Rational design of isonicotinic acid hydrazide derivatives with anti-tubercular activity: Machine learning, molecular docking, synthesis and biological testing

Vasyl Kovalishyn[a], Julie Grouleff[b], Ivan Semenyuta[a], Vitaliy O. Sinenko[a], Sergiy R. Slivchuk[a], Diana Hodyna[a], Volodymyr Brovarets[a], Volodymyr Blagodatny[c], Gennady Poda[b,d], Igor V. Tetko[e,f,*], Larysa Metelytsia[a]

[a]Institute of Bioorganic Chemistry and Petrochemistry, National Academy of Science of Ukraine, 1 Murmanska Street, 02660, Kyiv, Ukraine

[b]Drug Discovery Program, Ontario Institute for Cancer Research, MaRS Centre, 661 University Avenue, Suite 510, Toronto, Ontario, M5G 0A3, Canada

[c]P.L. Shupyk National Medical Academy of Postgraduate Education, 9 Dorohozhytska Street, 04112, Kyiv, Ukraine

[d]Leslie Dan Faculty of Pharmacy, University of Toronto, 144 College Street, Toronto, Ontario, M5S 3M2, Canada

[e]Helmholtz Zentrum München - German Research Center for Environmental Health (GmbH), Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany

[f]BIGCHEM GmbH, Ingolstädter Landstraße 1, b. 60w, D-85764 Neuherberg, Germany

[*]Corresponding author:

Dr. Igor V. Tetko
Helmholtz Zentrum München - German Research Center for Environmental Health (GmbH), Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany

Tel: +49-8931873575
Fax: +49-89-3187 4584
Email: itetko@vcclab.org

ABSTRACT

The problem of designing new anti-tubercular drugs against multiple-drug-resistant tuberculosis (MDR-TB) was addressed using advanced machine learning methods. Since there are only few published measurements against MDR-TB, we collected a large literature dataset and developed models against the non-resistant H37Rv strain. The predictive accuracy of these models had a coefficient of determination q^2 = 0.7-0.8 (regression models), and balanced accuracies of about 80% (classification models) with cross-validation and independent test sets. The models were applied to screen a virtual chemical library, which was designed to have MDR-TB activity. The seven most promising compounds were identified, synthesized and tested. All of them showed activity against the H37Rv strain, and three molecules demonstrated activity against the MDR-TB strain. The docking analysis indicated that the discovered molecules could bind enoyl reductase, InhA, which is required in mycobacterial cell wall development. The models are freely available online (http://ochem.eu/article/103868) and can be used to predict potential anti-TB activity of new chemicals.

Keywords:

- Isonicotinic acid hydrazide derivatives
- Anti-tubercular activity
- *Mycobacterium tuberculosis* (Mtb)
- Machine learning
- OCHEM
- Molecular docking

1. Introduction

Tuberculosis (TB) is an infectious disease primarily caused by the bacterium *Mycobacterium tuberculosis* (Mtb). While it usually affects the lungs, it can also perturb other organs. Almost one third of the world's population is infected with TB, primarily in developing countries. In 2015 about 1.4 million deaths were reported as a direct result of TB infection.[1] Streptomycin, discovered in 1943, was the first antibiotic active against TB. Since then, multiple new antibiotics against TB have been developed, with isoniazid, rifampicin, pyrazinamide and ethambutol being the most commonly used. Unfortunately, Mtb has developed resistance against these antibiotics, and spread of the TB epidemic increases due to multiple drug-resistant tuberculosis (MDR-TB), which cannot be easily cured with existing drugs.[1] Treatment of MDR-TB requires the use of four or more antibiotics over a period of up to 18-24 months. Moreover, the number of cases of MDR-TB resistant to all currently available drugs is on the rise.

The challenges in development of new tuberculosis drugs were recently reviewed.[2] Several chemical series such as quinolone-carbohydrazides,[3] pyrimidines, [4] isoniazid derivatives,[5,6] pyrazolopyridones[7] and others[8] were reported to have inhibitory activity against MDR-TB. Although several new compounds are currently in different stages of clinical trials, only one new drug, Bedaquiline (TMC207), was approved by the FDA as part of a combination therapy for the treatment of pulmonary TB tuberculosis resistant to standard drugs.[9] Thus, the search for new antibiotics, effective against MDR-TB, is very important in drug discovery.[10]

