

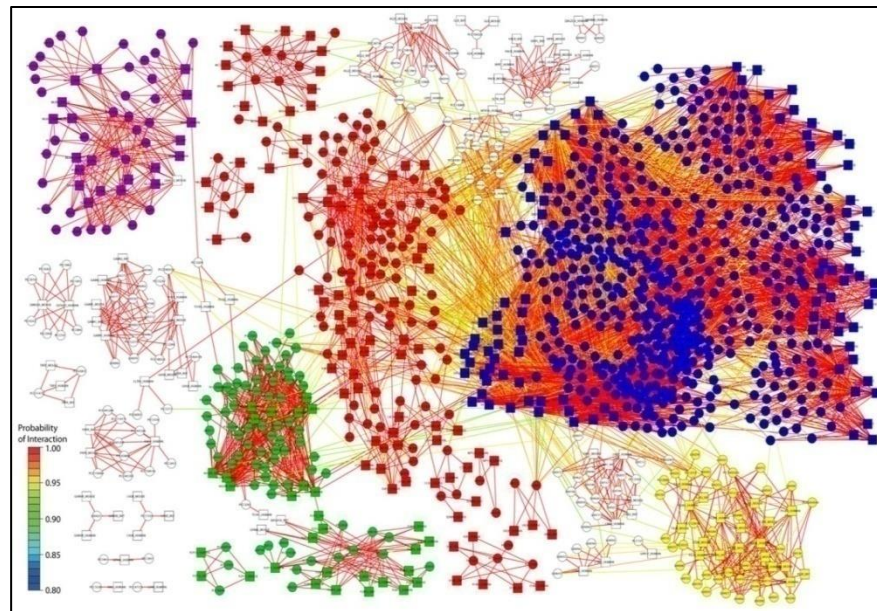
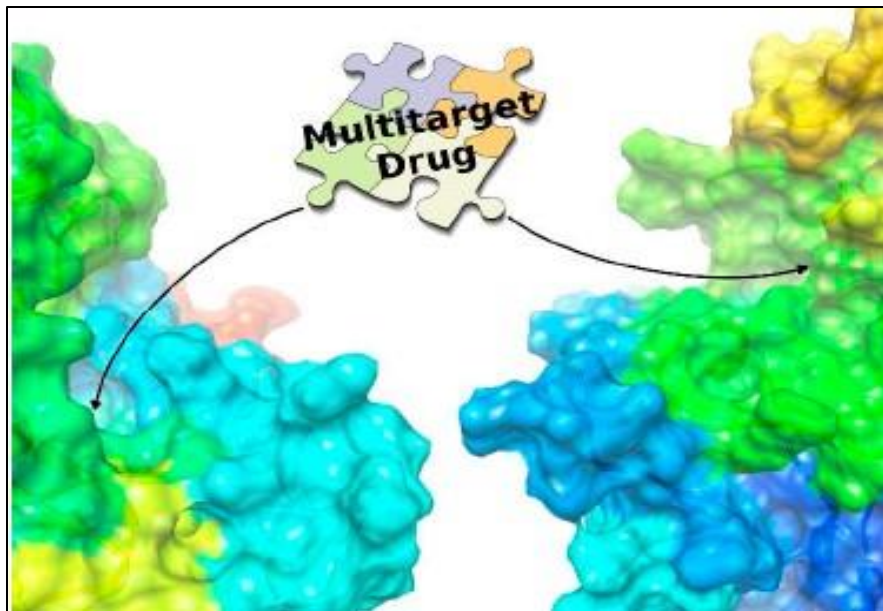
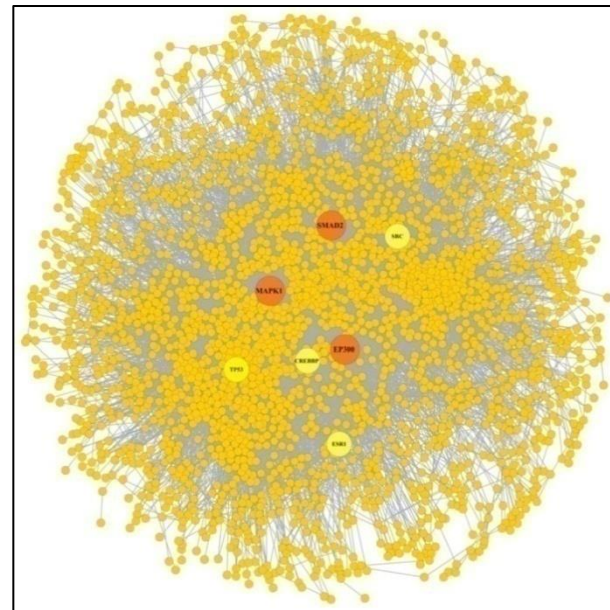
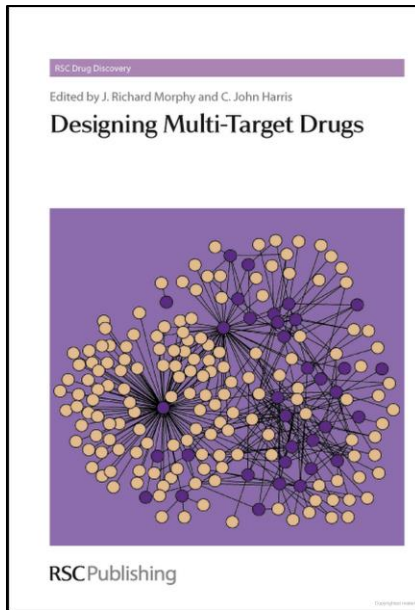
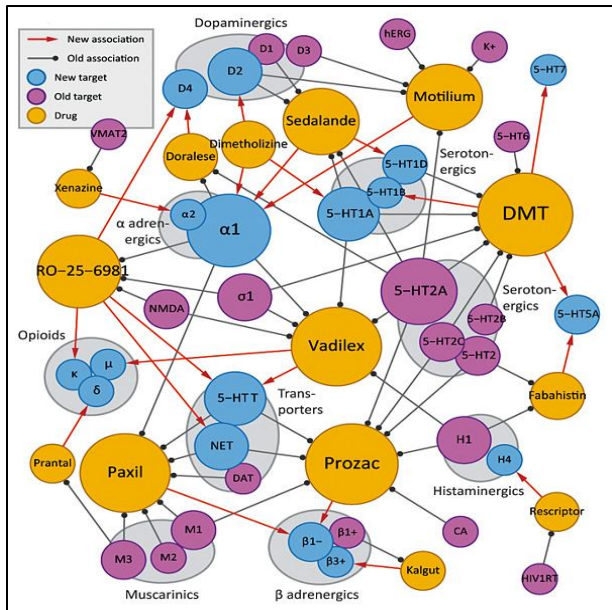
Volgograd State Medical University
Vorozhtsov Institute of Organic Chemistry



