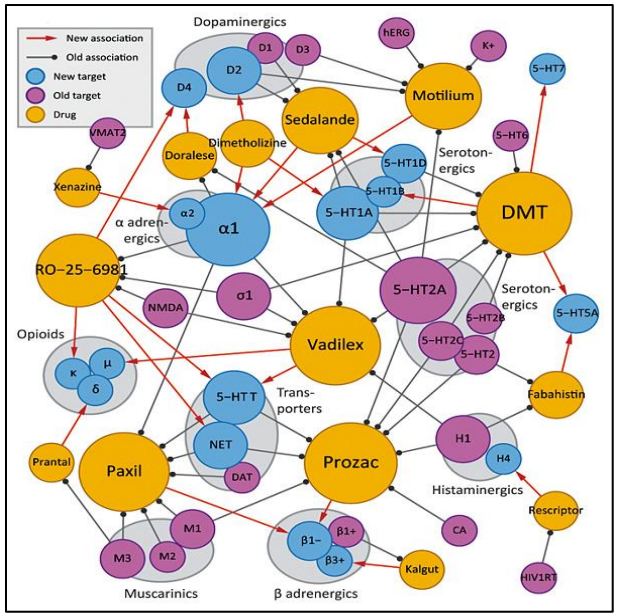
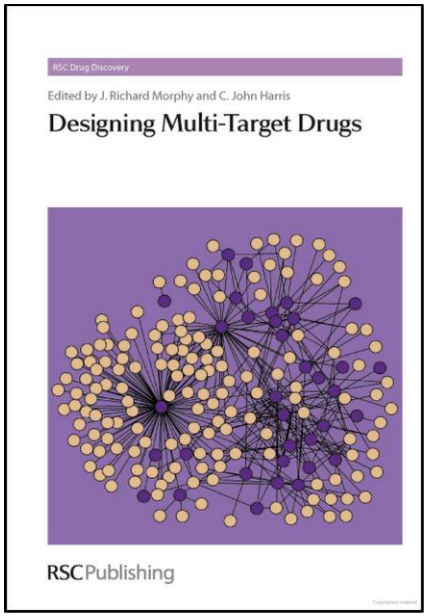
Volgograd State Medical University
Vorozhtsov Institute of Organic Chemistry

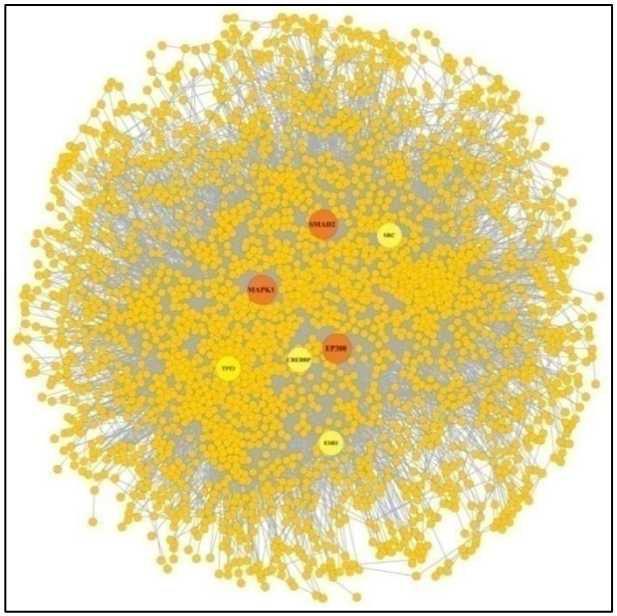
THE CONSENSUS ENSEMBLE MULTIDESCRIPTOR MULTITARGET NEURAL NETWORK MODELING OF PHARMACOLOGICAL ACTIVITY OF CHEMICAL COMPOUNDS

