Cytisine derivatives as new anti-Escherichia coli agents: *in silico* and *in vitro* studies

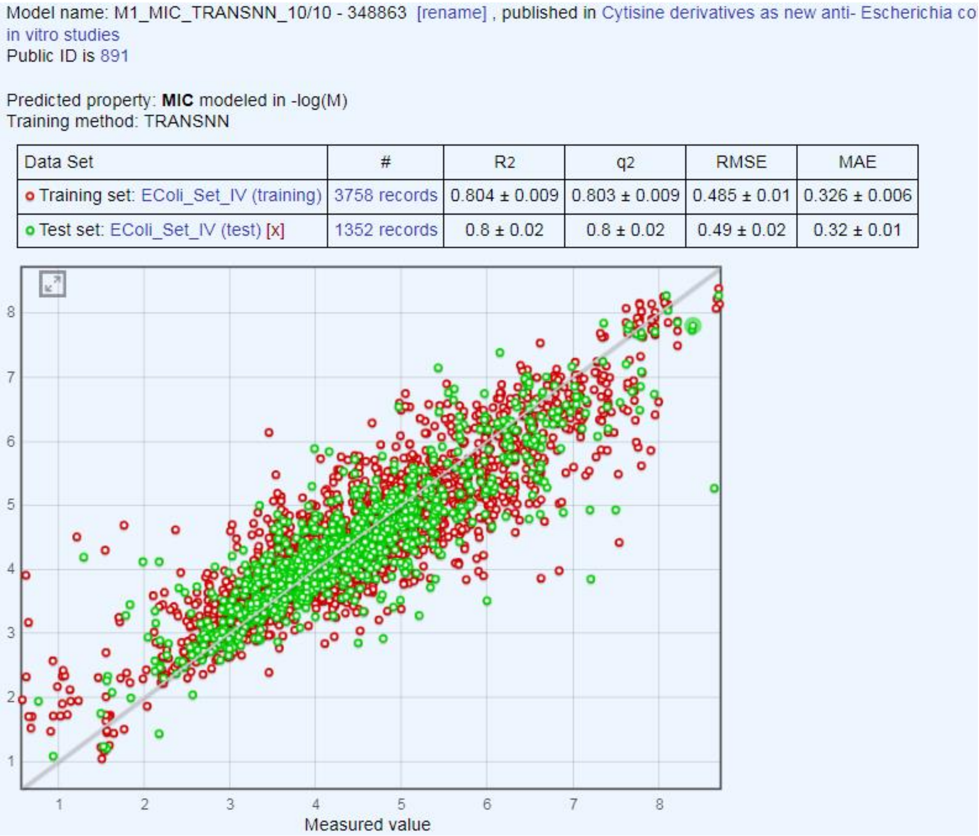
Diana M. Hodyna [1], Vasyl V. Kovalishyn [1], Volodymyr M. Blagodatnyi [1], Svitlana P. Bondarenko [2], Galyna P. Mrug [1], Mykhaylo S. Frasinyuk [1], Larysa O. Metelytsia [1]*

[1] *V.P. Kukhar Institute of Bioorganic Chemistry and Petrochemistry of the NAS of Ukraine, 1 Murmanska St., Kyiv, 02094, Ukraine*

[2] *National University of Food Technologies, 68 Volodymyrska St., Kyiv, 01601, Ukraine*