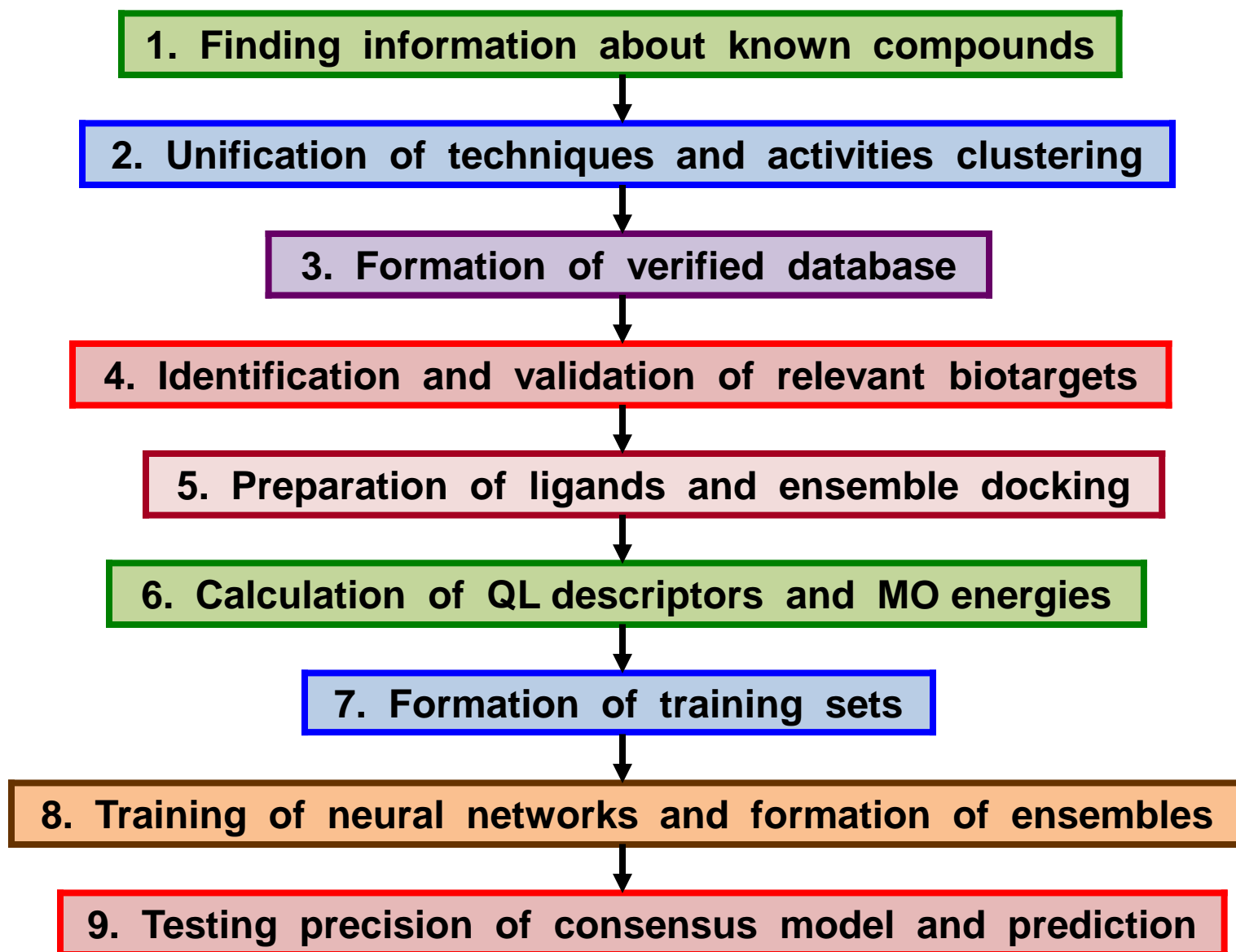
**THE CONSENSUS ENSEMBLE
MULTIDESCRIPTOR MULTITARGET
NEURAL NETWORK MODELING OF
PHARMACOLOGICAL ACTIVITY OF
CHEMICAL COMPOUNDS**