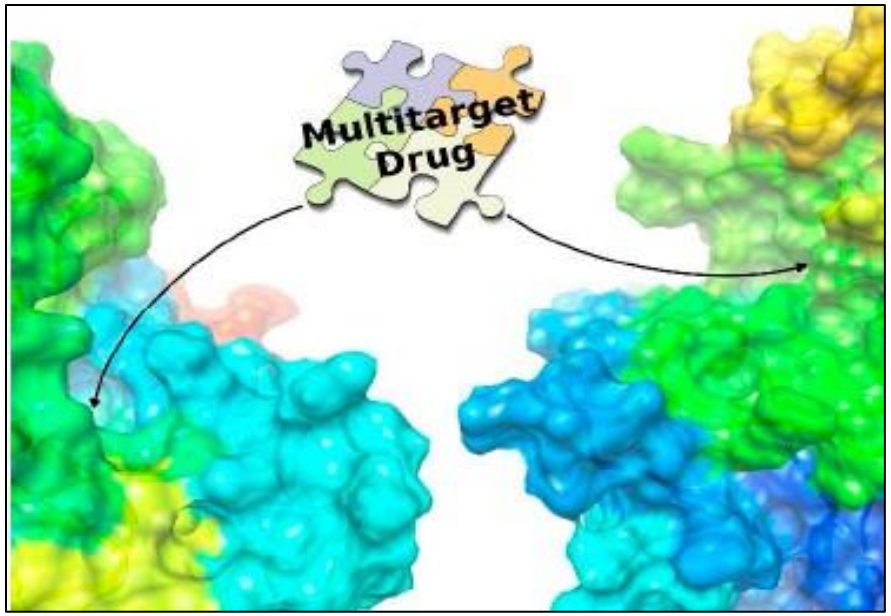
Vassiliev Pavel Mikhailovich

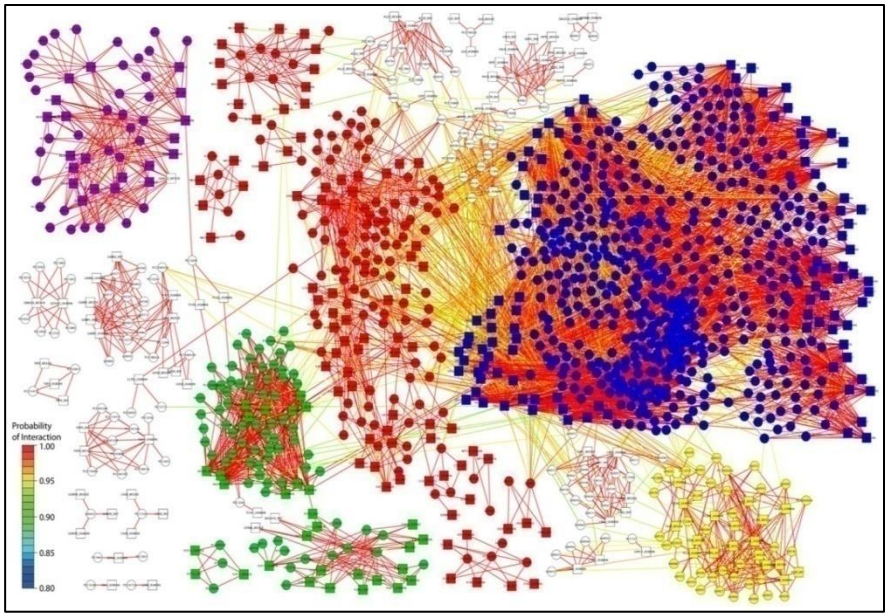
All drugs are multitarget



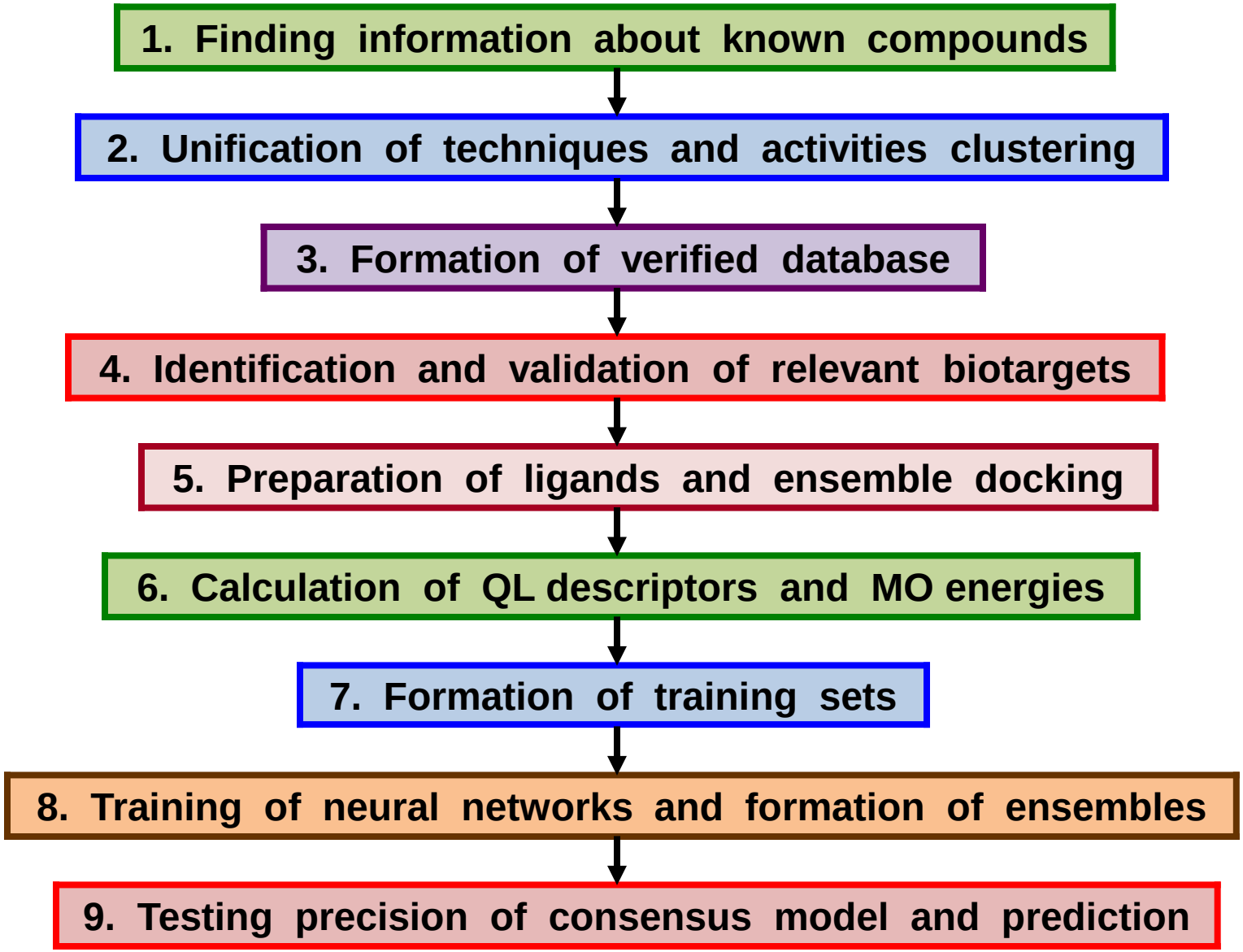




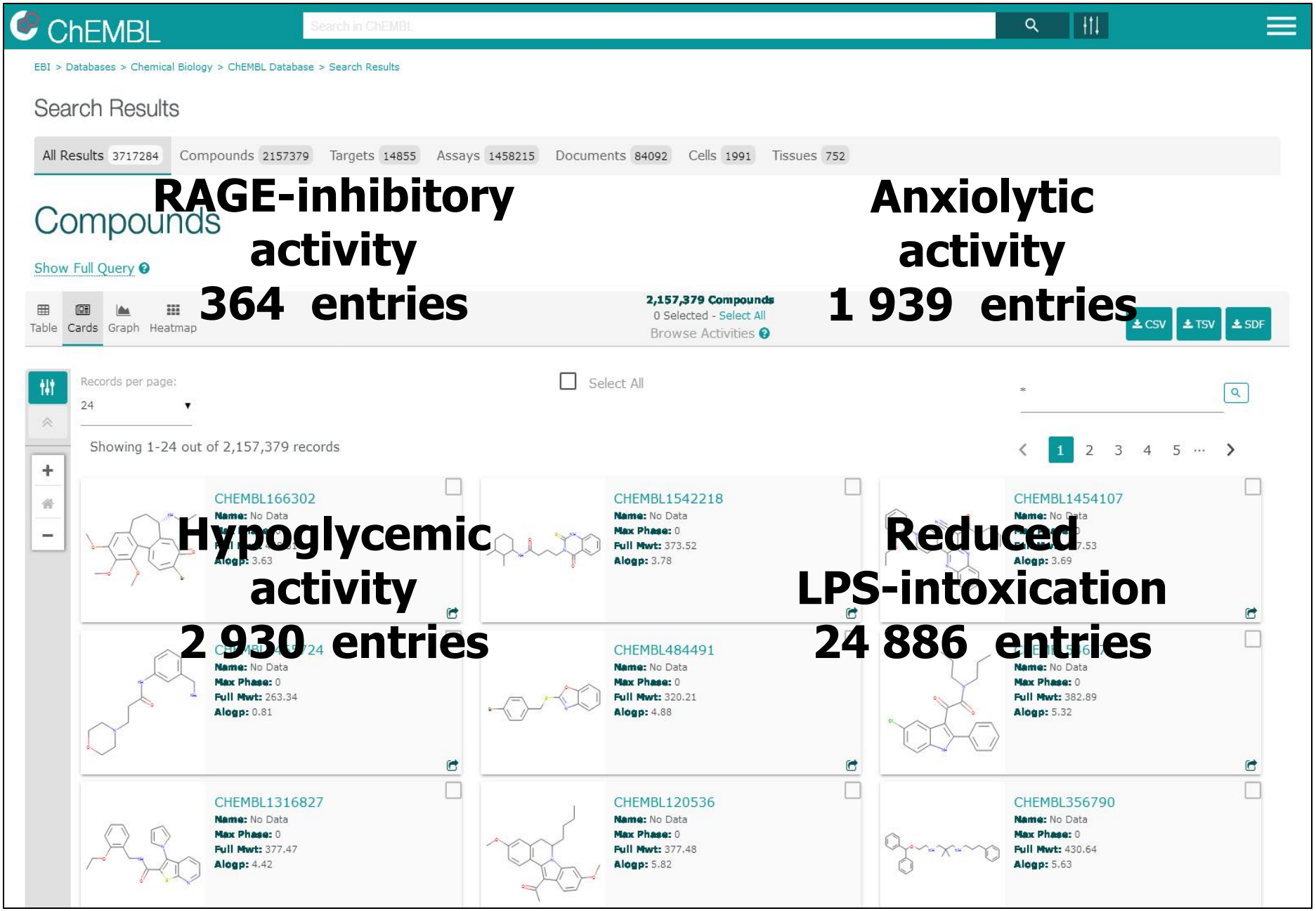




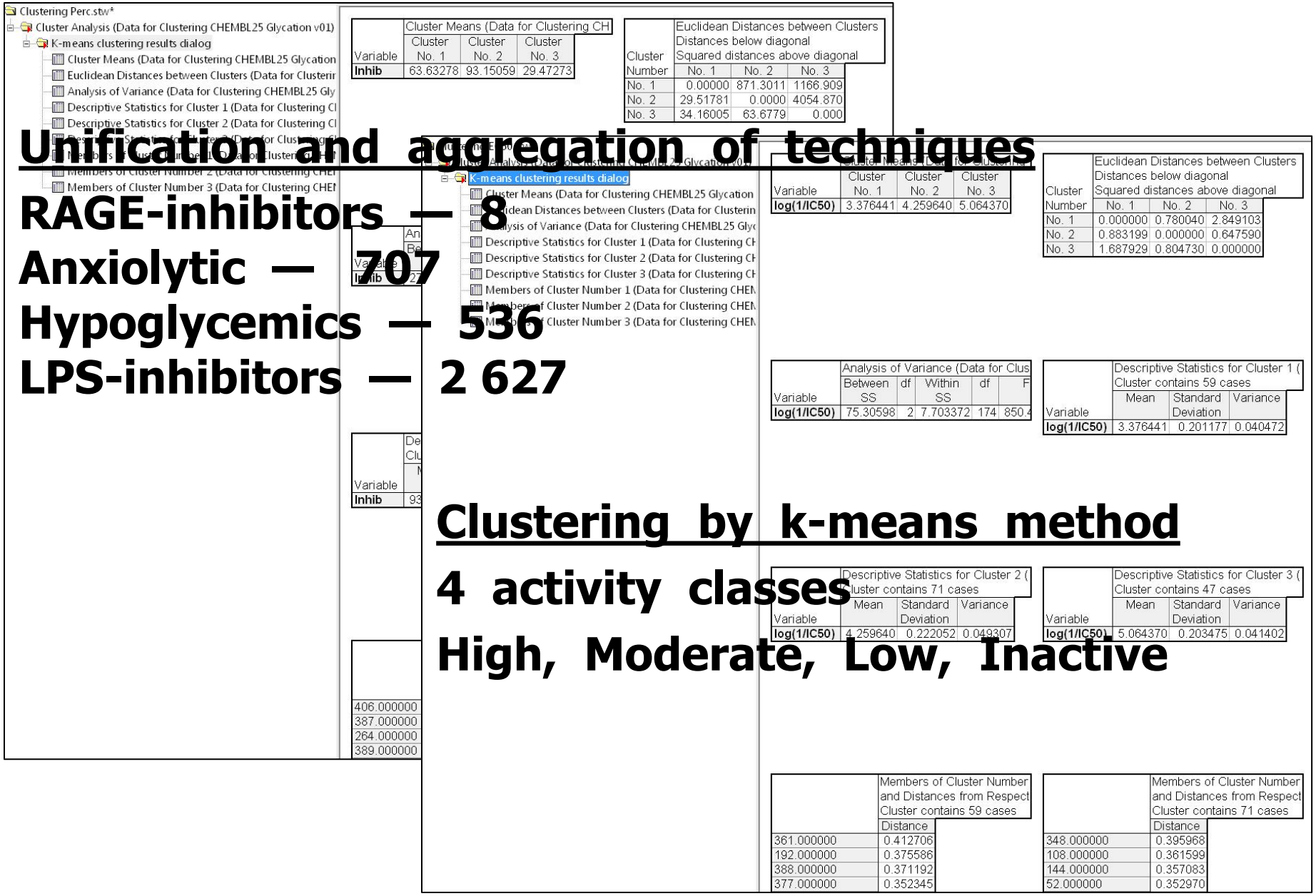
General model building workflow



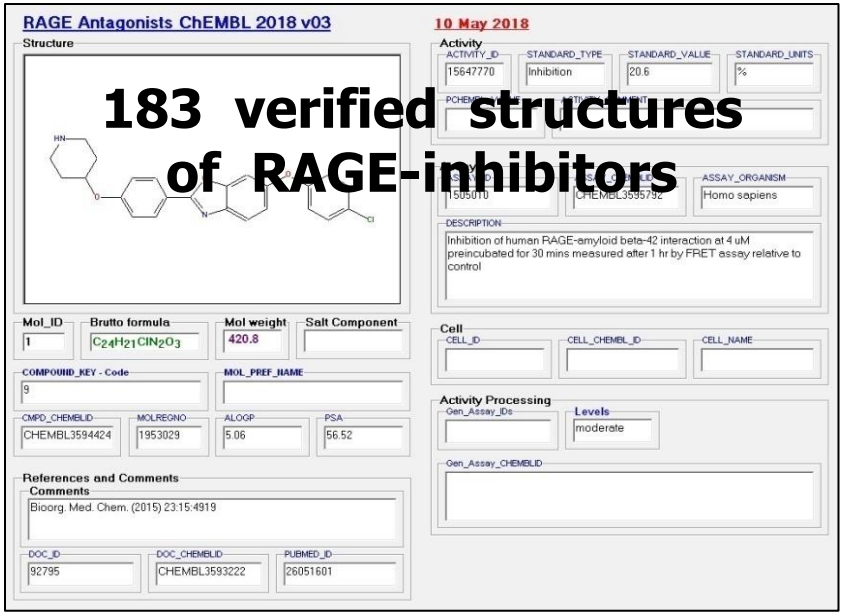
Finding information about known compounds