Isonicotinic acid hydrazide (isoniazid or INH) is one of the most efficient drugs for the treatment of Mtb infection.[1] INH is a prodrug that is activated in the Mtb cells by KatG (catalase-peroxidase enzyme). The main target of the activated INH is the enoyl reductase, InhA, which participates in the synthesis of mycolic acid, one of the building blocks of the mycobacterial cell wall.[11] Unfortunately, mutations in KatG decrease activation of INH and cause resistance of Mtb to this drug. Previously,[12,13] direct inhibition of InhA by diazole- and triazole-based compounds and their *in vitro* and *in vivo* anti-tubercular activity was reported. It is known that thiazole is one of the privileged cores in medicinal chemistry, and its derivatives were reported to show antimicrobial, anti-inflammatory, analgesic, antitubercular and anticancer properties.[14,15] In this study we propose to link the isoniazid and thiazole entities into single derivatives. The synergistic effect of such derivatives on two targets can help in overcoming the problem of Mtb resistance to INH.

In order to address this goal, we applied machine learning, molecular docking, virtual screening, synthesis and experimental testing of novel INH derivatives. Since there are no sufficient experimental literature data to develop a model against MDR-TB, we collected data and developed models against Mtb H37Rv. Our assumptions were that such an approach can help us identify active compounds, while linking the isoniazid and thiazole entities could help overcome resistance against the MDR-TB strain. The machine learning was done using classification and regression methods provided by the Online Chemical Modeling Environment (OCHEM).[16] The models along with the data used are freely available online for the scientific community at http://ochem.eu/article/103868.

2. Materials and methods

2.1 Datasets

The data for our analysis were compounds evaluated against the H37Rv TB strain, which were obtained from multiple publications. These data were uploaded into the OCHEM database, which is a user-friendly, web-based platform designed for storing experimental properties and chemical activities with the primary goal of *in silico* modeling.[16]

Two different datasets were used to build the models. The initial dataset **I** (6337 compounds) consisted of diverse chemical series, such as derivatives of azoles, isoniazids, indoles and others. The minimum inhibitory concentration (MIC) values of the molecules ranged from 0.0015 to 99.9 μM. The data were divided into high activity (2705 compounds with MIC ≤ 10 μM) and low activity molecules (3632 compounds with MIC > 10 μM). This dataset was used for the development of global classification models against the H37Rv strain, which were used for a preliminary evaluation of the anti-TB activity of the investigated compounds.

A more specific dataset **II** was generated from the previous set by selection of 510 compounds. This dataset consisted of INH and thiazole derivatives inhibiting the H37Rv TB strain, with MIC values ranging from 0.013 to 98.9 μM. MICs were converted into log(1/MIC) values and were used as the target variable to develop regression models.
For all datasets, about 20-25 % of the compounds were randomly selected by OCHEM to form independent test sets while the remaining molecules were used as training sets. The chemical structures and the corresponding anti-TB activity data of the compounds used in the training and test sets as well as the full list of publications are publicly accessible at http://ochem.eu/article/103868.

2.2 Machine learning methods

Several machine-learning methods available on OCHEM were used to construct *in silico* models based on different descriptor sets.

Associative Neural Network (ASNN). ASNN was inspired by the organization of neuronal connections in the brain[17] and represents a combination of an ensemble of the Feed-Forward Backpropagation Neural Networks and the *k*-Nearest Neighbors (*k*NN) method. While neural networks build an ensemble of global models, *k*NN provides a local correction of the global model set.[18] This combination corrects the bias of the neural network ensemble and increases its accuracy. The ensemble included 100 neural networks, which were developed using the default parameters provided by OCHEM.

k-Nearest Neighbors (kNN). The kNN classifies the target molecule to the majority class of it *k* nearest neighbors. The parameter *k* is automatically selected to maximize kNN performance for the training set.

XGBoost. Extreme Gradient Boosting uses a weighted sum of individual trees in an ensemble.[19] The problem of overfitting is addressed by minimization of the norm of the learnt weights. A greedy algorithm is used to add new branches that most improve the objective function of the algorithm.[19]

WEKA-RF. WEKA (Waikato Environment for Knowledge Analysis)[20] is a collection of machine learning algorithms for data mining tasks. The Random Forest (WEKA-RF) model consists of a set of decision trees, each of which is built using a bootstrap replica of the training set and randomly selected subsets of descriptors.[21] The random forest makes predictions by majority votes of the individual trees.

The OCHEM provides an estimation of the applicability domain and accuracy of predictions,[22] which was used in this study to select reliable predictions of virtual compounds.