Vassiliev Pavel Mikhailovich

All drugs are multitarget



General model building workflow



Finding information about known compounds

ChEMBL Search in ChEMBL

EBI > Databases > Chemical Biology > ChEMBL Database > Search Results

Search Results

All Results 3717284 Compounds 2157379 Targets 14855 Assays 1458215 Documents 84092 Cells 1991 Tissues 752

Compounds

Show Full Query

Table Cards Graph Heatmap

RAGE-inhibitory activity
364 entries

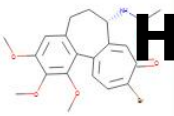
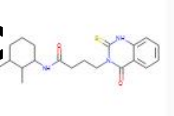

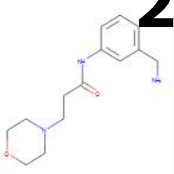
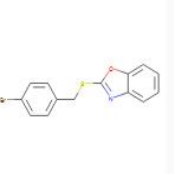
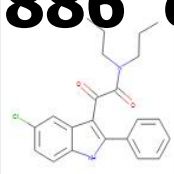
Anxiolytic activity
1 939 entries

2,157,379 Compounds
0 Selected - Select All
Browse Activities

CSV TSV SDF

Records per page: 24 Select All

Showing 1-24 out of 2,157,379 records

 <p>CHEMBL166302 Name: No Data Max Phase: 0 Full Mwt: 263.34 Alogp: 3.63</p> <p>Hypoglycemic activity 2 930 entries</p>	 <p>CHEMBL1542218 Name: No Data Max Phase: 0 Full Mwt: 373.52 Alogp: 3.78</p>	 <p>CHEMBL1454107 Name: No Data Max Phase: 0 Full Mwt: 382.89 Alogp: 3.69</p> <p>Reduced LPS-intoxication 24 886 entries</p>
 <p>CHEMBL484491 Name: No Data Max Phase: 0 Full Mwt: 320.21 Alogp: 4.88</p>	 <p>CHEMBL120536 Name: No Data Max Phase: 0 Full Mwt: 377.48 Alogp: 5.82</p>	 <p>CHEMBL356790 Name: No Data Max Phase: 0 Full Mwt: 430.64 Alogp: 5.63</p>

Unification of techniques and activities clustering

Unification and aggregation of techniques

RAGE-inhibitors — 8

Anxiolytic — 707

Hypoglycemics — 536

LPS-inhibitors — 2 627

Clustering by k-means method

4 activity classes

High, Moderate, Low, Inactive

Cluster Means (Data for Clustering CHEMBL25 Glycation v01)			
Variable	Cluster No. 1	Cluster No. 2	Cluster No. 3
Inhib	63.63278	93.15059	29.47273

Euclidean Distances between Clusters			
Distances below diagonal			
Squared distances above diagonal			
Cluster Number	No. 1	No. 2	No. 3
No. 1	0.00000	871.3011	1166.909
No. 2	29.51781	0.0000	4054.870
No. 3	34.16005	63.6779	0.000

K-means clustering results dialog			
Variable	Cluster No. 1	Cluster No. 2	Cluster No. 3
log(1/IC50)	3.376441	4.259640	5.064370

Analysis of Variance (Data for Clustering CHEMBL25 Glycation v01)					
Variable	Between SS	df	Within SS	df	F
log(1/IC50)	75.30598	2	7.703372	174	850.4

Descriptive Statistics for Cluster 1 (Cluster contains 59 cases)			
Variable	Mean	Standard Deviation	Variance
log(1/IC50)	3.376441	0.201177	0.040472

