G. C. Ippolito, C. R. Walkley, S. D. Maika, Y. Fujiwara, M. Ito, M. Groudine, M. A. Bender, P. W. Tucker, S. H. Orkin, *Nature* **2009**, *460*, 1093-1097.

[23] D. Zhou, K. Liu, C.-W. Sun, K. M. Pawlik, T. M. Townes, **2010**, *42*, 742.

[24] J. Jiang, S. Best, S. Menzel, N. Silver, M. I. Lai, G. L. Surdulescu, T. D. Spector, S. L. Thein, *Blood* **2006**, *108*, 1077-1083.

[25] X. S. Xu, X. Hong, G. Wang, *Journal of Hematology & Oncology* **2009**, *2*, 15-15.

[26] A. Finotti, J. Gasparello, G. Breveglieri, L. C. Cosenza, G. Montagner, A. Bresciani, S. Altamura, N. Bianchi, E. Martini, E. Gallerani, M. Borgatti, R. Gambari, *Experimental hematology* **2015**, *43*, 1062-1071.e1063.

[27] K. Trakarnsanga, M. C. Wilson, W. Lau, B. K. Singleton, S. F. Parsons, P. Sakuntanaga, R. Kurita, Y. Nakamura, D. J. Anstee, J. Frayne, *Haematologica* **2014**, *99*, 1677-1685.

[28] S. L. Thein, S. Menzel, M. Lathrop, C. Garner, *Human Molecular Genetics* **2009**, *18*, R216-R223.

[29] J. Xu, C. Peng, V. G. Sankaran, Z. Shao, E. B. Esrick, B. G. Chong, G. C. Ippolito, Y. Fujiwara, B. L. Ebert, P. W. Tucker, S. H. Orkin, *Science* **2011**, *334*, 993-996.

[30] R. Renella, A. Perlov, C. E. Harris, D. E. Bauer, J. Xu, S. Guda, M. D. Milsom, S. H. Orkin, D. A. Williams, *Blood* **2012**, *120*, 753-753.

[31] M. Roosjen, B. McColl, B. Kao, L. J. Gearing, M. E. Blewitt, J. Vadolas, *The FASEB Journal* **2014**, *28*, 1610-1620.

[32] S. Guda, C. Brendel, R. Renella, P. Du, D. E. Bauer, M. C. Canver, J. K. Grenier, A. W. Grimson, S. C. Kamran, J. Thornton, H. de Boer, D. E. Root, M. D. Milsom, S. H. Orkin, R. I. Gregory, D. A. Williams, *Molecular Therapy* **2015**, *23*, 1465-1474.

[33] D. Samid, A. Yeh, P. Prasanna, *Blood* **1992**, *80*, 1576-1581.

[34] O. Witt, S. Mönkemeyer, G. Rönndahl, B. Erdlenbruch, D. Reinhardt, K. Kanbach, A. Pekrun, *Blood* **2003**, *101*, 2001-2007.

[35] E. R. Macari, C. H. Lowrey, *Blood* **2011**, *117*, 5987-5997.

[36] Y. He, G. Rank, M. Zhang, J. Ju, R. Liu, Z. Xu, F. Brown, L. Cerruti, C. Ma, R. Tan, S. M. Jane, Q. Zhao, *Journal of Translational Medicine* **2013**, *11*, 14-14.

[37] N. Y. H. Ng, C. H. Ko, *International Scholarly Research Notices* **2014**, *2014*, 123257.

[38] N. Bianchi, C. Zuccato, I. Lampronti, M. Borgatti, R. Gambari, *Evidence-based complementary and alternative medicine : eCAM* **2009**, *6*, 141-151.

[39] S. Zein, W. Li, V. Ramakrishnan, T.-F. Lou, S. Sivanand, A. Mackie, B. Pace, *Experimental Biology and Medicine* **2010**, *235*, 1385-1394.

[40] M. V. R. Reddy, V. R. Pallela, S. C. Cosenza, M. R. Mallireddigari, R. Patti, M. Bonagura, M. Truongcao, B. Akula, S. S. Jatiani, E. P. Reddy, *Bioorgan Med Chem* **2010**, *18*, 2317-2326.

[41] V. Desplat, M. Vincenzi, R. Lucas, S. Moreau, S. Savrimoutou, S. Rubio, N. Pinaud, D. Bigat, E. Enriquez, M. Marchivie, S. Routier, P. Sonnet, F. Rossi, L. Ronga, J. Guillon, *ChemMedChem* **2017**, *12*, 940-953.

[42] R. Rondanin, D. Simoni, M. Maccesi, R. Romagnoli, S. Grimaudo, R. M. Pipitone, M. Meli, A. Cascio, M. Tolomeo, *ChemMedChem* **2017**, *12*, 1183-1190.