Results and discussion

Regression models from Table 1

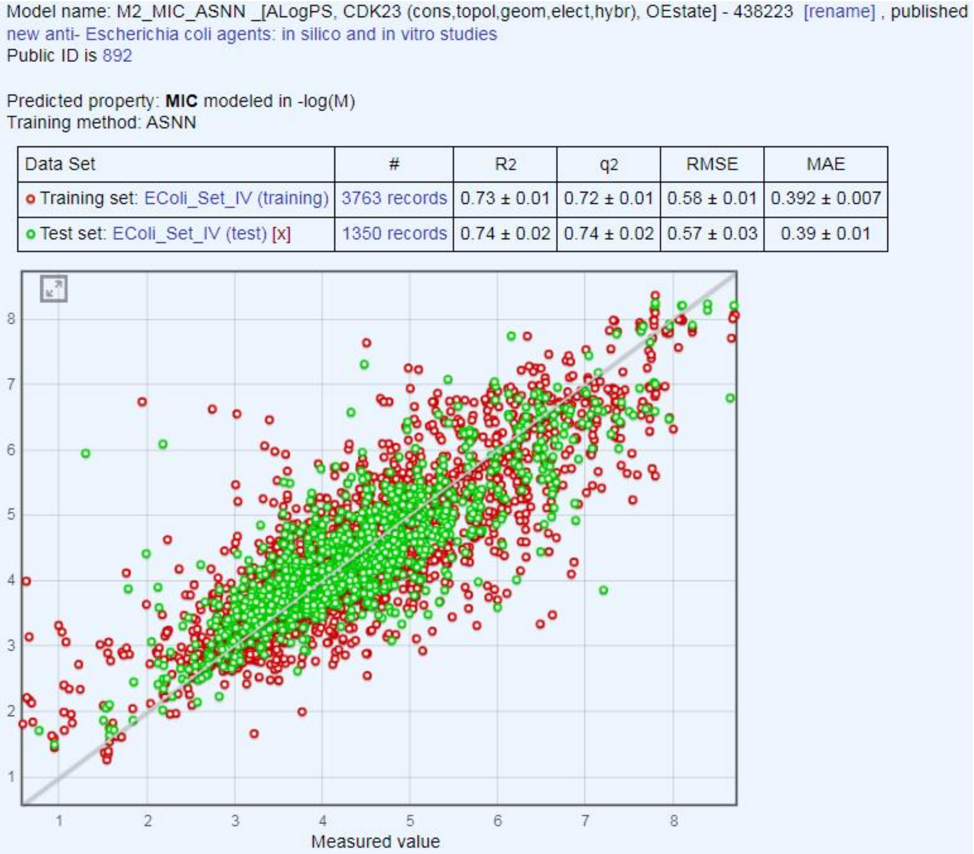


Model name: M1_MIC_TRANSNN_10/10 - 348863 [rename] , published in Cytisine derivatives as new anti- Escherichia co
in vitro studies
Public ID is 891

Predicted property: **MIC** modeled in -log(M)
Training method: TRANSNN

Data Set	#	R2	q2	RMSE	MAE
o Training set: EColi_Set_IV (training)	3758 records	0.804 ± 0.009	0.803 ± 0.009	0.485 ± 0.01	0.326 ± 0.006
o Test set: EColi_Set_IV (test) [x]	1352 records	0.8 ± 0.02	0.8 ± 0.02	0.49 ± 0.02	0.32 ± 0.01

a)