Descriptive Statistics for Cluster 2 (Cluster contains 71 cases)			
Variable	Mean	Standard Deviation	Variance
log(1/IC50)	4.259640	0.222052	0.049307

Descriptive Statistics for Cluster 3 (Cluster contains 47 cases)			
Variable	Mean	Standard Deviation	Variance
log(1/IC50)	5.064370	0.203475	0.041402

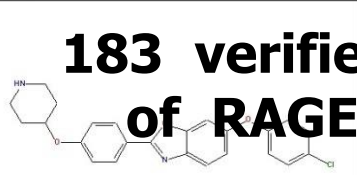
Members of Cluster Number and Distances from Respect Cluster contains 59 cases	
Distance	
361.000000	0.412706
192.000000	0.375586
388.000000	0.371192
377.000000	0.352345

Members of Cluster Number and Distances from Respect Cluster contains 71 cases	
Distance	
348.000000	0.395968
108.000000	0.361599
144.000000	0.357083
52.000000	0.352970

Formation of verified databases

RAGE Antagonists ChEMBL 2018 v03 10 May 2018

Structure



183 verified structures of RAGE-inhibitors

Activity

ACTIVITY_ID	STANDARD_TYPE	STANDARD_RELATION	STANDARD_VALUE	STANDARD_UNITS
15647770	Inhibition	">"	20.6	%

ASSAY_ORGANSIM: Homo sapiens

DESCRIPTION: Inhibition of human RAGE-amyloid beta-42 interaction at 4 μM preincubated for 30 mins measured after 1 hr by FRET assay relative to control


Mol_ID: 1
Brutto formula: C₂₄H₂₁ClN₂O₃
Mol weight: 420.8

Activity Processing: Levels: moderate

References and Comments: Bioorg. Med. Chem. (2015) 23:15:4919

Anxiolytic Substances v11 - Corrected 24 January 2022

Structure



216 verified structures of anxiolytic substances

Activity

STANDARD_TYPE	STANDARD_RELATION	STANDARD_VALUE	STANDARD_UNITS
ED50	">"	0.66	mg.kg-1

ASSAY_ORGANSIM: Rattus norvegicus

ASSAY_DESCRIPTION: Anxiolytic response measured by antagonism of pentylenetetrazole discriminative stimuli in rats.

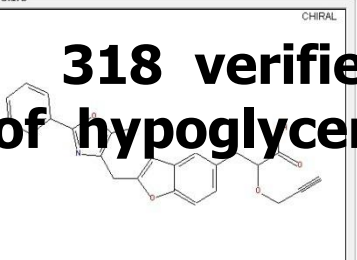
Mol_ID: 1
Brutto formula: C₂₁H₁₈FN₅O
Mol weight: 375.3

Activity Processing: Levels: high

References: Document_CHEMBL_ID: CHEMBL1125833, Document_Year: 1991

Hypoglycemic Substances v16 - MW <= 1000 02 December 2020

Structure



318 verified structures of hypoglycemic substances

Activity

STANDARD_TYPE	STANDARD_RELATION	STANDARD_VALUE	STANDARD_UNITS
Glucose	">"	100	%

ASSAY_ORGANSIM: Mus musculus

ASSAY_DESCRIPTION: In vivo hypoglycemic activity expressed as percent glucose normalization in db/db mice at 5 mg/Kg

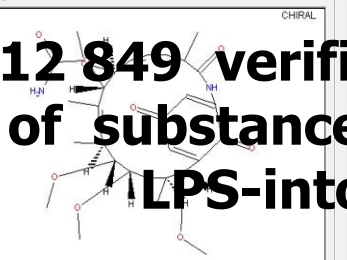
Mol_ID: 1
Brutto formula: C₂₅H₂₁N₅
Mol weight: 415.4