Unification of techniques and activities clustering

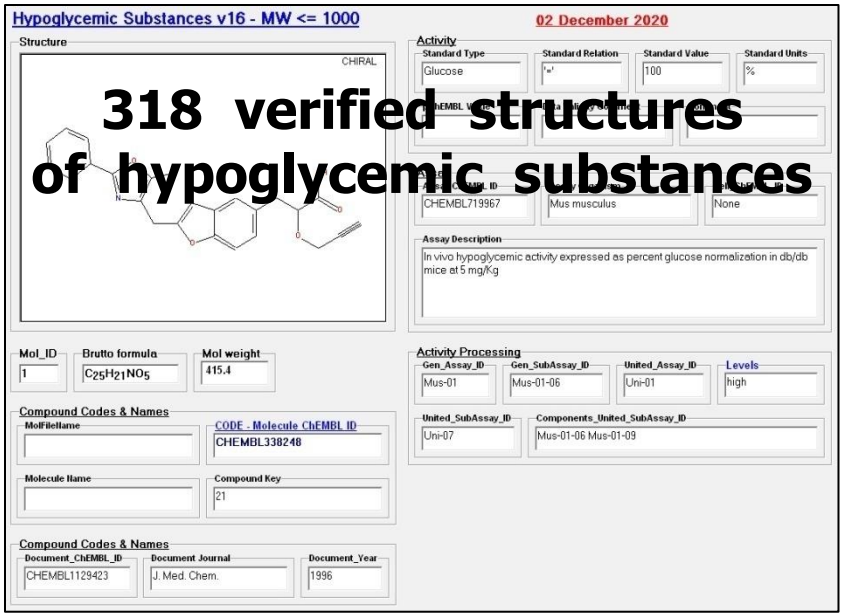


Formation of verified databases

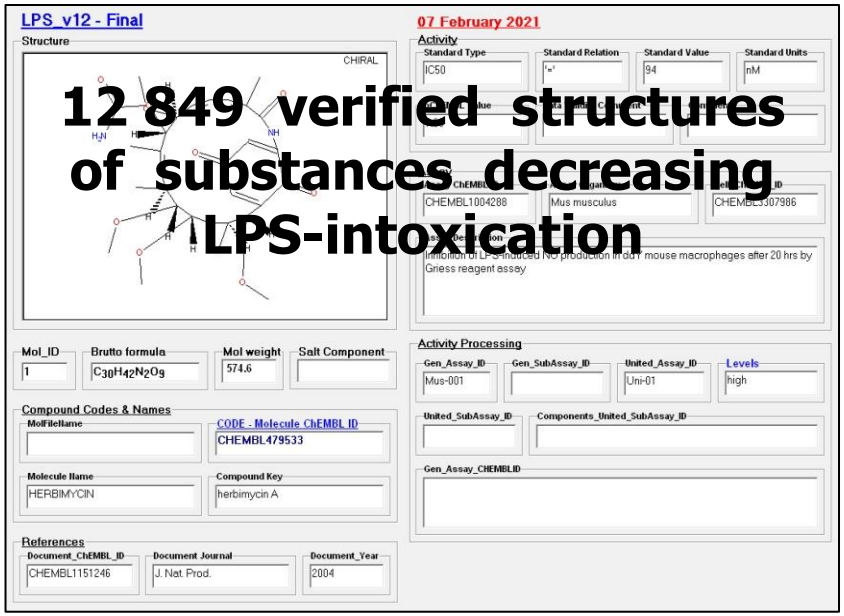


183 verified structures of RAGE-inhibitors

216 verified structures of anxiolytic substances

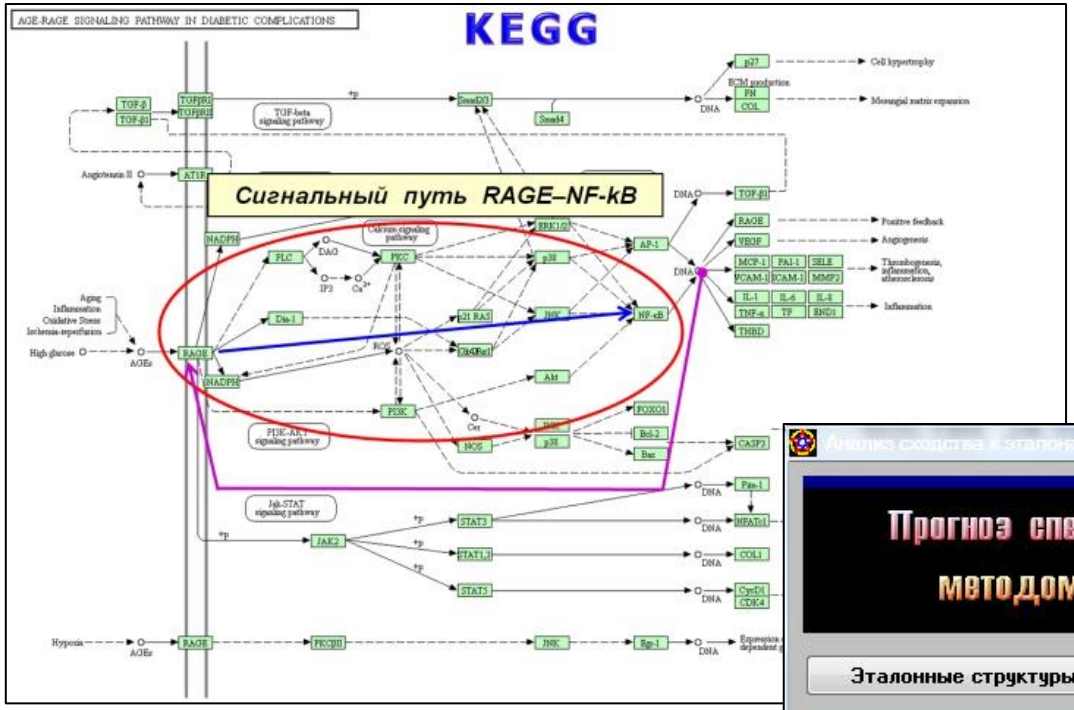


318 verified structures of hypoglycemic substances



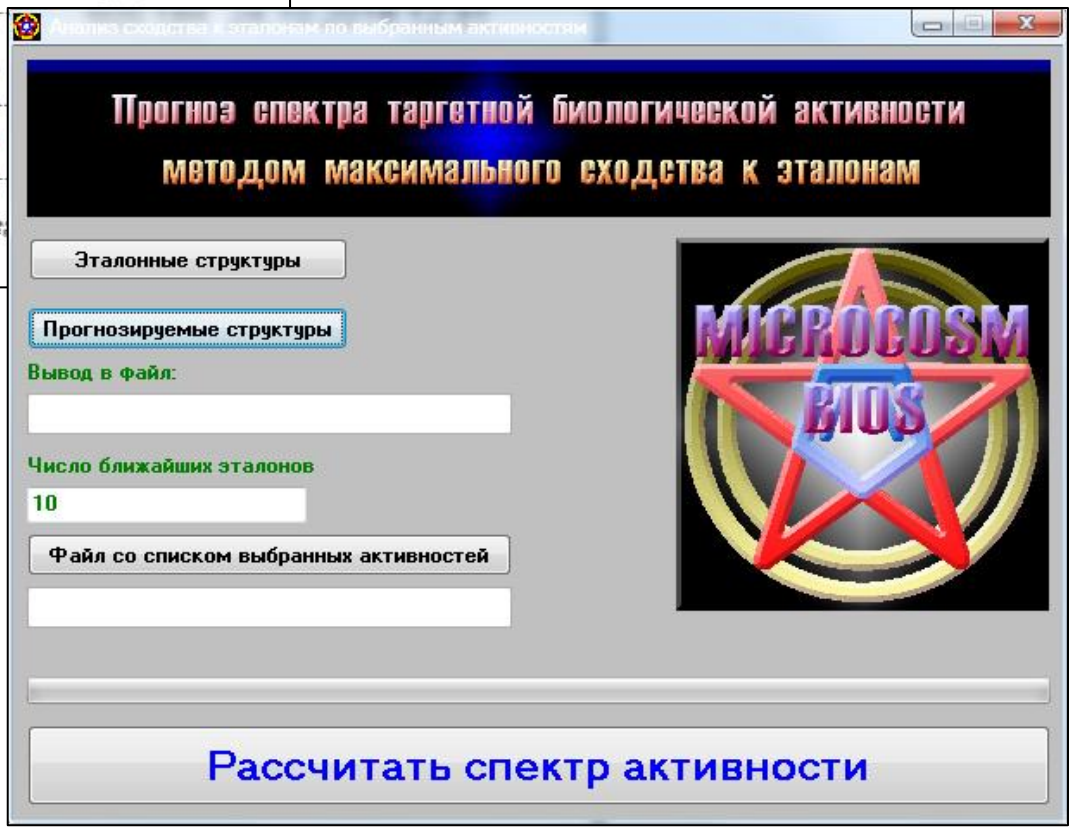
12 849 verified structures of substances decreasing LPS-intoxication

Identification of relevant biotargets



Signaling pathway analysis

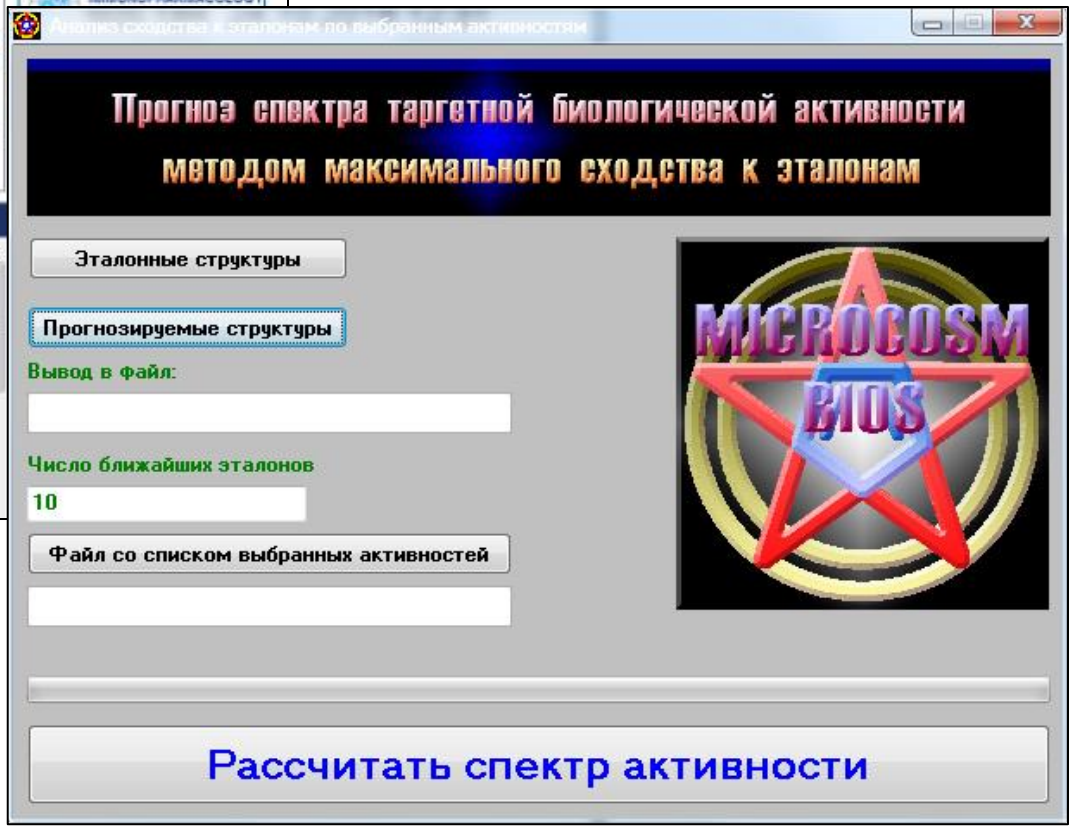
Context prediction of target activity using chemotypes