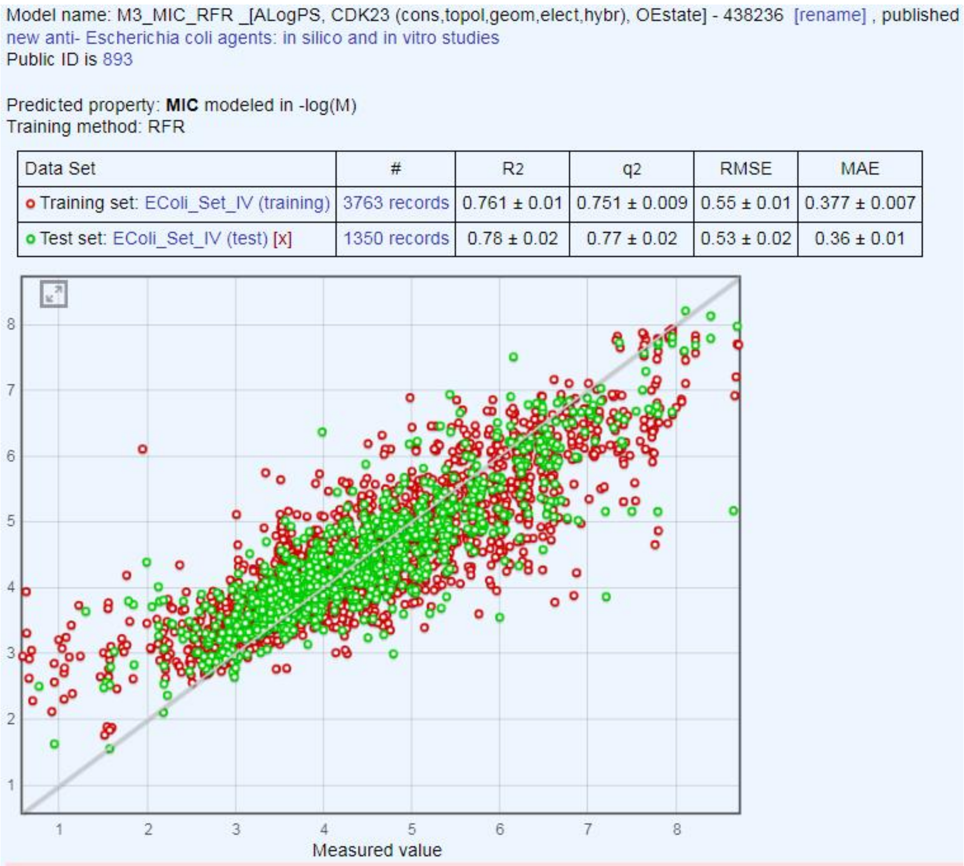


Model name: M2_MIC_ASNN _[ALogPS, CDK23 (cons,topol,geom,elect,hybr), OEstate] - 438223 [rename] , published
new anti- Escherichia coli agents: in silico and in vitro studies
Public ID is 892

Predicted property: **MIC** modeled in -log(M)
Training method: ASNN

Data Set	#	R2	q2	RMSE	MAE
o Training set: EColi_Set_IV (training)	3763 records	0.73 ± 0.01	0.72 ± 0.01	0.58 ± 0.01	0.392 ± 0.007
o Test set: EColi_Set_IV (test) [x]	1350 records	0.74 ± 0.02	0.74 ± 0.02	0.57 ± 0.03	0.39 ± 0.01

b)



Model name: M3_MIC_RFR _[ALogPS, CDK23 (cons,topol,geom,elect,hybr), OEstate] - 438236 [rename] , published
new anti- Escherichia coli agents: in silico and in vitro studies
Public ID is 893

Predicted property: **MIC** modeled in -log(M)
Training method: RFR

Data Set	#	R2	q2	RMSE	MAE
○ Training set: EColi_Set_IV (training)	3763 records	0.761 ± 0.01	0.751 ± 0.009	0.55 ± 0.01	0.377 ± 0.007
○ Test set: EColi_Set_IV (test) [x]	1350 records	0.78 ± 0.02	0.77 ± 0.02	0.53 ± 0.02	0.36 ± 0.01

c)

Model name: M4_Consensus MIC - 438239 [rename] , published in Cytisine derivatives as new anti- Escherichia
studies
Public ID is 894