Activity Processing: Levels: high

References: Document_CHEMBL_ID: CHEMBL1129423, Document_Year: 1996

LPS_v12 - Final 07 February 2021

Structure



12 849 verified structures of substances decreasing LPS-intoxication

Activity

STANDARD_TYPE	STANDARD_RELATION	STANDARD_VALUE	STANDARD_UNITS
IC50	">"	94	nM

ASSAY_ORGANSIM: Mus musculus

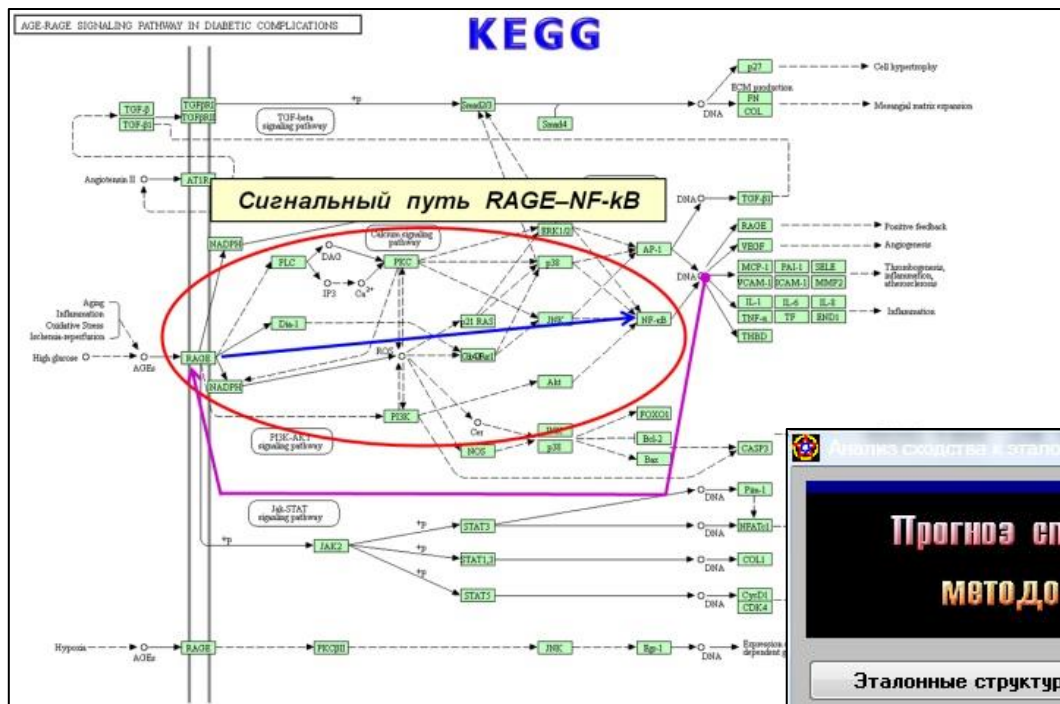
ASSAY_DESCRIPTION: Inhibition of LPS-induced NO production in rat mouse macrophages after 20 hrs by Griess reagent assay

Mol_ID: 1
Brutto formula: C₃₀H₄₂N₂O₈
Mol weight: 574.6

Activity Processing: Levels: high

References: Document_CHEMBL_ID: CHEMBL1151246, Document_Year: 2004

Identification of relevant biotargets



Signaling pathway analysis

Context prediction of target activity using chemotypes

Методы сходства в эталонах по выбранным активностям

Прогноз спектра целевой биологической активности
методом максимального сходства к эталонам

Эталонные структуры

Прогнозируемые структуры

Вывод в файл:

Число ближайших эталонов
10

Файл со списком выбранных активностей

Расчитать спектр активности

MICROCOSM BIOS

Searching standard compounds

IUPHAR/BPS Guide to PHARMACOLOGY
An expert-curated resource of pharmacological targets and the substances that act on them

Home About Targets Ligands Diseases Resources Advanced search Immuno Portal Malaria Portal

OtoPdb home page tour

Quick links

Targets

- G protein-coupled receptors
- Ion channels
- Nuclear hormone receptors
- Kinases
- Catalytic receptors
- Transporters
- Enzymes
- Other protein targets

Ligands

- Approved drugs
- Synthetic organics
- Metabolites
- Natural products
- Endogenous peptides
- Other peptides
- Inorganics
- Antibodies
- Labeled ligands

Resources

- Help documentation
- FAQ
- Tutorial
- Download data & reports
- REST web services

Recent Twitter activity

Sorry, this content is unavailable until you click to accept cookies from this site.

Selection of
structure similarity
inactive compounds

Прогноз спектра целевой биологической активности
методом максимального сходства к эталонам

Эталонные структуры

Прогнозируемые структуры

Вывод в файл:

Число ближайших эталонов
10

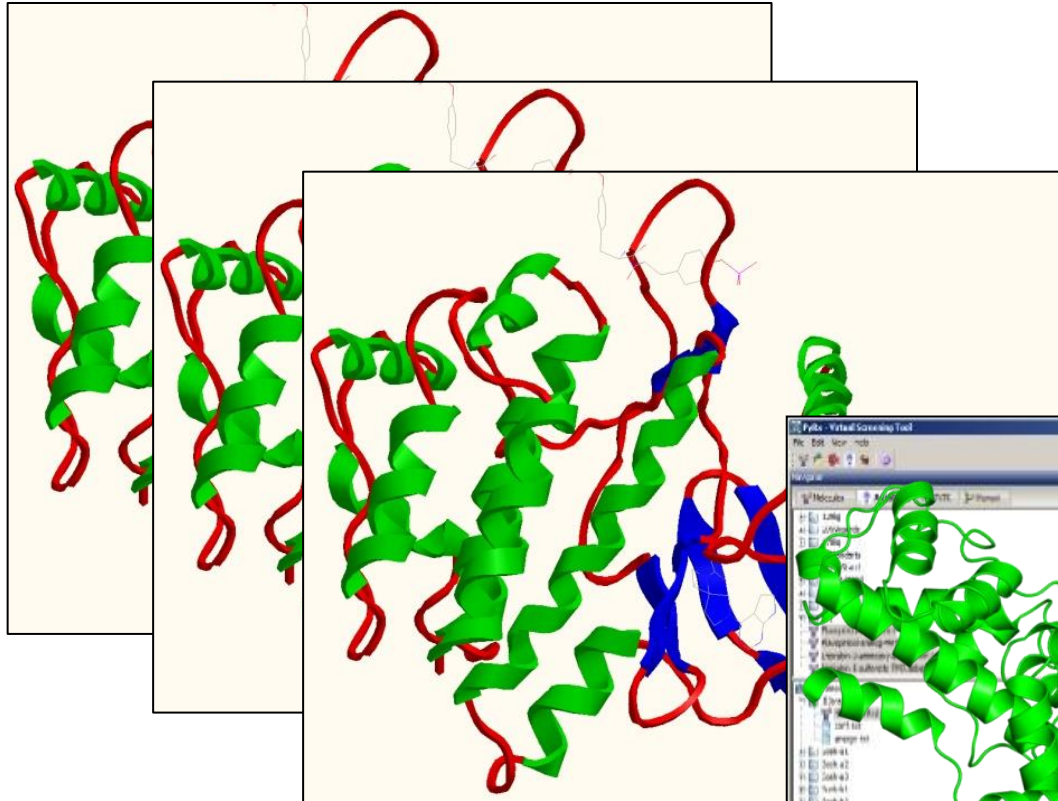
Файл со списком выбранных активностей

Рассчитать спектр активности

MICROCOSM BIOS

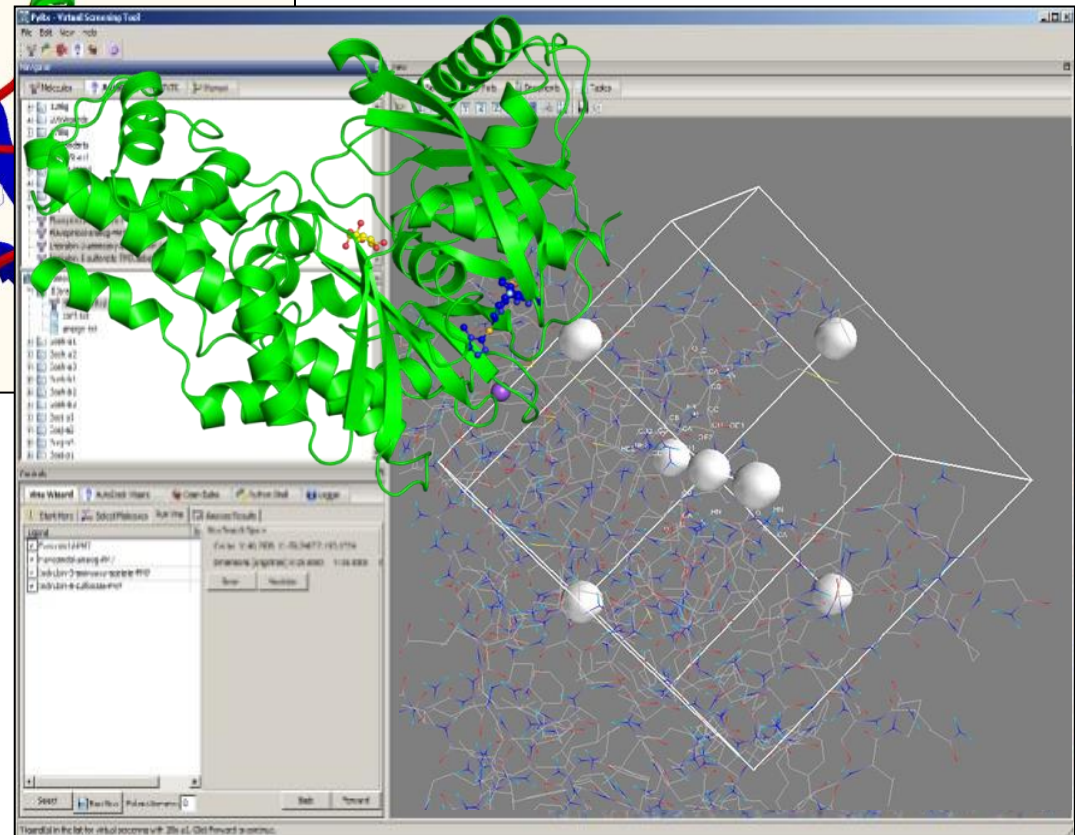
Selection of known
active reference
compounds

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**

EMBL-EBI

Advanced Search

Help Contact

Other tutorials and videos

Services Research Training About us

Search

Advanced search

Feedback

Quick links

- 3iW4 overview
- Citations
- Structure analysis

Released: 03 Nov 2009

Model geometry

Fit model/data

Crystal structure of PKC alpha in complex with NVP-AEB071

Source organism: *Homo sapiens*

Primary publication:

Discovery of 3-((1H-indol-3-yl)-4-[2-(4-methylpiperazin-1-yl)quinazolin-4-yl]pyrrole-2,5-dione (AEB071), a potent and selective inhibitor of protein kinase C isotypes.