2.3 Validation of models

A five-fold cross-validation approach and external validation sets were used to evaluate the models.[23] Frequently, the validation of models in quantitative structure-activity relationship (QSAR) studies is performed after the variable selection. Such an approach can result in incorrect estimation of the predictive power of models due to over-fitting by variable selection.[23] OCHEM avoids this problem by repeating all steps of model development for each cross-validation step. Moreover, the aforementioned test sets were used to further confirm the quality of the models.

2.4 Statistical parameters

Sensitivity and specificity of classification models are calculated as:

$$SN = TP / (TP + FN) \quad (1)$$

$$SP = TN / (TN + FP) \quad (2)$$

Here *TP*, *FP*, *TN* and *FN* denote the number of true positives, false positives, true negatives and false negatives, respectively. In this study we evaluated the performance of the models using the balanced accuracy (also sometimes referred to as the correct classification rate, *BA*), which is calculated as:

$$BA = (SN + SP) /2 \quad (3)$$

In addition to these parameters, OCHEM also provides the confusion matrix (see *Supplementary materials*), which can be used to derive other statistical measures.

The accuracy of the regression models was evaluated using the root mean square error (RMSE), the mean absolute error (MAE), the squared correlation coefficient R^2, and the coefficient of determination q^2.

2.4 Molecular descriptors

OCHEM provides several commonly used software packages for the calculation of vast collections of molecular descriptors. Five packages were used in this study:

E-State indices. The electro-topological state indices are 2D-descriptors that combine both electronic and topological characteristics of the analyzed compounds.[24]

ALogPS. The program calculates the 1-octanol/water partition coefficient and aqueous solubility.[25]

ChemAxon descriptors. The ChemAxon Calculator Plugin supports the calculation of seven descriptor groups, ranging from 0D to 3D: elemental analysis, charge, geometry, partitioning, protonation states and others.

ADRIANA.Code. The software uses a series of methods for the generation of 3D-structures and the calculation of physico-chemical descriptors and molecular properties based on empirical models.[26]

Unsupervised filtering was used with each descriptor set before they were used as an input for the machine learning algorithms. Descriptors with fewer than two unique variables or with a coefficient of variance less than 0.01 were excluded. Moreover, descriptors with a pairwise non-parametric Pearson's correlation coefficient $R > 0.95$ were grouped.

3. Results and discussion

*3.1 Classification models (dataset **I**)*

The initial set of 6337 compounds was randomly split into training (4753) and test (1584) sets. The numerical values of activity were discretized as described in section 2.1. The models were developed after the unsupervised filtering of descriptors as described in section 2.4. Additionally, the Unsupervised Forward Selection (UFS)[27] was used in classification models 1, 2 and 4 (Table 1) to further filter descriptors. The RF algorithm uses random subsets of descriptors to build each tree in the forest and thus it is less subject to the problem of correlations between the descriptors. Therefore, the best WEKA-RF model (see Table 1) was obtained without the UFS.

We preliminarily investigated all descriptors sets available at the OCHEM website. The final models with the highest prediction accuracies were calculated using all four sets of descriptors described in section 2.4.

Table 1. Statistical coefficients calculated for classification models obtained for dataset **I**.

N	Method	Sensitivity (%)		Specificity (%)		Balanced Accuracy (%)	
		Training[a]	Test[a]	Training	Test	Training	Test
1	ASNN	77.9	79.0	83.0	81.0	80.3 ± 0.6	80.0 ± 1.0
2	*k*NN	75.9	77.0	81.5	79.0	78.5 ± 0.6	78.0 ± 1.0
3	WEKA-RF	83.7	85.0	81.6	80.0	81.0 ± 0.6	81.0 ± 1.0
4	XGBOOST	78.8	82.2	82.0	82.0	80.0 ± 0.6	81.2 ± 1.0
5	Consensus[b]	82.5	85.0	82.9	82.0	81.7 ± 0.6	82.2 ± 0.9

[a]The training and test datasets included 4705 and 1569 molecules, respectively. [b]The consensus model was built by averaging four models.

The developed models are summarized in Table 1 and in **Fig. 1S** in the *Supplementary materials*. All the models show similar results in terms of sensitivity, specificity and balanced

accuracy (BA). The cross-validated BAs for the training sets were in the range of 78.5-81 % (Table 1). Similar accuracies were also calculated the test sets *BA* = 78-81.2 % (see **Fig. 1S** of the *Supplementary materials*). The consensus model calculated the highest performance. This model was applied to identify the most promising anti-TB agents in the virtual dataset as described in section 3.3.