Predicted property: **MIC** modeled in -log(M)
Training method: Consensus

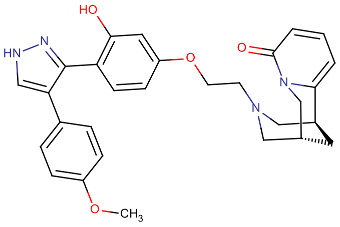
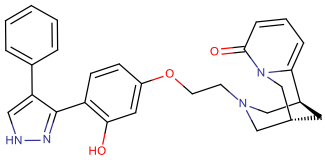
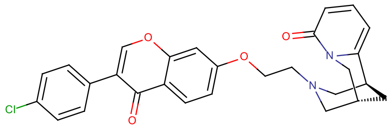
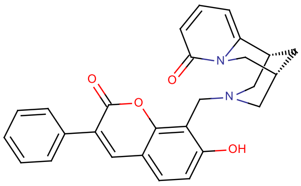
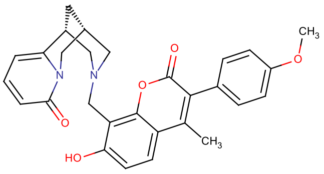
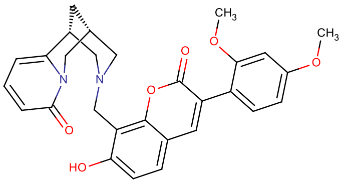
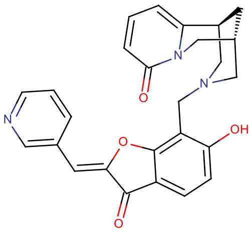
Data Set	#	R2	q2	RMSE	MAE
○ Training set: EColi_Set_IV (training)	3775 records	0.795 ± 0.01	0.792 ± 0.009	0.5 ± 0.01	0.341 ± 0.006
○ Test set: EColi_Set_IV (test) [x]	1360 records	0.8 ± 0.02	0.79 ± 0.01	0.5 ± 0.02	0.33 ± 0.01

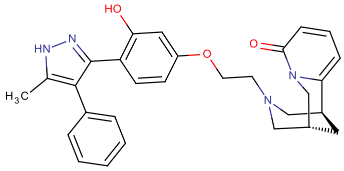
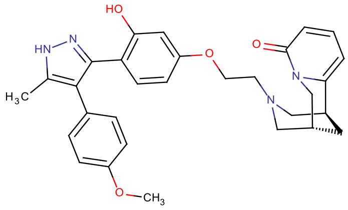
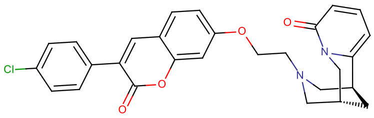
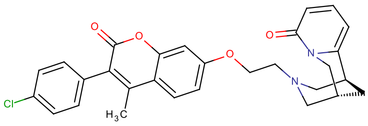
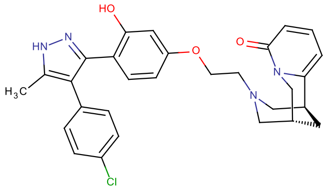
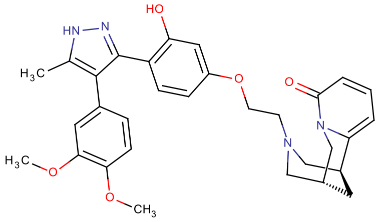
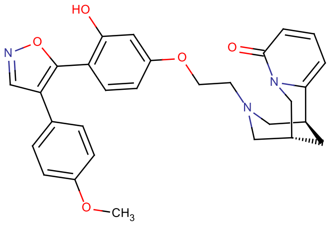
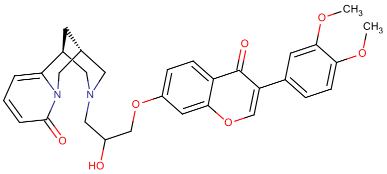
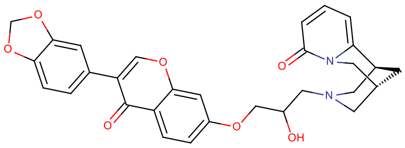
d)