Wagner J, von Matt P, Sedrani R, Albert R, Cooke N, Ehrhardt C, Geiser F, Strauss A, Cowan-Jacob SW, Beerli C, Weckbecker G, Evenou JP, Zenke J. *J. Med. Chem.* 52 6193-6 (2009)

PMID: 19827831

Function and Biology

Details

Reaction catalysed:

ATP + a protein = ADP + a phosphoprotein

Biochemical function:

- not assigned

Biological process:

- not assigned

Cellular component:

- not assigned

Sequence domains:

- AGC-kinase, C-terminal
- Protein kinase domain
- Protein kinase, C-terminal
- Protein kinase, ATP binding site
- Serine/threonine-protein kinase, active site
- Classical Protein Kinase C alpha, catalytic domain
- Protein kinase-like domain superfamily

Model Overview

Go to Sequence Overview

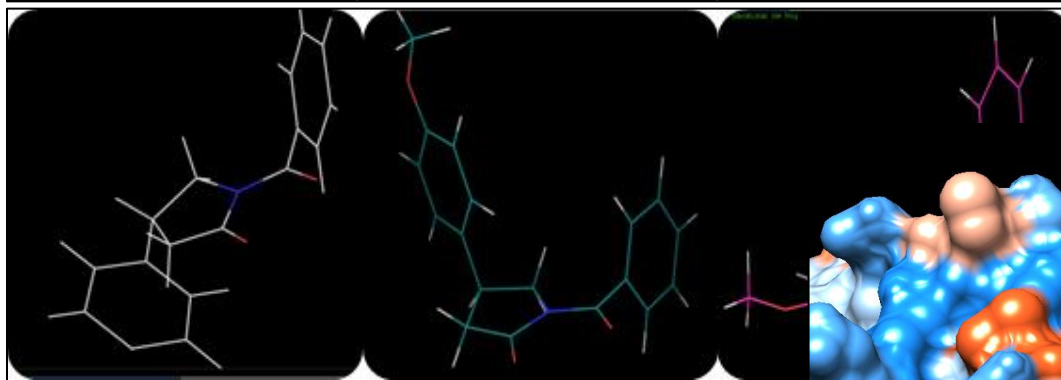
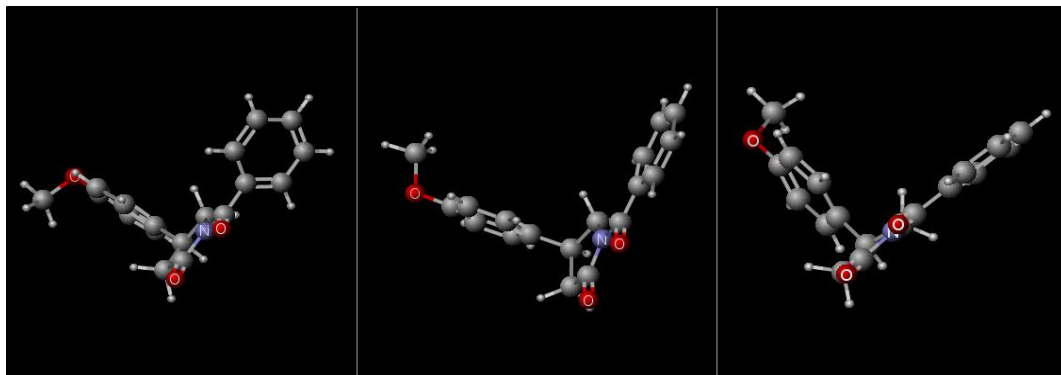
Search Summary

Perform Action on Selected Models

Ligand

13;