*3.2 Regression models (dataset **II**)*

The initial dataset of 510 compounds was split by chance into training (408) and test (102) sets as described. The regression models built by the ASNN and XGBOOST methods (see Table 2) gave the best performance. For this analysis Adriana, E-state and ALOGPS descriptors systematically contributed the top performing models for both methods.

Table 2. Statistical coefficients of the regression models.

N	Method	Training Set[a]			Test Set[a]		
		R^2	q^2	RMSE[c]	R^2	q^2	RMSE
1	ASNN	0.78 ± 0.02	0.77 ± 0.02	0.51 ± 0.02	0.72 ± 0.05	0.70 ± 0.06	0.54 ± 0.05
2	XGBOOST	0.71 ± 0.03	0.71 ± 0.03	0.57 ± 0.03	0.74 ± 0.05	0.73 ± 0.05	0.51 ± 0.04
3	Consensus[b]	0.78 ± 0.02	0.78 ± 0.02	0.50 ± 0.02	0.76 ± 0.04	0.75 ± 0.05	0.49 ± 0.04

[a]The training and test datasets included 408 and 102 molecules, respectively. The same descriptors calculated by the Adriana, E-state and ALOGPS programs were used for both models. [b]The consensus model was simple average of ASNN and XGBOOST models. [c]RMSE is the root mean square error. R^2 and q^2 are the squared linear correlation and coefficient of determination, respectively.

The q^2 values were 0.71-0.78 and 0.70-0.75 for training and test sets, respectively. Other statistical parameters of the models are summarized in Table 2 as well as in **Fig. 2S** of the *Supplementary materials*. A consensus model, which was an average of both models, gave the best performance. It was used to provide a quantitative evaluation of potential antitubercular agents as described in the next section.

3.3 Prediction activity of new compounds

A virtual database of drug-like compounds was generated. It included 165 isoniazid derivatives with different substitution patterns (see **Fig. 1** and *Supplementary materials*, Table 1S). All of these compounds were screened using the consensus classification models, and only compounds predicted to be active were selected for further evaluation (see *Supplementary materials*, Table 2S). The 18 compounds with the most confident predictions (>70%) were selected for the next step.

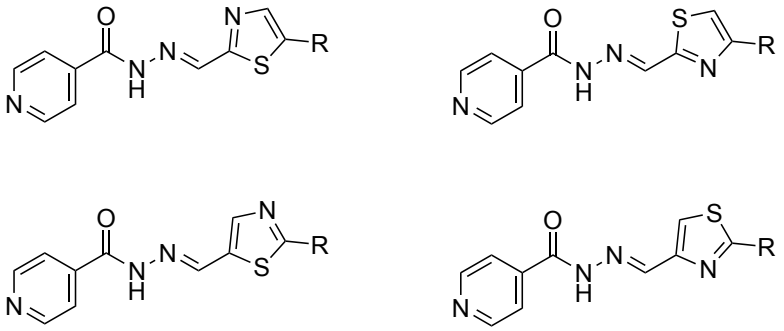


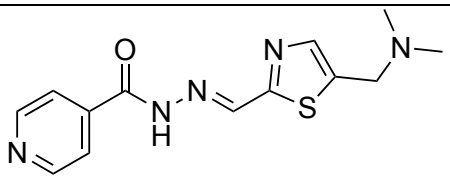
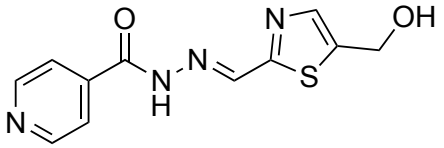
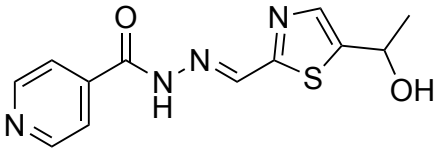
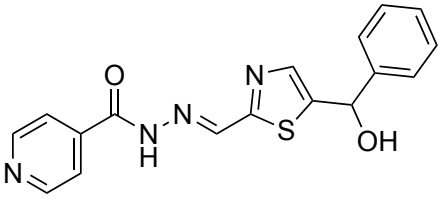
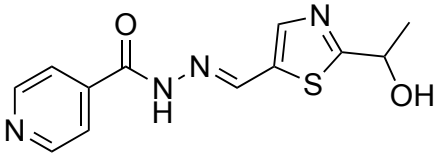
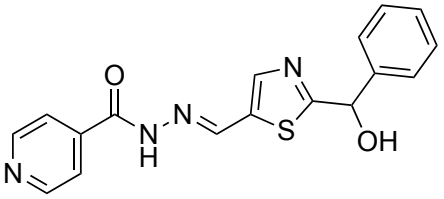
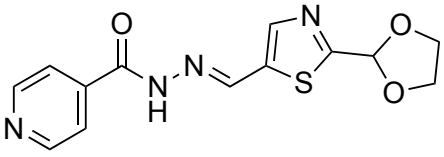
Fig. 1. Virtual library of the 165 thiazole-containing isoniazid derivatives. Substituent "R" was chosen based on the availability of the starting materials for synthesis.