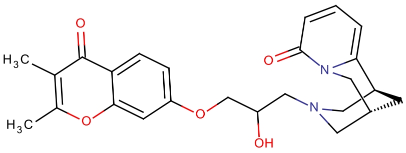
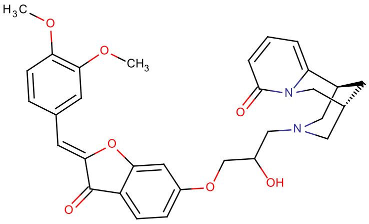
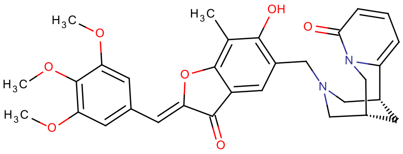
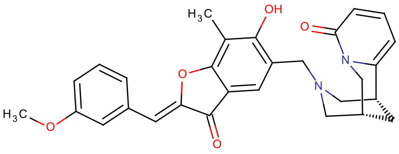
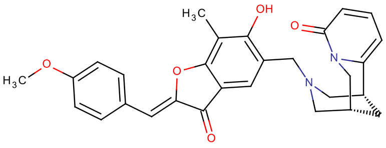
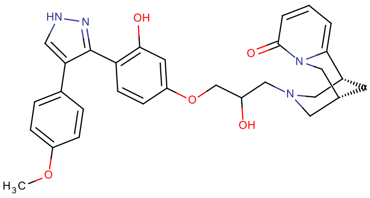
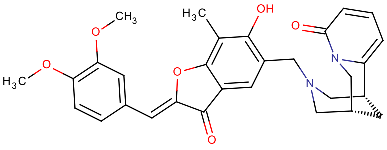
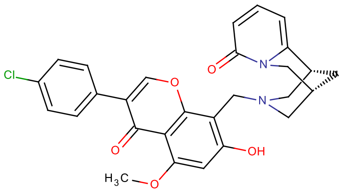
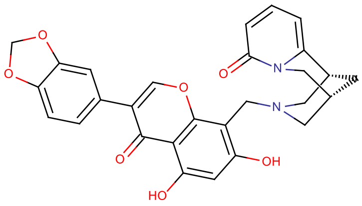
Fig. S1. Regression machine learning models built by the OCHEM server (http://ochem.eu). a-c) Statistical coefficients calculated for regression models by a different MLT; d) Consensus model calculated by averaging the previous three models.

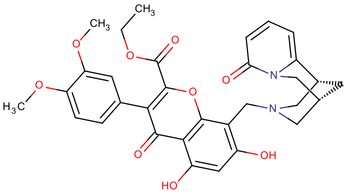
Evaluation activity of new compounds

Table 2S. Anti-*Escherichia coli* activity calculated by using the consensus regression model for 26 virtual compounds

Comp. No	**Chemical Structure**	**log(1/MIC), M**	**CONSENSUS STD**	**Estimated RMSE**	**AD**
1		5.65	0.99	1.34	FALSE
2		5.51	1.1	1.34	FALSE
3		4.59	0.54	1.34	FALSE
4[b]		4.64	0.28	0.61	TRUE
5[b]		4.57	0.25	0.61	TRUE
6[b]		4.61	0.2	0.41	TRUE
7[a]		**5.05**	**0.43**	**0.74**	**TRUE**

8		5.46	1.12	1.34	FALSE
9		5.54	1.11	1.34	FALSE
10		4.39	0.53	1.34	FALSE
11		4.51	0.55	1.34	FALSE
12		5.52	1.01	1.34	FALSE
13[a]		**5.55**	**1.09**	**1.34**	**TRUE**
14[a]		**4.53**	**0.39**	**0.61**	**TRUE**
15[a]		**4.60**	**0.24**	**0.61**	**TRUE**
16		4.67	0.51	1.34	FALSE

17[a]		**4.34**	**0.24**	**0.61**	**TRUE**
18[a]		**4.88**	**0.16**	**0.41**	**TRUE**
19		5.11	0.55	1.34	FALSE
20		5.22	0.66	1.34	FALSE
21		5.25	0.73	1.34	FALSE
22		5.69	1	1.34	FALSE
23		5.07	0.54	1.34	FALSE
24[a]		**4.91**	**0.21**	**0.41**	**TRUE**
25		5.02	0.64	1.34	FALSE

26[b]		4.61	0.32	0.61	TRUE

[a]Final set compounds are represented in bold. [b]Compound predicted as high toxic; MIC - minimal inhibitory concentration; CONSENSUS-STD - the standard deviation of the predictions, obtained from an ensemble of models; RMSE is the root mean square error; AD –applicability domain.