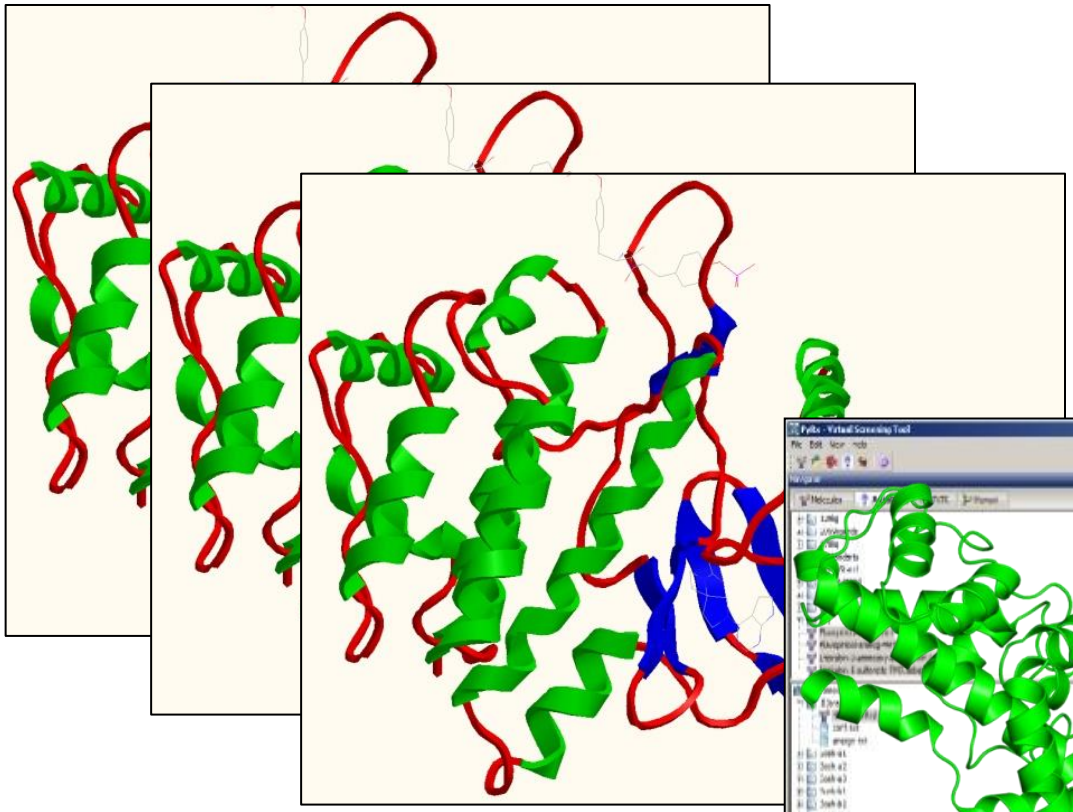
Searching standard compounds

Selection of known active reference compounds

Selection of structure similarity inactive compounds

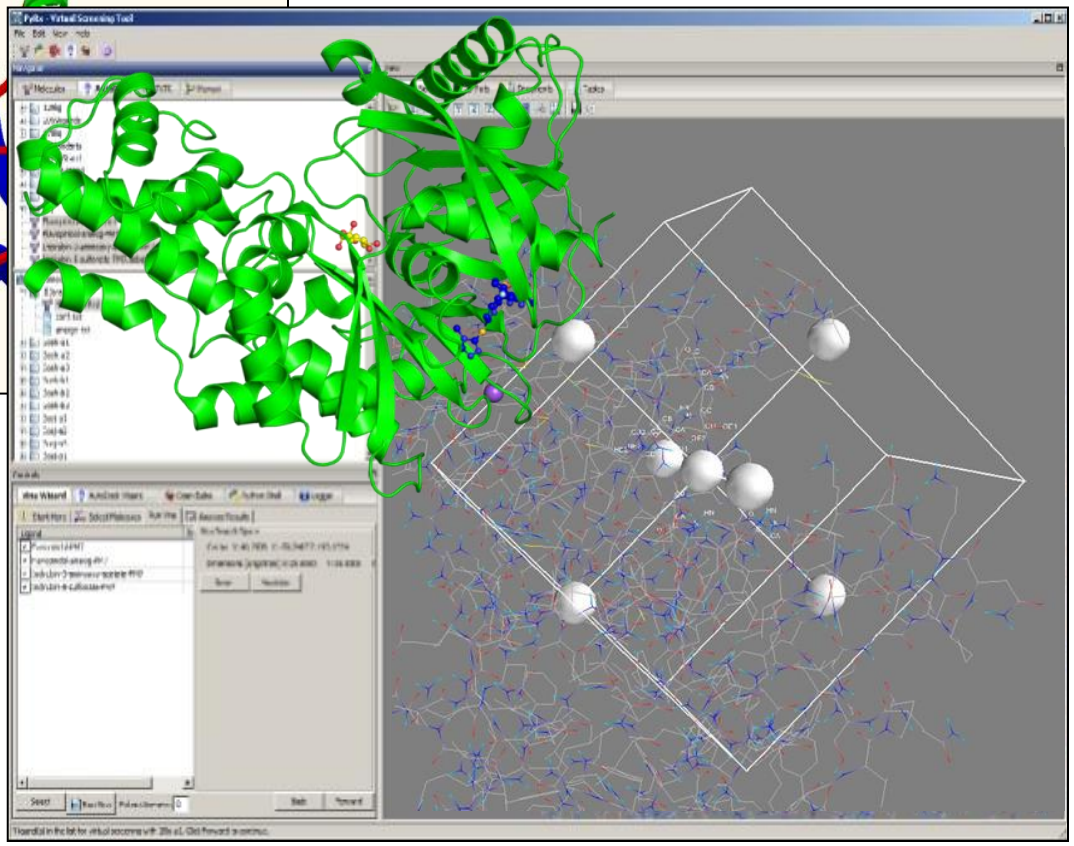


Validation of 3D-models of biotargets



Formation of initial sets of 3D-models

Docking of reference substances and selection of validated 3D-models



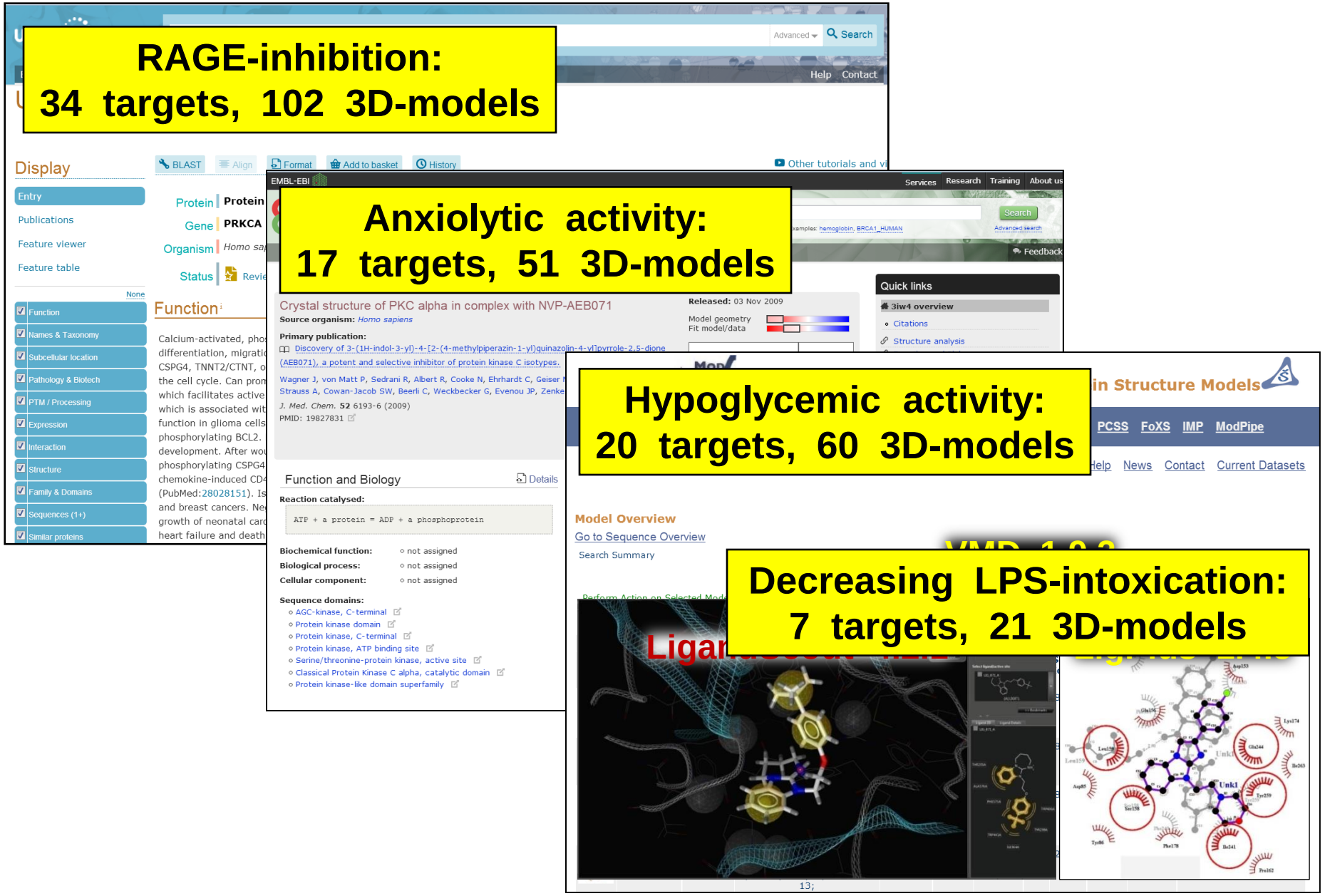
Validated 3D-models of relevant biotargets

RAGE-inhibition:
34 targets, 102 3D-models

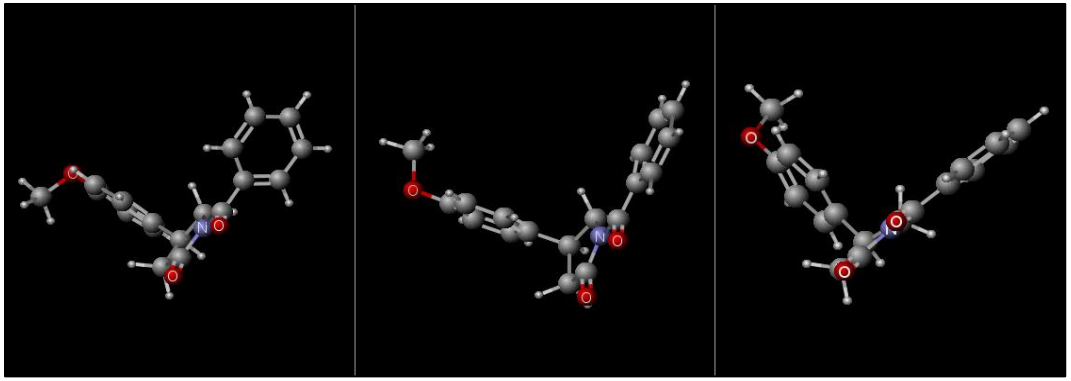
Anxiolytic activity:
17 targets, 51 3D-models

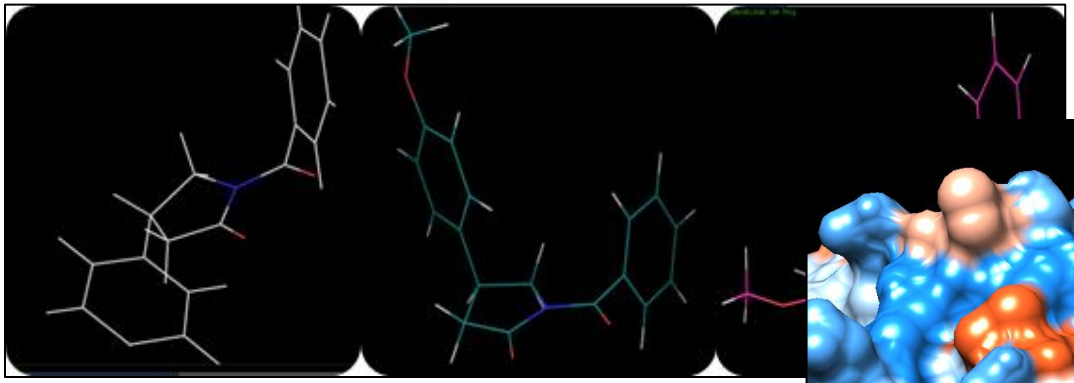
Hypoglycemic activity:
20 targets, 60 3D-models