All 18 compounds were also predicted to have anti-TB activities in the range 1-10 μM (i.e., 5-6 on the log(1/MIC) scale) using the consensus regression model (see Table 4 and Table 3S of *Supplementary materials*). Since all of them were active no compounds were excluded in this step.

The next analysis was screening to flag potentially reactive, chemically unstable compounds and compounds with other liabilities using the OCHEM ToxAlerts[28]. The ten compounds with the lowest number of liabilities were proposed for synthesis. From this set, based on their synthetic feasibility, seven compounds were synthesized and tested for their anti-TB activity (Table 3). The results of the biological testing of the synthesized compounds confirmed the QSAR predictions (see Table 4).

3.4 Chemistry

Table 3. Chemical structures of seven synthesized compounds tested for their anti-TB activity**.**

Comp. No.	Molecular weight	Chemical Structure	Chemical Name
8	289.36		*N*'-({5-[(dimethylamino)methyl]-1,3-thiazol-2-yl}methylene)isonicotinohydrazide
9	262.29		*N*'-{[5-(hydroxymethyl)-1,3-thiazol-2-yl]methylene}isonicotinohydrazide
10	276.32		*N*'-{[5-(1-hydroxyethyl)-1,3-thiazol-2-yl]methylene}isonicotinohydrazide
11	338.39		*N*'-({5-[hydroxy(phenyl)methyl]-1,3-thiazol-2-yl}methylene)isonicotinohydrazide
15	276.32		*N*'-{[2-(1-hydroxyethyl)-1,3-thiazol-5-yl]methylene}isonicotinohydrazide
16	338.39		*N*'-({2-[hydroxy(phenyl)methyl]-1,3-thiazol-5-yl}methylene)isonicotinohydrazide
17	304.33		*N*'-{[2-(1,3-dioxolan-2-yl)-1,3-thiazol-5-yl]methylene}isonicotinohydrazide

The detailed description of the chemical synthesis of selected compounds (Table 3) is provided in the supplementary materials. Compound **15** was synthetized by us earlier[29] but had not been tested against Mtb.

It is known that the acylhydrazone-like compounds may not be stable in acidic conditions. We checked the chemical stability of all seven compounds and found that they were stable both at neutral pH as well as under the biological assay conditions.

3.5 Anti-mycobacterial activity

The antimicrobial activity of the selected compounds against the H37Rv and MDR Mtb strains was determined using the proportional method of Canetti (see also supplementary materials).[30] This approach is the most commonly used to study the sensitivity and resistance of tuberculosis strains.[1]

Table 4. Experimental and predicted antimycobacterial activity of the tested compounds.[a]

Compound	Predicted activity	Biological testing	
No.	log(1/MIC)	H37Rv[b]	MDR Mtb[c]
8	4.80 ± 1.70	S[d]	S
9	5.85 ± 0.49	S	S
10	5.95 ± 0.35	S	I[e]
11	5.95 ± 0.21	S	R[f]
15	5.80 ± 0.42	S	I
16	6.00 ± 0.42	S	R
17	4.70 ± 1.60	S	S

[a]The compound concentration was 10 μM. [b]The H37Rv strain is sensitive to isoniazid and rifampicin. [c]The MDR Mtb strain is resistant to isoniazid and rifampicin. [d]S is the sensitive culture. [e]I is the intermediate sensitive culture. [f]R is the resistant culture.

The results of experimental testing are summarized in Table 4, where S, I and R indicate cultures sensitive, intermediate and resistant to the tested compound. All synthesized compounds showed anti-TB activity against H37Rv (Table 4) thus validating the results of the computational analysis. Importantly, the culture of MDR Mtb strains was sensitive to compounds **8, 9** and **17** using operating concentrations (10 μM), which are similar to that of the reference compound rifampicin. Thus, these compounds could serve as prototypes for potentially effective anti-TB compounds, in particular for the treatment of drug resistant forms of tuberculosis.