[43] R. K. Tiwari, A. Brown, N. Sadeghiani, A. N. Shirazi, J. Bolton, A. Tse, G. Verkhivker, K. Parang, G. Sun, *ChemMedChem* **2017**, *12*, 86-99.

[44] C. Lozzio, B. Lozzio, *Blood* **1975**, *45*, 321-334.

[45] J. F. Dorsey, J. M. Cunnick, S. M. Mane, J. Wu, *Blood* **2002**, *99*, 1388-1397.

[46] A. Jacquel, M. Herrant, L. Legros, N. Belhacene, F. Luciano, G. Pages, P. Hofman, P. Auberger, *The FASEB Journal* **2003**.

[47] E. Brognara, I. Lampronti, G. Breveglieri, A. Accetta, R. Corradini, A. Manicardi, M. Borgatti, A. Canella, C. Multineddu, R. Marchelli, R. Gambari, *European Journal of Pharmacology* **2011**, *672*, 30-37.

[48] T. R. Rutherford, J. B. Clegg, D. J. Weatherall, *Nature* **1979**, *280*, 164-165.

[49] T. Rutherford, J. B. Clegg, D. R. Higgs, R. W. Jones, J. Thompson, D. J. Weatherall, *P Natl Acad Sci USA* **1981**, *78*, 348-352.

[50] N. F. Olivieri, D. J. Weatherall, *Human Molecular Genetics* **1998**, *7*, 1655-1658.

[51] G. I. Passeri, D. Trisciuzzi, D. Alberga, L. Siragusa, F. Leonetti, G. F. Mangiatordi, O. Nicolotti, *International Journal of Quantitative Structure-Property Relationships (IJQSPR)* **2018**, *3*, 134-160.

[52] E. Vrontaki, G. Melagraki, S. Voskou, M. S. Phylactides, T. Mavromoustakos, M. Kleanthous, A. Afantitis, *Mini-Reviews in Medicinal Chemistry* **2017**, *17*, 188-204.

[53] G. Rivera, S. Adrade-Ochoa, M. S. O. Romero, I. Palos, A. Monge, L. E. Sanchez-Torres, *Anti-cancer agents in medicinal chemistry* **2017**, *17*, 682-691.

[54] J. Monga, S. L. Khokra, A. Husain, *Medicinal Chemistry Research* **2013**, *22*, 1837-1845.

[55] P. Yi, J. Yang, D. Huang, *Advanced Materials Research* **2012**, *554-556*, 1853-1856.

[56] A. T. Taher, K. M. Musallam, M. D. Cappellini, D. J. Weatherall, *Brit J Haematol* **2011**, *152*, 512-523.

[57] G. J. Dover, S. Brusilow, D. Samid, *N Engl J Med* **1992**, *327*, 569-570.

[58] H. Fathallah, G. F. Atweh, *ASH Education Program Book* **2006**, *2006*, 58-62.

[59] G. P. Rodgers , G. J. Dover , N. Uyesaka , C. T. Noguchi , A. N. Schechter , A. W. Nienhuis *New Engl J Med* **1993**, *328*, 73-80.

[60] P. Rigano, A. Pecoraro, R. Calzolari, A. Troia, S. Acuto, D. Renda, G. R. Pantalone, A. Maggio, R. D. Marzo, *Brit J Haematol* **2010**, *151*, 509-515.

[61] L. B. Aguilar-Lopez, J. L. Delgado-Lamas, B. Rubio-Jurado, F. Javier Perea, B. Ibarra, *Blood Cells, Molecules, and Diseases* **2008**, *41*, 136-137.

[62] N. Masera, L. Tavecchia, M. Capra, G. Cazzaniga, C. Vimercati, L. Pozzi, A. Biondi, G. Masera, *Blood Transfusion* **2010**, *8*, 63-65.

[63] R. S. Weinberg, X. Ji, M. Sutton, S. Perrine, Y. Galperin, Q. Li, S. A. Liebhaber, G. Stamatoyannopoulos, G. F. Atweh, *Blood* **2005**, *105*, 1807-1809.

[64] K. Bourantas, G. Economou, J. Georgiou, *European Journal of Haematology* **1997**, *58*, 22-25.

[65] S. T. Singer, E. P. Vichinsky, N. Sweeters, E. Rachmilewitz, *Brit J Haematol* **2011**, *154*, 281-284.

[66] G. Melagraki, E. Ntougkos, V. Rinotas, C. Papaneophytou, G. Leonis, T. Mavromoustakos, G. Kontopidis, E. Douni, A. Afantitis, G. Kollias, *PLOS Computational Biology* **2017**, *13*, e1005372.

[67] G. Melagraki, A. Afantitis, *Current topics in medicinal chemistry* **2015**, *15*, 1827-1836.

[68] *Enalos+ KNIME nodes*, http://enalosplus.novamechanics.com/.