Decreasing LPS-intoxication:
7 targets, 21 3D-models



Preparation of ligands and ensemble docking

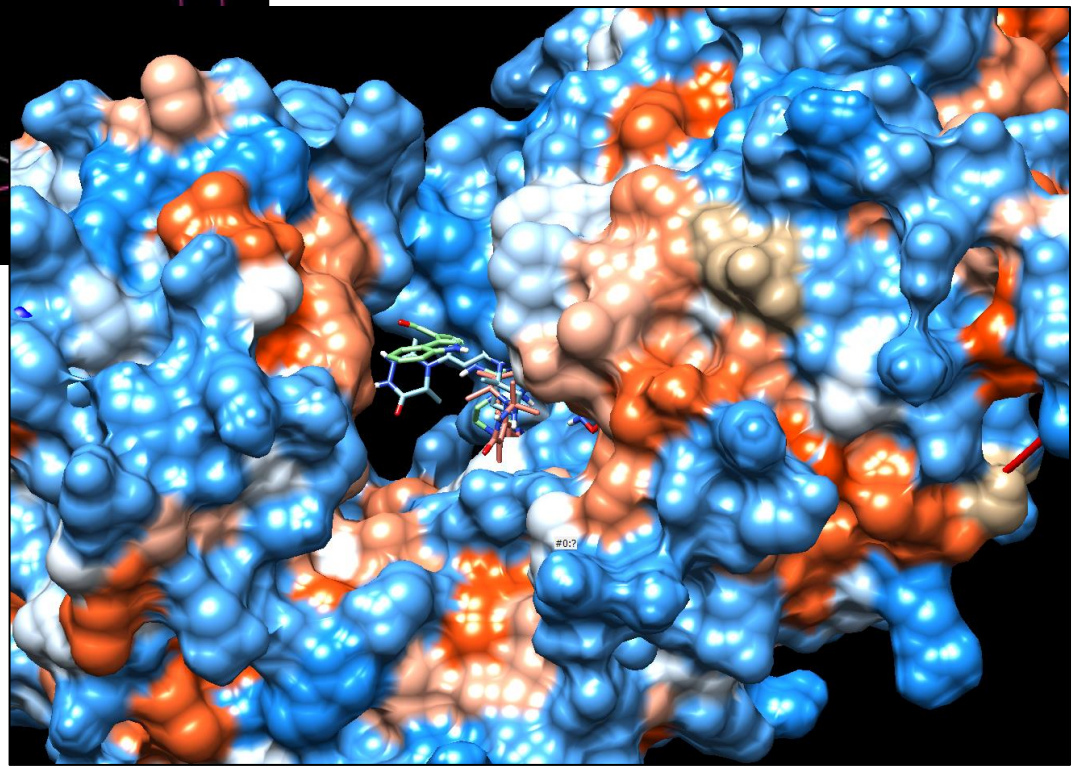




Each ligand in each 3D-model 5 times in 10 conformations

$$\Delta E_{ij} = \min(\Delta E_{ijkl})$$

~138 000 000 docking energies

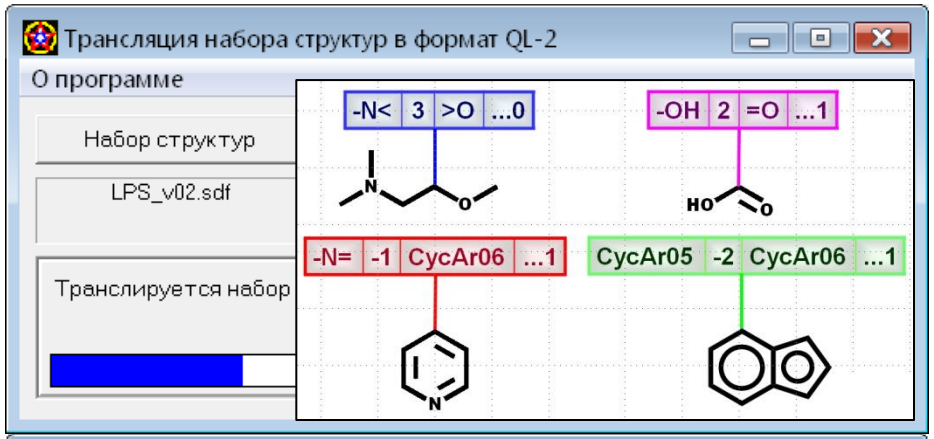


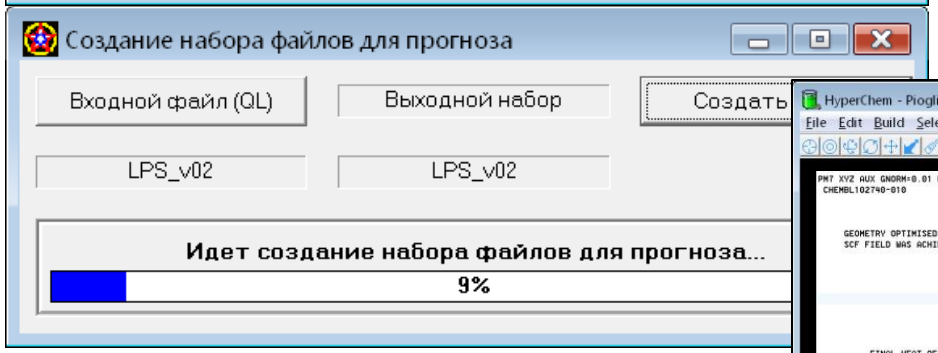
<u>Molecular mechanics</u>
10 conformers

<u>Quantum chemistry PM7</u>
Optimization and selection of the best conformation

~1 300 000 conformers

Calculation of QL descriptors and MO energies

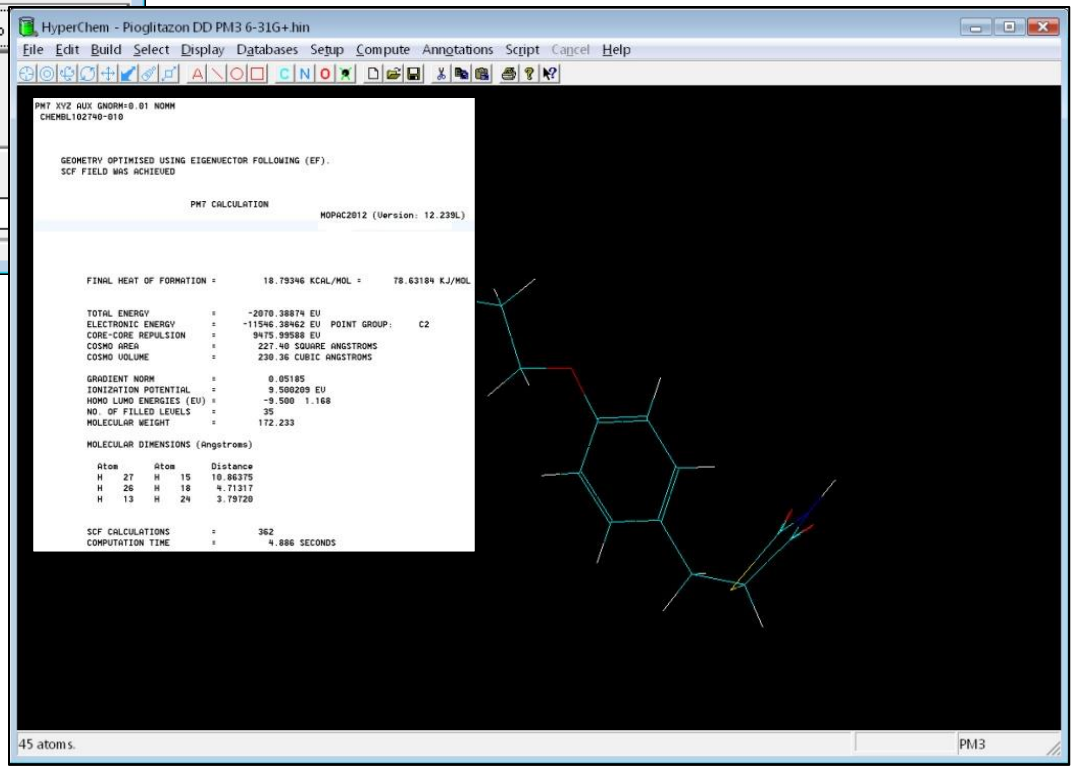




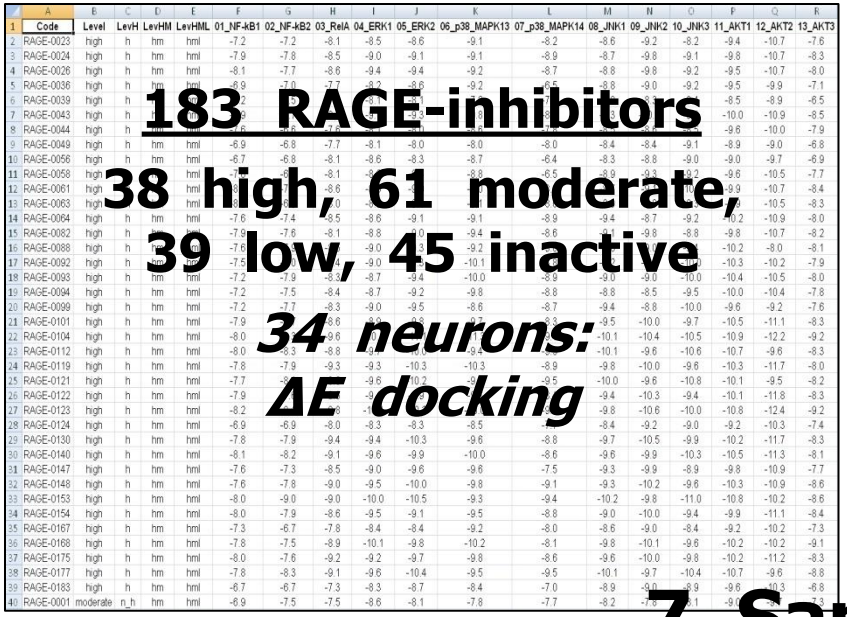
IT Microcosm
QL-descriptors
1st rank

Visual Basic script

10 E_{HOMO} energies
10 E_{LUMO} energies



Formation of training sets



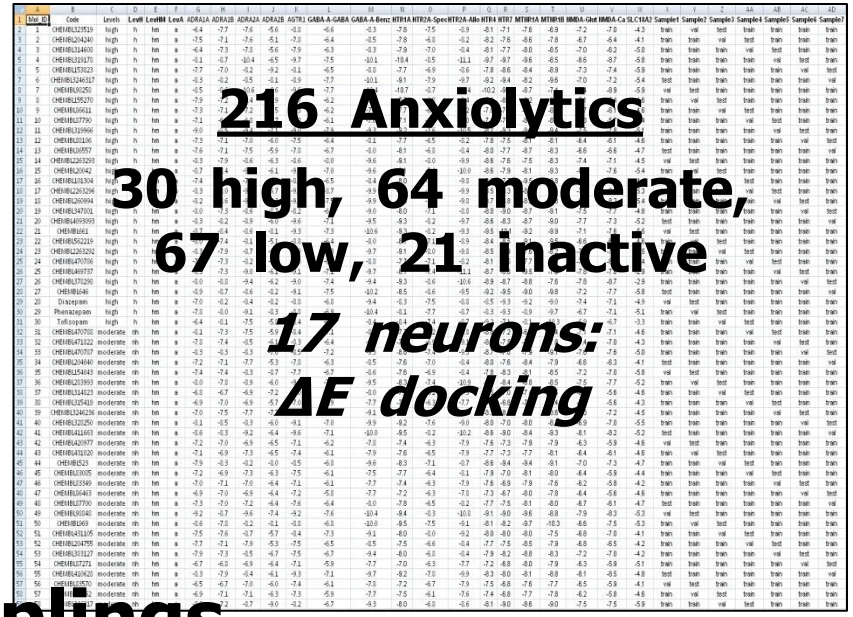
183 RAGE-inhibitors

38 high, 61 moderate,

39 low, 45 inactive

34 neurons:

ΔE docking



216 Anxiolytics

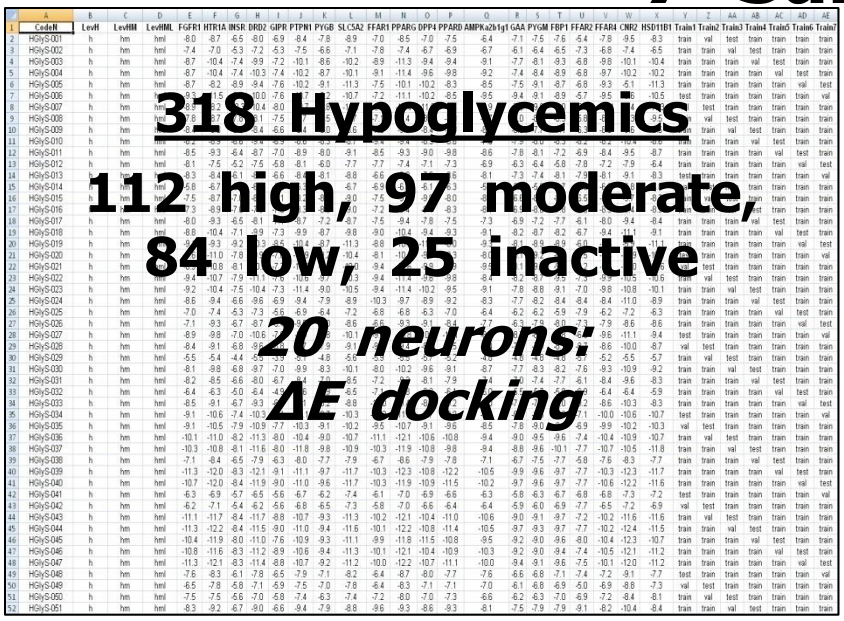
30 high, 64 moderate,

67 low, 21 inactive

17 neurons:

ΔE docking

7 Samplings



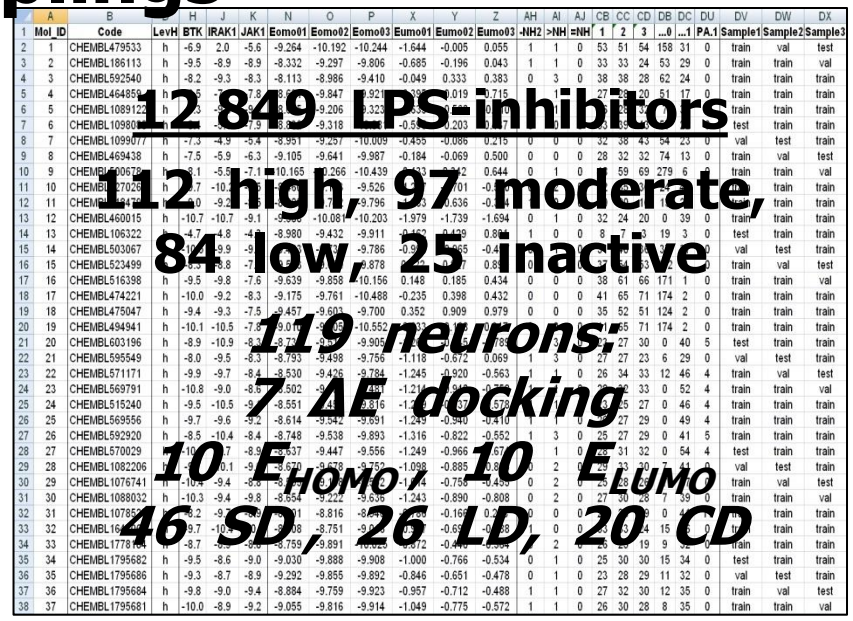
318 Hypoglycemics

112 high, 97 moderate,

84 low, 25 inactive

20 neurons:

ΔE docking



12 849 LPS-inhibitors

112 high, 97 moderate,

84 low, 25 inactive

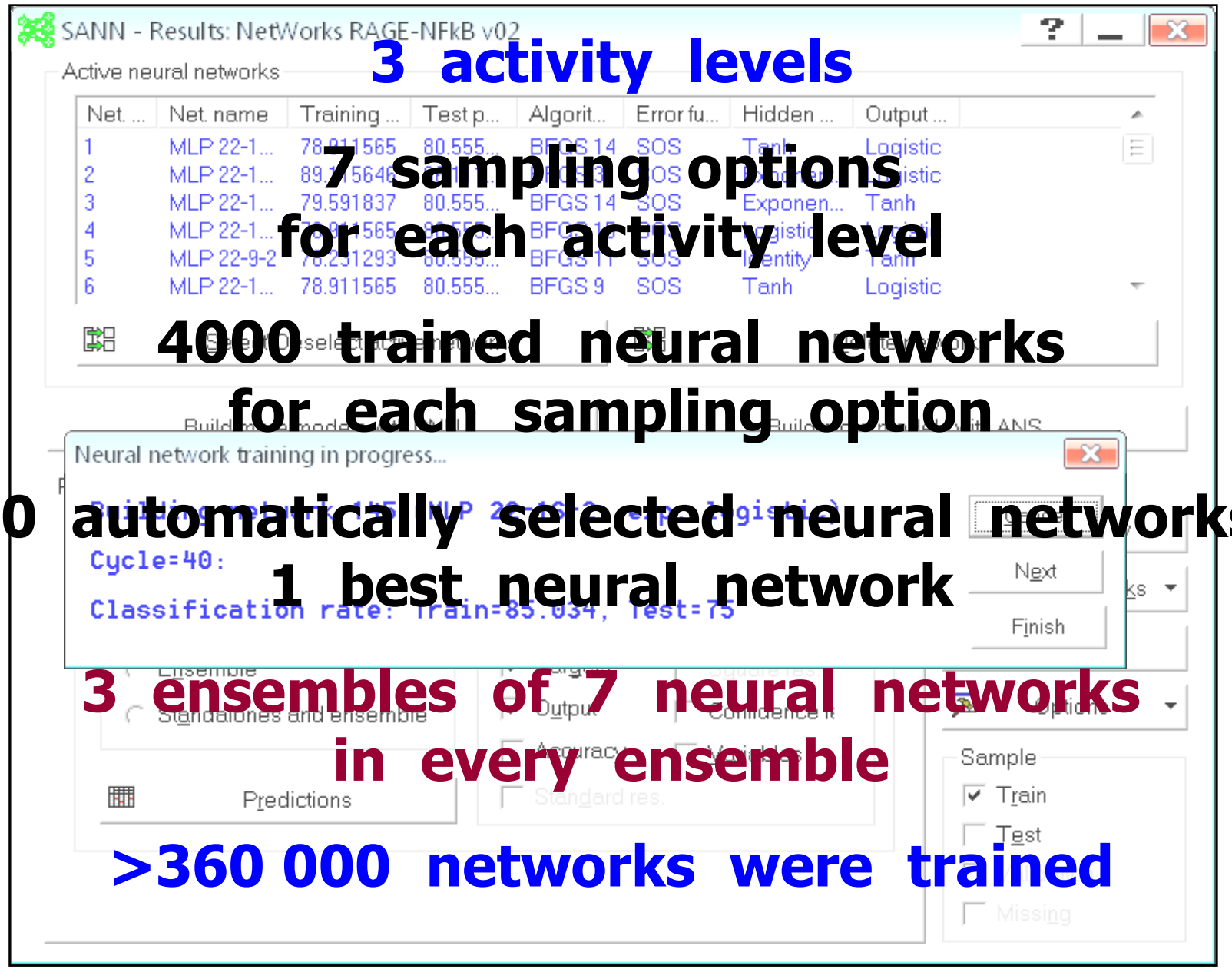
119 neurons:

7 ΔE docking

10 E_{HOMO} / 10 E_{LUMO}

46 SD, 26 LD, 20 CD

Training of neural networks and formation of ensembles



3 activity levels

7 sampling options for each activity level

4000 trained neural networks for each sampling option

50 automatically selected neural networks

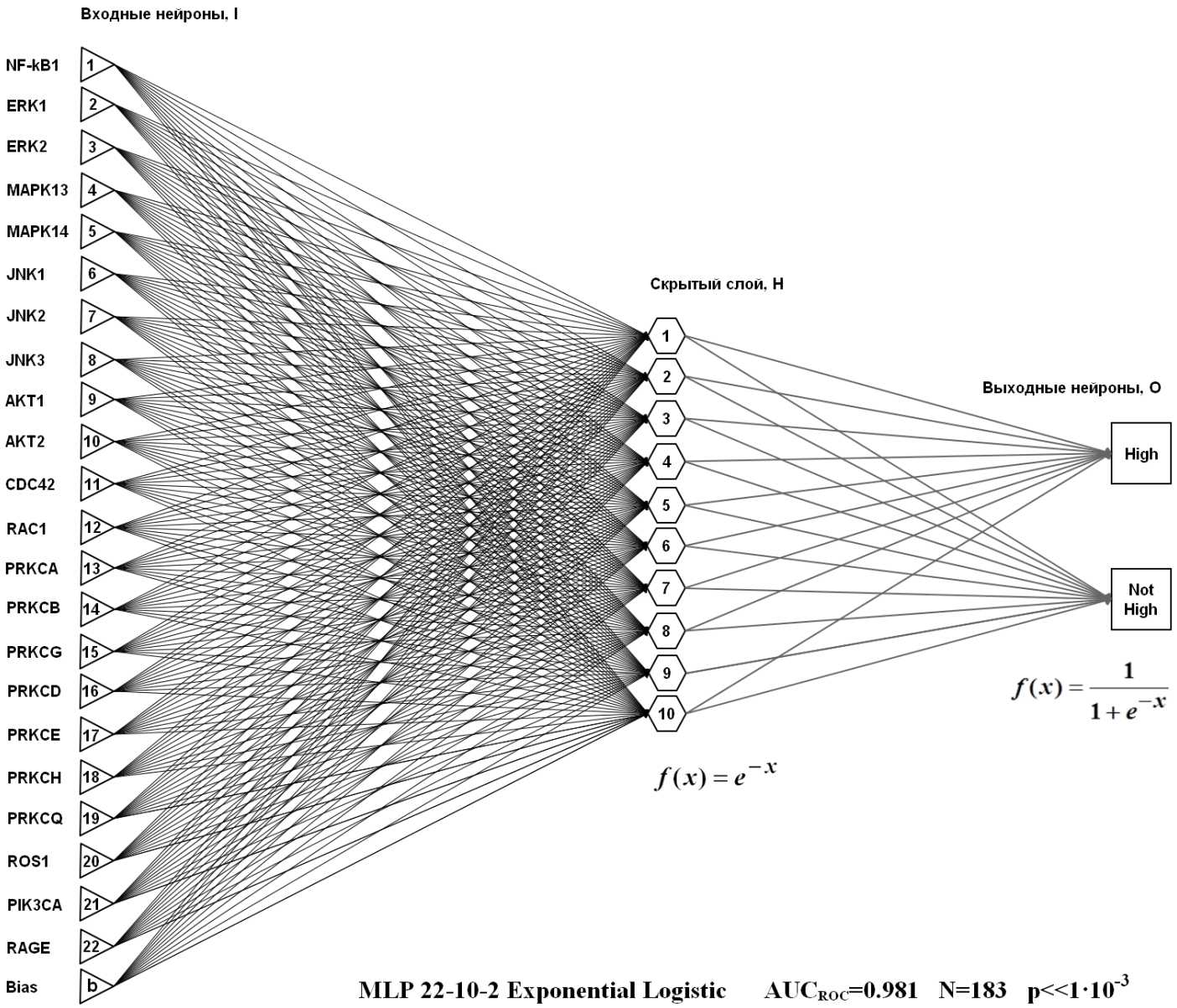
1 best neural network

3 ensembles of 7 neural networks in every ensemble

>360 000 networks were trained

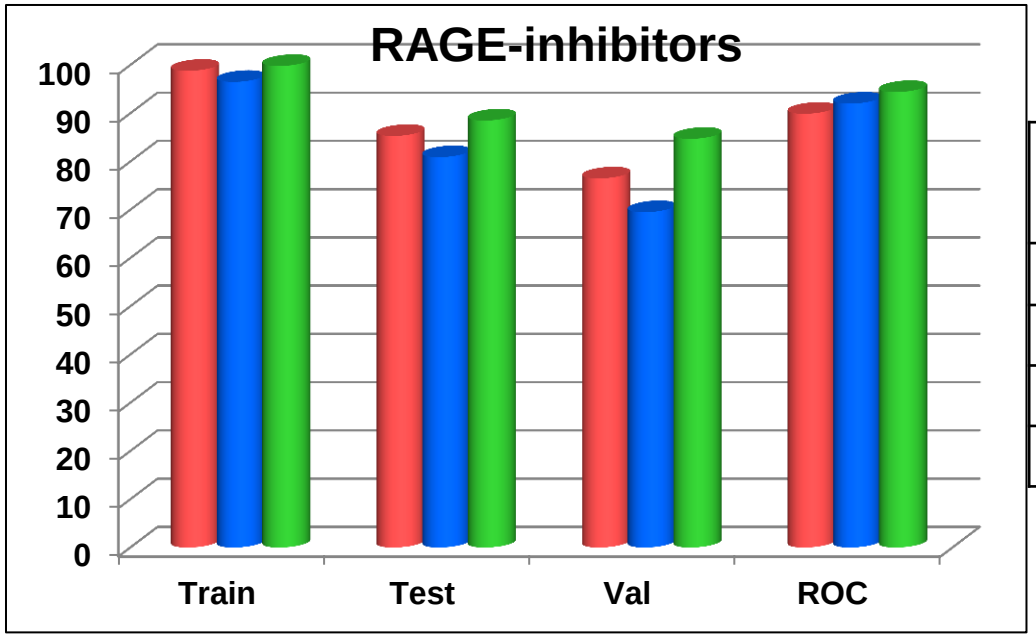
An example of a neural network architecture

RAGE-inhibition from docking energies for 34 biotargets



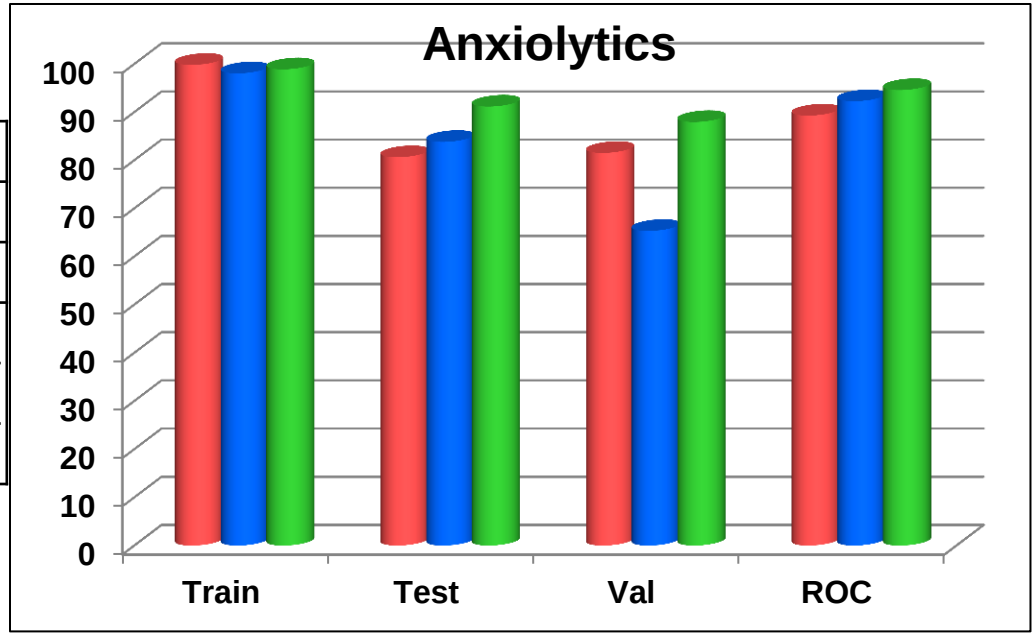
MLP 22-10-2 Exponential Logistic AUC_{ROC}=0.981 N=183 $p<<1\cdot10^{-3}$

Accuracy of neural network ensembles

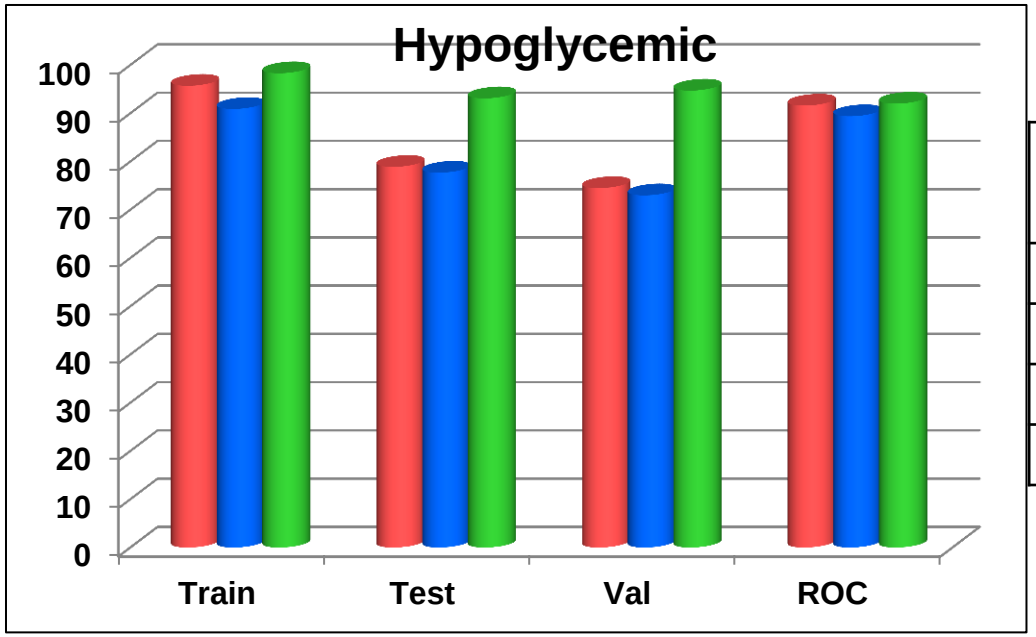


Activity level	Accuracy, %			
	Train	Test	Val	ROC
RAGE-inhibitory activity				
High	98.8	85.3	76.5	89.9
High or Moderate	96.5	80.9	69.5	92.0
Active	99.8	88.5	84.7	94.4

Activity level	Accuracy, %			
	Train	Test	Val	ROC
Anxiolytic activity				
High	99.8	80.6	81.5	89.3
High or Moderate	98.1	83.8	65.3	92.3
Active	98.9	91.2	87.9	94.6

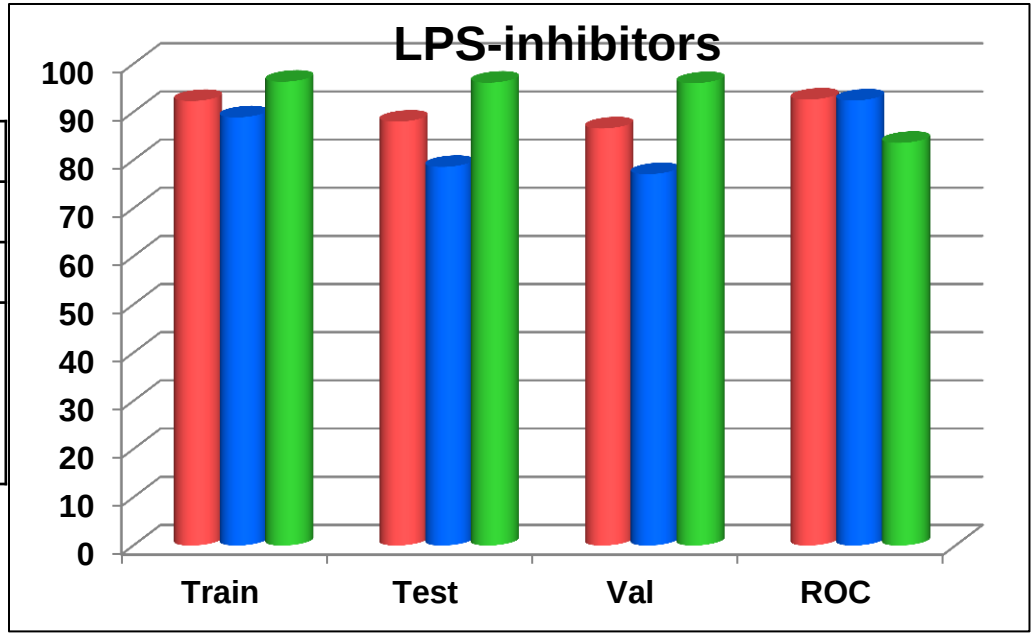


Accuracy of neural network ensembles



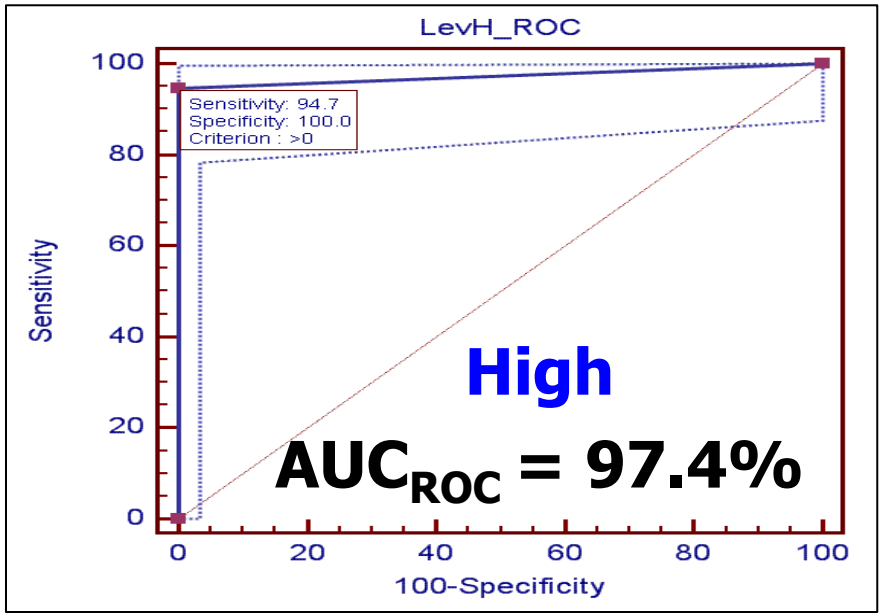
Activity level	Accuracy, %			
	Train	Test	Val	ROC
Hypoglycemic activity				
High	95.7	78.9	74.5	91.6
High or Moderate	90.9	77.7	73.0	89.4
Active	98.3	93.1	94.7	92.0