3.6 Molecular docking studies with enoyl acyl carrier protein reductase (InhA)

After discovery of compounds active against the MDR-TB strain, we hypothesized as to their mechanism of action (MOA). It is known that INH acts as a prodrug that requires catalytic activation by KatG to form an isonicotinic acyl radical that eventually reacts with NADH.[5] The resulting adduct binds strongly to InhA and inhibits its function. Since all the active compounds contain the isoniazid moiety, it is possible that they have similar MOA. However, the compounds can also inhibit InhA directly, which was also hypothesized as the MOA for other types of compounds.[12,13] To investigate whether our compounds could inhibit InhA directly, we performed molecular docking of compounds **8**, **9** and **17** against a crystal structure of InhA (PDB code 4TZK, 1.62 Å) as described in Supplementary Materials. The docking calculations resulted in a similar binding mode for all three compounds (see **Fig. 2**, **3S** and **4S**). The carboxamide moiety forms two H-bonds, one with the 2'-hydroxy group of the NAD ribose and another with the Y158 phenolic group. The pyridine ring forms a favorable charge-π interaction with the

K165 side chain. In addition the complex is stabilized by interactions between the thiazole ring and the aromatic side chains of F149 and Y158. The hydroxyl group of compound **9** also forms a H-bond with the backbone carbonyl of P156. An overlay of the docking poses of all three compounds and the co-crystallized compound from the 4TZK crystal structure of InhA is shown in **Fig. 5S**.

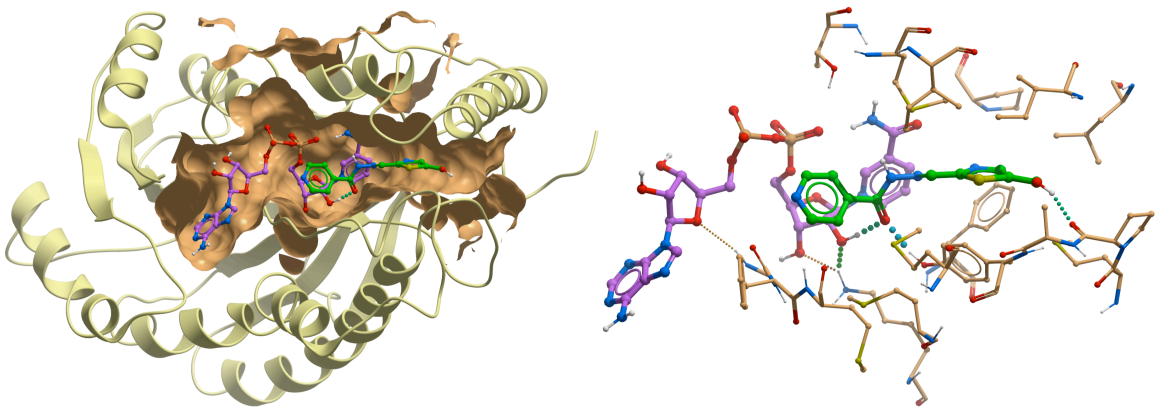


Fig. 2. Molecular docking of compound **9** in the InhA crystal structure (PDB code 4TZK, 1.62 Å). A) Ribbon representation of InhA with NAD (purple) bound and compound **9** (green) docked in the active site; the active site crevice is represented by the orange surface, partially clipped for clarity. B) Details of the interaction between compound **9** (green) and NAD-bound InhA (orange). H-bonds are depicted by dotted lines. NAD is shown in purple.

Since all three compounds fit well in the active site of InhA, it is likely that the observed anti-TB MDR activity of isonicotinic acid hydrazide derivatives (**8**, **9** and **17**) is due to the direct inhibition of InhA. Further experiments are necessary to validate this hypothesis.

4. Conclusions

A set of predictive *in silico* models based on different machine learning techniques and a broad range of molecular descriptors were built using the OCHEM web-based platform. The models demonstrated good stability, robustness and predictive power when verified by cross-validation, prediction of external test sets and by prospective validation involving synthesis and biological testing of seven isonicotinic acid hydrazide derivatives. Of note, all compounds predicted to be active as a result of the machine learning models were found to be active against the H37Rv strain. This result confirms that the expert use of machine learning approaches facilitates a rational search for active molecules within budget and time constrains, which are especially tight in academic settings.[31] The developed models are publicly available and can be used to predict the anti-TB inhibitor activity of new compounds. Importantly, three compounds showed activity against the Mtb resistant strain. They are possible candidates for the development of novel anti-mycobacterial agents. The docking studies suggested that these compounds might directly inhibit the InhA.