[69] *Enalos+ PubChem KNIME nodes*, http://enalosplus.novamechanics.com/index.php/enalosplusnodes/pubchem/.

[70] *KNIME Analytics Platform*, www.knime.org.

[71] *Enalos+ Molecular Descriptors KNIME Nodes*, http://enalosplus.novamechanics.com/index.php/enalosplusnodes/molecular-descriptors/.

[72] H. Hong, Q. Xie, W. Ge, F. Qian, H. Fang, L. Shi, Z. Su, R. Perkins, W. Tong, *Journal of Chemical Information and Modeling* **2008**, *48*, 1337-1344.

[73] V. C. R. Todeschini, *Molecular Descriptors for Chemoinformatics*, Wiley, Weinheim, **2009**.

[74] B. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, Germany, **2000**.

[75] A. Tropsha, *Mol Inform* **2010**, *29*, 476-488.

[76] A. Afantitis, G. Melagraki, H. Sarimveis, P. A. Koutentis, J. Markopoulos, O. Igglessi-Markopoulou, *QSAR & Combinatorial Science* **2006**, *25*, 928-935.

[77] *Enalos+ Modelling KNIME Nodes*, http://enalosplus.novamechanics.com/index.php/enalosplusnodes/modelling/.

[78] B. Bienfait, P. Ertl, *Journal of Cheminformatics* **2013**, *5*, 24.

[79] G. Melagraki, A. Afantitis, *Comb Chem High Throughput Screen* **2016**, *19*, 260-261.

[80] I. V. Tetko, *J Comput Aid Mol Des* **2012**, *26*, 135-136.

[81] *Enalos KNIME nodes*, http://enalosplus.novamechanics.com/index.php/enalos-nodes/.

[82] G. Melagraki, A. Afantitis, *Chemometrics and Intelligent Laboratory Systems* **2013**, *123*, 9-14.

[83] A. P. Toropova, A. A. Toropov, M. Marzo, S. E. Escher, J. L. Dorne, N. Georgiadis, E. Benfenati, *Food and Chemical Toxicology* **2017**.

[84] P. K. Ojha, K. Roy, *Chemometrics and Intelligent Laboratory Systems* **2011**, *109*, 146-161.

[85] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, *SIGKDD Explor. Newsl.* **2009**, *11*, 10-18.

[86] I. H. Witten, E. Frank, M. a. Hall, *Annals of Physics* **2011**, *54*, 664.

[87] Y. Zhao, Y. Zhang, *Advances in Space Research* **2008**, *41*, 1955-1959.

[88] L. Breiman, *Mach. Learn.* **2001**, *45*, 5-32.

[89] f. r. p. o. OECD Principles for the validation, Q. S. A. R. Models., h. w. o. o. a. 12/20/2013).

[90] P. K. Ojha, K. Roy, *Food and Chemical Toxicology* **2017**.

[91] A. Afantitis, G. Melagraki, P. A. Koutentis, H. Sarimveis, G. Kollias, *Eur J Med Chem* **2011**, *46*, 497-508.

[92] V. D. Mouchlis, G. Melagraki, T. Mavromoustakos, G. Kollias, A. Afantitis, *J Chem Inf Model* **2012**, *52*, 711-723.

[93] H. Liu, X. Yao, P. Gramatica, *Comb Chem High Throughput Screen* **2009**, *12*, 490-496.

[94] E. Papa, S. Kovarich, P. Gramatica, *QSAR & Combinatorial Science* **2009**, *28*, 790-796.

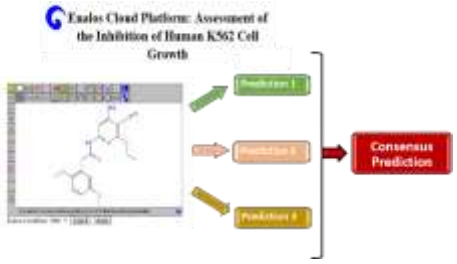
[95] S. Zhang, A. Golbraikh, S. Oloff, H. Kohn, A. Tropsha, *Journal of Chemical Information and Modeling* **2006**, *46*, 1984-1995.

[96] G. Melagraki, A. Afantitis, *RSC Advances* **2014**, *4*, 50713-50725.

[97] G. Melagraki, A. Afantitis, *Current topics in medicinal chemistry* **2015**, *15*, 1827-1836.

10.1002/cmdc.201700675

Entry for the Table of Contents


Enalos Cloud Platform: Assessment of the Inhibition of Human K562 Cell Growth
Prediction 1
Consensus Prediction

Enalos Cloud Platform: A large dataset of K562 biological inhibitors has beeen computationally treated to identify compounds that possibly have therapeutic action against β–thalassemia. A predictive computational model for K562 inhibition was developed and validated. The model facilitates fast and reliable virtual screening of new molecules and is freely available online.