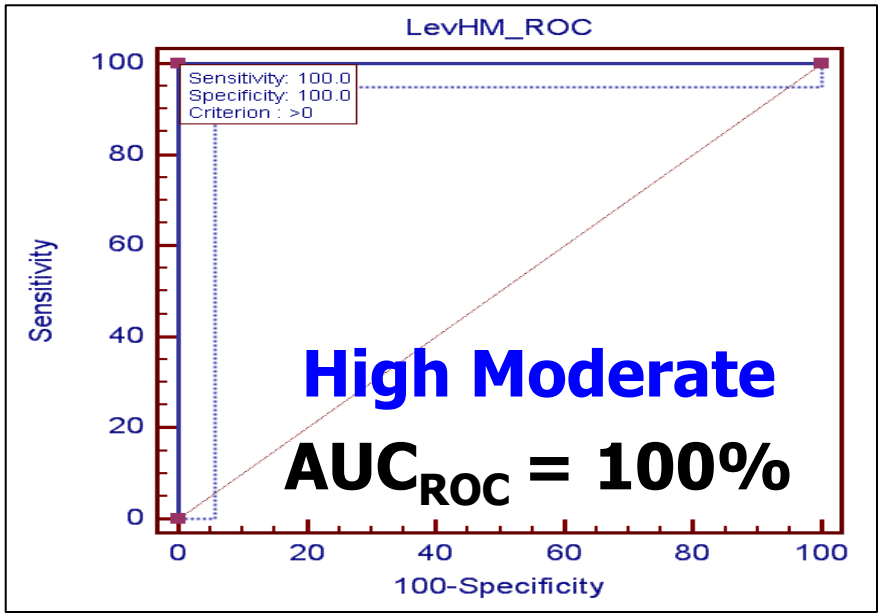
Activity level	Accuracy, %			
	Train	Test	Val	ROC
Decreased LPS-intoxication				
High	92.3	88.0	86.6	92.6
High or Moderate	88.9	78.6	77.1	92.4
Active	96.3	96.1	96.0	83.6

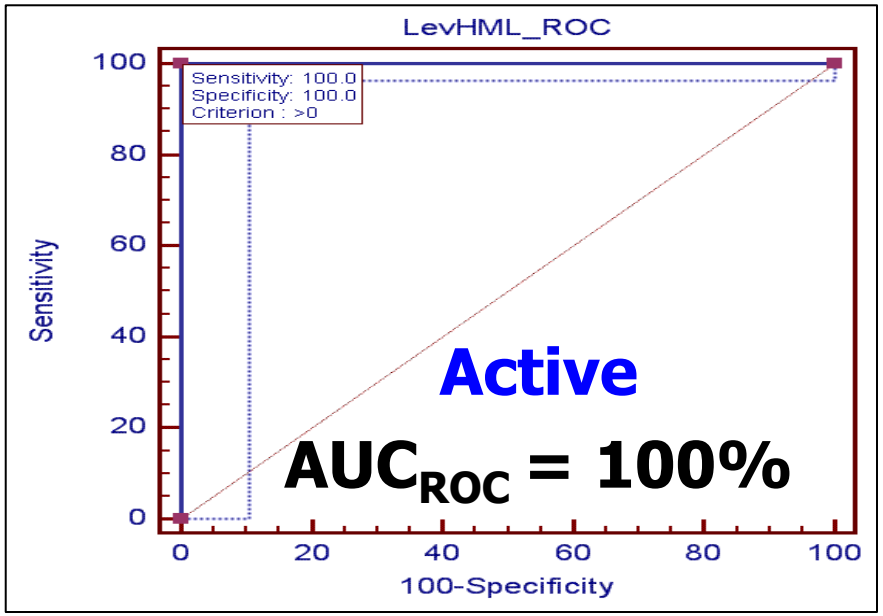


ROC analysis of a consensus neural network model

RAGE-inhibitors







Consensus prediction of RAGE-inhibitory activity

First level consensus — simple vote

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Code	Sampl1	Sampl2	Sampl3	Sampl4	Sampl5	Sampl6	Sampl7	Ind1	Ind2	Ind3	Ind4	Ind5	Ind6	Ind7	Cons>=4	Cons>=5	Cons>=6	Cons=7
2	10	n_h	n_h	n_h	h	n_h	n_h	n_h				1							
3	11	n_h	n_h	h	h	h	h	h			1	1	1	1	1	h	h		
4	14080	n_h	h	n_h	n_h	n_h	n_h	h		1					1				
5	14086	n_h	h	n_h	n_h	n_h	n_h	n_h		1									
6	14087	n_h	n_h	n_h	n_h	n_h	n_h	n_h											
7	14091	n_h	n_h	n_h	n_h	n_h	n_h	h							1				
8	14093	h	n_h	n_h	n_h	n_h	n_h	h	1						1				
9	14102	n_h	h	n_h	n_h	n_h	n_h	n_h		1									
10	14120	n_h	n_h	n_h	n_h	h	n_h	n_h					1						
11	14138	n_h	n_h	n_h	n_h	n_h	n_h	n_h											
12	14144	n_h	n_h	n_h	n_h	n_h	n_h	n_h											
13	14160	h	n_h	n_h	n_h	n_h	n_h	h	1						1				
14	14	n_h	n_h	n_h	h	h	n_h	h				1	1		1				
15	16	n_h	n_h	n_h	n_h	n_h	n_h	n_h											
16	17	n_h	n_h	n_h	n_h	n_h	n_h	n_h											
17	20	n_h	n_h	h	n_h	n_h	n_h	n_h			1								
18	23	h	n_h	n_h	h	n_h	h	n_h	1			1		1					
19	26	n_h	n_h	h	h	h	n_h	n_h			1	1	1						
20	2	n_h	n_h	n_h	n_h	n_h	n_h	n_h											
21	35	n_h	n_h	h	h	h	n_h	n_h			1	1	1						
22	37	n_h	n_h	n_h	h	n_h	n_h	n_h				1							
23	38	n_h	n_h	n_h	n_h	n_h	h	h						1	1				
24	41	n_h	n_h	n_h	n_h	n_h	n_h	n_h											
25	44	n_h	n_h	n_h	n_h	n_h	n_h	h							1				
26	47	n_h	n_h	n_h	h	h	n_h	n_h				1	1						
27	55	h	n_h	n_h	n_h	n_h	n_h	n_h	1										
28	56	n_h	n_h	n_h	h	n_h	h	h				1		1	1				
29	5	n_h	n_h	h	h	h	h	n_h			1	1	1	1		h			
30	7	n_h	n_h	n_h	n_h	n_h	n_h	h							1				
31	8	h	n_h	n_h	n_h	n_h	n_h	n_h	1										
32	AB-0016	n_h	n_h	n_h	n_h	n_h	n_h	n_h											
33	AB-0019	n_h	n_h	n_h	n_h	n_h	n_h	n_h											
34	AB-0020	n_h	n_h	n_h	n_h	n_h	n_h	n_h											
35	AB-0045	n_h	n_h	n_h	n_h	n_h	n_h	n_h											
36	AB-0054	n_h	n_h	n_h	n_h	n_h	n_h	n_h											
37	AB-0067	n_h	n_h	n_h	n_h	n_h	n_h	h							1				
38	AZH-0119	n_h	n_h	n_h	n_h	n_h	n_h	n_h											
39	AZH-0143	n_h	h	h	h	h	h	h		1	1	1	1	1	1	h	h	h	
40	BIF-0016	h	n_h	h	h	h	h	h	1		1	1	1	1	1	h	h	h	
41	BIF-0023	h	n_h	h	h	h	h	h	1		1	1	1	1	1	h	h	h	
42	BIF-0064	n_h	n_h	n_h	n_h	n_h	n_h	n_h											

Consensus prediction of RAGE-inhibitory activity

Second level consensus — noncontradiction check

Code	High				HighMod				HighModLow				High Full Consensus				Moderate Full Consensus			
	Cons>=4	Cons>=5	Cons>=6	Cons=7	Cons>=4	Cons>=5	Cons>=6	Cons=7	Cons>=4	Cons>=5	Cons>=6	Cons=7	Cons>=4	Cons>=5	Cons>=6	Cons=7	Cons>=4	Cons>=5	Cons>=6	Cons=7
14080					hm	hm			hml	hml	hml	hml					1	1		
14091					hm				hml	hml	hml	hml					1			
14102					hm	hm	hm	hm	hml	hml	hml	hml					1	1	1	1
23					hm	hm			hml	hml	hml	hml					1	1		
26					hm				hml	hml	hml						1			
35					hm	hm			hml	hml	hml						1	1		
38					hm				hml								1			
5	**h**				**hm**				**hml**	**hml**	**hml**		**1**							
AZH-0143	**h**	**h**	**h**		**hm**				**hml**	**hml**	**hml**		**1**							
K-215					hm	hm	hm	hm	hml	hml							1	1		
SUM-0018					hm	hm	hm		hml	hml	hml						1	1	1	
SUM-0026					hm				hml								1			
TONS-0462	**h**				**hm**	**hm**			**hml**	**hml**			**1**					**1**		
Osha01					hm	hm			hml	hml							1	1		
Osha02					hm	hm	hm		hml	hml	hml						1	1	1	
Osha03					hm	hm	hm		hml	hml	hml						1	1	1	
Osha04					hm	hm	hm		hml	hml	hml						1	1	1	
Osha07	**h**	**h**	**h**		**hm**	**hm**	**hm**		**hml**	**hml**	**hml**	**hml**	**1**	**1**	**1**					
Osha10					hm	hm	hm	hm	hml								1			
Osha11	**h**	**h**	**h**		**hm**	**hm**	**hm**	**hm**	**hml**	**hml**	**hml**	**hml**	**1**	**1**	**1**					**1**
Osha12					hm				hml	hml							1			
Osha21					hm	hm	hm	hm	hml	hml	hml						1	1	1	
Osha22	**h**				**hm**				**hml**	**hml**			**1**							
Osha26					hm				hml	hml							1			
Oshi09	**h**				**hm**	**hm**			**hml**	**hml**	**hml**		**1**					**1**		
Oshi10	**h**	**h**	**h**		**hm**	**hm**	**hm**	**hm**	**hml**	**hml**	**hml**		**1**	**1**	**1**					
Oshi11	**h**				**hm**	**hm**			**hml**	**hml**	**hml**		**1**					**1**		
Oshi12					hm	hm			hml	hml	hml	hml					1	1		
Oshi21					hm	hm			hml	hml	hml						1	1		
Oshi37					hm	hm	hm		hml	hml	hml	hml					1	1	1	
1sha30	**h**	**h**			**hm**	**hm**			**hml**	**hml**	**hml**		**1**	**1**						
1sha35					hm	hm			hml								1			
1sha36					hm				hml	hml	hml						1			
1sha44					hm	hm	hm		hml	hml	hml	hml					1	1	1	
1sha45					hm	hm			hml	hml							1	1		
1sha46					hm	hm			hml	hml	hml						1	1		
1sha53	**h**	**h**	**h**	**h**	**hm**	**hm**	**hm**		**hml**	**hml**	**hml**	**hml**	**1**	**1**	**1**					
1shadr01	**h**				**hm**	**hm**			**hml**	**hml**	**hml**	**hml**	**1**					**1**		

high = h & hm & a
moderate = nh & hm & a
low = nh & nhm & a
inactive = nh & nhm & na

incorrect = h & nhm & a
incorrect = h & nhm & na
incorrect = nh & hm & na
... et cetera

Creative team

P.M. Vassiliev

A.A. Spasov

D.A. Babkov

D.V. Maltsev

A.N. Kochetkov

M.A. Perfilev

A.R. Koroleva

A.V. Golubeva

O.A. Luzina

Acknowledgments

The research was supported by the following grants:

RFBR № 18-015-00499

RFBR № 20-015-00164

RSF № 20-013-00029

Ministry of Education and Science of the Russian Federation

№ 075-15-20-777

Thank You for Your attention!