Acknowledgements

GP and JG would like to acknowledge the support of the Ontario Institute for Cancer Research, and its funding from the Government of Ontario. We also thank ChemAxon (http://www.chemaxon.com) for academic license of their software used in this study, and David Smil and Michael Withnall for critical reading of the manuscript.

Declaration of interest

IVT is CEO of BIGCHEM GmbH, which licenses the OCHEM software. Other authors declare no conflict of interest.

Supporting Information

The supporting information document contains technical description of experimental (Schemes 1 – 3) and computational procedures (Figures 1S – 5S), and tables with additional information (Tables 1S – 3S).

FIGURE CAPTIONS

Fig. 1. Virtual library of the 165 thiazole-containing isoniazid derivatives. Substituent "R" was chosen based on the availability of the starting materials for synthesis.

Fig. 2. Molecular docking of compound **9** in the InhA crystal structure (PDB code 4TZK, 1.62 Å). A) Ribbon representation of InhA with NAD (purple) bound and compound **9** (green) docked in the active site; the active site crevice is represented by the orange surface, partially clipped for clarity. B) Details of the interaction between compound **9** (green) and NAD-bound InhA (orange). H-bonds are depicted by dotted lines. NAD is shown in purple.

References

[1] Global tuberculosis report 2017. http://www.who.int/tb/publications/global_report/en/

[2] S. Tiberi, R. Buchanan, J. A. Caminero, R. Centis, M. A. Arbex, M. Salazar, J. Potter, G. B. Migliori, *Presse Med.* **2017**, *46*, e41-e51.

[3] K. D. Thomas, A. V. Adhikari, S. Telkar, I. H. Chowdhury, R. Mahmood, N. K. Pal, G. Row, E. Sumesh, *Eur. J. Med. Chem.* **2011**, *46*, 5283-5292.

[4] E. V. Verbitskiy, S. A. Baskakova, M. A. Kravchenko, S. N. Skornyakov, G. L. Rusinov, O. N. Chupakhin, V. N. Charushin, *Bioorg. Med. Chem.* **2016**, *24*, 3771-3780.

[5] F. Martins, S. Santos, C. Ventura, R. Elvas-Leitao, L. Santos, S. Vitorino, M. Reis, V. Miranda, H. F. Correia, J. Aires-de-Sousa, V. Kovalishyn, D. A. Latino, J. Ramos, M. Viveiros, *Eur. J. Med. Chem.* **2014**, *81*, 119-138.

[6] V. Kovalishyn, V. Brovarets, V. Blagodatnyi, I. Kopernyk, D. Hodyna, S. Chumachenko, O. Shablykin, O. Kozachenko, M. Vovk, M. Barus, M. Bratenko, L. Metelytsia, *Curr. Drug Discov. Technol.* **2017**, *14*, 25-38.

[7] M. Panda, S. Ramachandran, V. Ramachandran, P. S. Shirude, V. Humnabadkar, K. Nagalapur, S. Sharma, P. Kaur, S. Guptha, A. Narayan, J. Mahadevaswamy, A. Ambady,

N. Hegde, S. S. Rudrapatna, V. P. Hosagrahara, V. K. Sambandamurthy, A. Raichurkar, *J. Med. Chem.* **2014**, *57*, 4761-4771.
[8] G. Fernandes, C. Man Chin, J. L. Dos Santos, *Pharmaceuticals (Basel)* **2017**, *10*.
[9] R. Mahajan, *Int J Appl Basic Med Res* **2013**, *3*, 1-2.
[10] M. M. Islam, H. M. Hameed, J. Mugweru, C. Chhotaray, C. Wang, Y. Tan, J. Liu, X. Li, S. Tan, I. Ojima, W. W. Yew, E. Nuermberger, G. Lamichhane, T. Zhang, *J Genet Genomics* **2017**, *44*, 21-37.
[11] D. Kumar, Beena, G. Khare, S. Kidwai, A. K. Tyagi, R. Singh, D. S. Rawat, *Eur. J. Med. Chem.* **2014**, *81*, 301-313.
[12] M. Martinez-Hoyos, E. Perez-Herran, G. Gulten, L. Encinas, D. Alvarez-Gomez, E. Alvarez, S. Ferrer-Bazaga, A. Garcia-Perez, F. Ortega, I. Angulo-Barturen, J. Rullas-Trincado, D. Blanco Ruano, P. Torres, P. Castaneda, S. Huss, R. Fernandez Menendez, S. Gonzalez Del Valle, L. Ballell, D. Barros, S. Modha, N. Dhar, F. Signorino-Gelo, J. D. McKinney, J. F. Garcia-Bustos, J. L. Lavandera, J. C. Sacchettini, M. S. Jimenez, N. Martin-Casabona, J. Castro-Pichel, A. Mendoza-Losana, *EBioMedicine* **2016**, *8*, 291-301.
[13] L. A. Spagnuolo, S. Eltschkner, W. Yu, F. Daryaee, S. Davoodi, S. E. Knudson, E. K. Allen, J. Merino, A. Pschibul, B. Moree, N. Thivalapill, J. J. Truglio, J. Salafsky, R. A. Slayden, C. Kisker, P. J. Tonge, *J. Am. Chem. Soc.* **2017**, *139*, 3417-3429.
[14] M. S. Shaikh, M. B. Palkar, H. M. Patel, R. A. Rane, W. S. Alwan, M. M. Shaikh, I. M. Shaikh, G. A. Hampannavar, R. Karpoormath, *RSC Advances* **2014**, *4*, 62308-62320.
[15] P. Makam, R. Kankanala, A. Prakash, T. Kannan, *Eur. J. Med. Chem.* **2013**, *69*, 564-576.
[16] I. Sushko, S. Novotarskyi, R. Korner, A. K. Pandey, M. Rupp, W. Teetz, S. Brandmaier, A. Abdelaziz, V. V. Prokopenko, V. Y. Tanchuk, R. Todeschini, A. Varnek, G. Marcou, P. Ertl, V. Potemkin, M. Grishina, J. Gasteiger, C. Schwab, I. I. Baskin, V. A. Palyulin, E. V. Radchenko, W. J. Welsh, V. Kholodovych, D. Chekmarev, A. Cherkasov, J. Aires-de-Sousa, Q. Y. Zhang, A. Bender, F. Nigsch, L. Patiny, A. Williams, V. Tkachenko, I. V. Tetko, *J. Comput. Aided. Mol. Des.* **2011**, *25*, 533-554.
[17] A. E. Villa, I. V. Tetko, P. Dutoit, Y. De Ribaupierre, F. De Ribaupierre, *J. Neurosci. Methods* **1999**, *86*, 161-178.
[18] I. V. Tetko, *Methods Mol. Biol.* **2008**, *458*, 185-202.
[19] T. Chen, C. Guestrin, *ArXiv e-prints* **2016**, *1603*, arXiv:1603.02754.
[20] E. Frank, M. Hall, L. Trigg, G. Holmes, I. H. Witten, *Bioinformatics* **2004**, *20*, 2479-2481.
[21] L. Breiman, *Machine Learn.* **2001**, *45*, 5-32.
[22] I. Sushko, S. Novotarskyi, R. Körner, A. K. Pandey, V. V. Kovalishyn, V. V. Prokopenko, I. V. Tetko, *J. Chemom.* **2010**, *24*, 202-208.
[23] I. V. Tetko, I. Sushko, A. K. Pandey, H. Zhu, A. Tropsha, E. Papa, T. Oberg, R. Todeschini, D. Fourches, A. Varnek, *J. Chem. Inf. Model.* **2008**, *48*, 1733-1746.
[24] L. B. Kier, L. H. Hall, *Pharm. Res.* **1990**, *7*, 801-807.
[25] I. V. Tetko, V. Y. Tanchuk, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1136-1145.
[26] J. Gasteiger, *J. Med. Chem.* **2006**, *49*, 6429-6434.
[27] D. C. Whitley, M. G. Ford, D. J. Livingstone, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1160-1168.
[28] I. Sushko, E. Salmina, V. A. Potemkin, G. Poda, I. V. Tetko, *J. Chem. Inf. Model.* **2012**, *52*, 2310-2316.

[29] V. O. Sinenko, S. R. Slivchuk, S. G. Pil'o, G. F. Raenko, V. S. Brovarets, *Russian Journal of General Chemistry* **2016**, *86*, 1597-1603.
[30] G. Canetti, S. Froman, J. Grosset, P. Hauduroy, M. Langerova, H. T. Mahler, G. Meissner, D. A. Mitchison, L. Sula, *Bull World Health Organ* **1963**, *29*, 565-578.
[31] F. Ban, K. Dalal, H. Li, E. LeBlanc, P. S. Rennie, A. Cherkasov, *J. Chem. Inf. Model.* **2017**, *57*, 1018-1028.