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

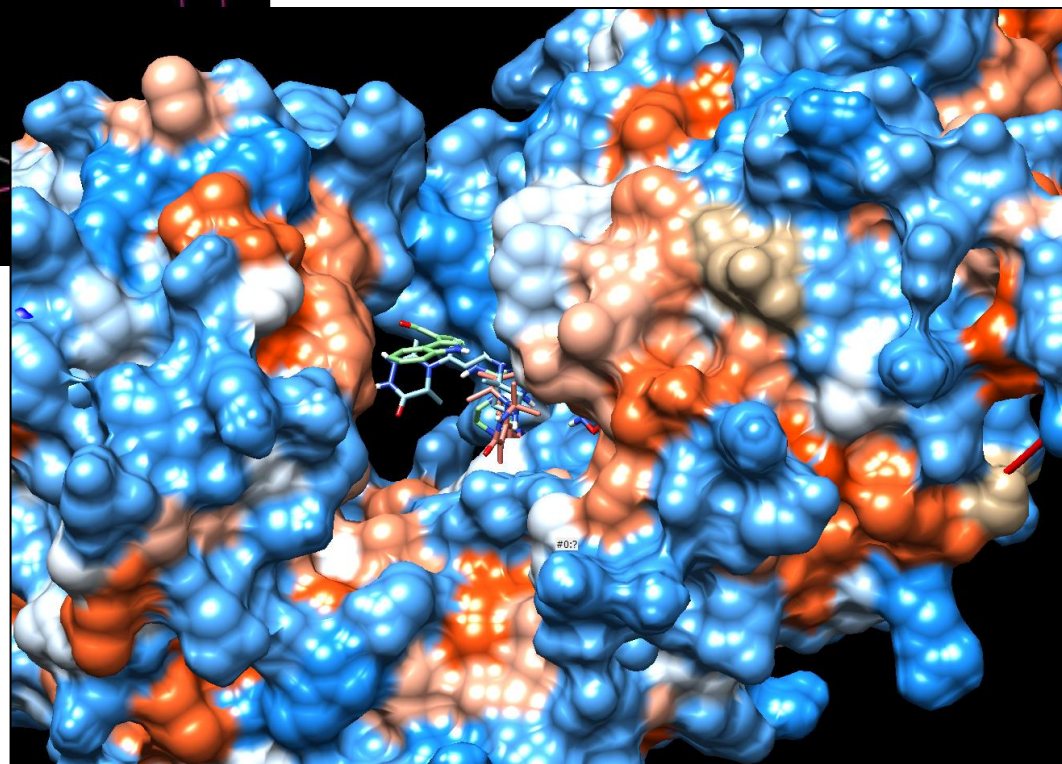
Molecular mechanics

10 conformers

Quantum chemistry PM7

Optimization and selection
of the best conformation

~1 300 000 conformers



Calculation of QL descriptors and MO energies

Трансляция набора структур в формат QL-2

О программе

Набор структур

LPS_v02.sdf

Транслируется набор

Создание набора файлов для прогноза

Входной файл (QL)

Выходной набор

Создать

LPS_v02

LPS_v02

Идет создание набора файлов для прогноза...

9%

IT Microcosm
QL-descriptors
1st rank

Visual Basic script

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

HyperChem - Proglitazon DD PM3 6-31G+.hin

File Edit Build Select Display Databases Setup Compute Annotations Script Cancel Help

PH7 XYZ AUX GNORM=0.01 NORM
CHEM1102740-010

GEOMETRY OPTIMISED USING EIGENVECTOR FOLLOWING (EF).
SCF FIELD WAS ACHIEVED

PM7 CALCULATION

MOPAC2012 (Version: 12.239L)

FINAL HEAT OF FORMATION = 18.79346 KCAL/MOL = 78.63184 KJ/MOL

TOTAL ENERGY = -2070.38874 EU
ELECTRONIC ENERGY = -11546.38462 EU POINT GROUP: C2
CORE-CORE REPELSION = 29175.05588 EU
COSMO AREA = 227.40 SQUARE ANGSTROMS
COSMO VOLUME = 230.36 CUBIC ANGSTROMS

GRADIENT NORM = 0.05185
IONIZATION POTENTIAL = 9.500209 EU
HOMO LUMO ENERGIES (EU) = -9.500 1.168
NO. OF FILLED LEVELS = 35
MOLECULAR WEIGHT = 172.233

MOLECULAR DIMENSIONS (Angstroms)

Atom	Atom	Distance
H 27	H 15	10.86375
H 26	H 18	4.71317
H 13	H 24	3.79720

SCF CALCULATIONS = 362
COMPUTATION TIME = 4.886 SECONDS

45 atoms. PM3

Training of neural networks and formation of ensembles

3 activity levels

Net. ...	Net. name	Training ...	Test p...	Algorit...	Error fu...	Hidden ...	Output ...
1	MLP 22-1...	78.911565	80.555...	BFGS 14	SOS	Tanh	Logistic
2	MLP 22-1...	89.75646	80.555...	BFGS 9	SOS	Exponen...	Logistic
3	MLP 22-1...	79.591837	80.555...	BFGS 14	SOS	Exponen...	Tanh
4	MLP 22-1...	78.911565	80.555...	BFGS 14	SOS	Logistic	Tanh
5	MLP 22-9-2	78.251293	80.555...	BFGS 11	SOS	Identity	Tanh
6	MLP 22-1...	78.911565	80.555...	BFGS 9	SOS	Tanh	Logistic

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

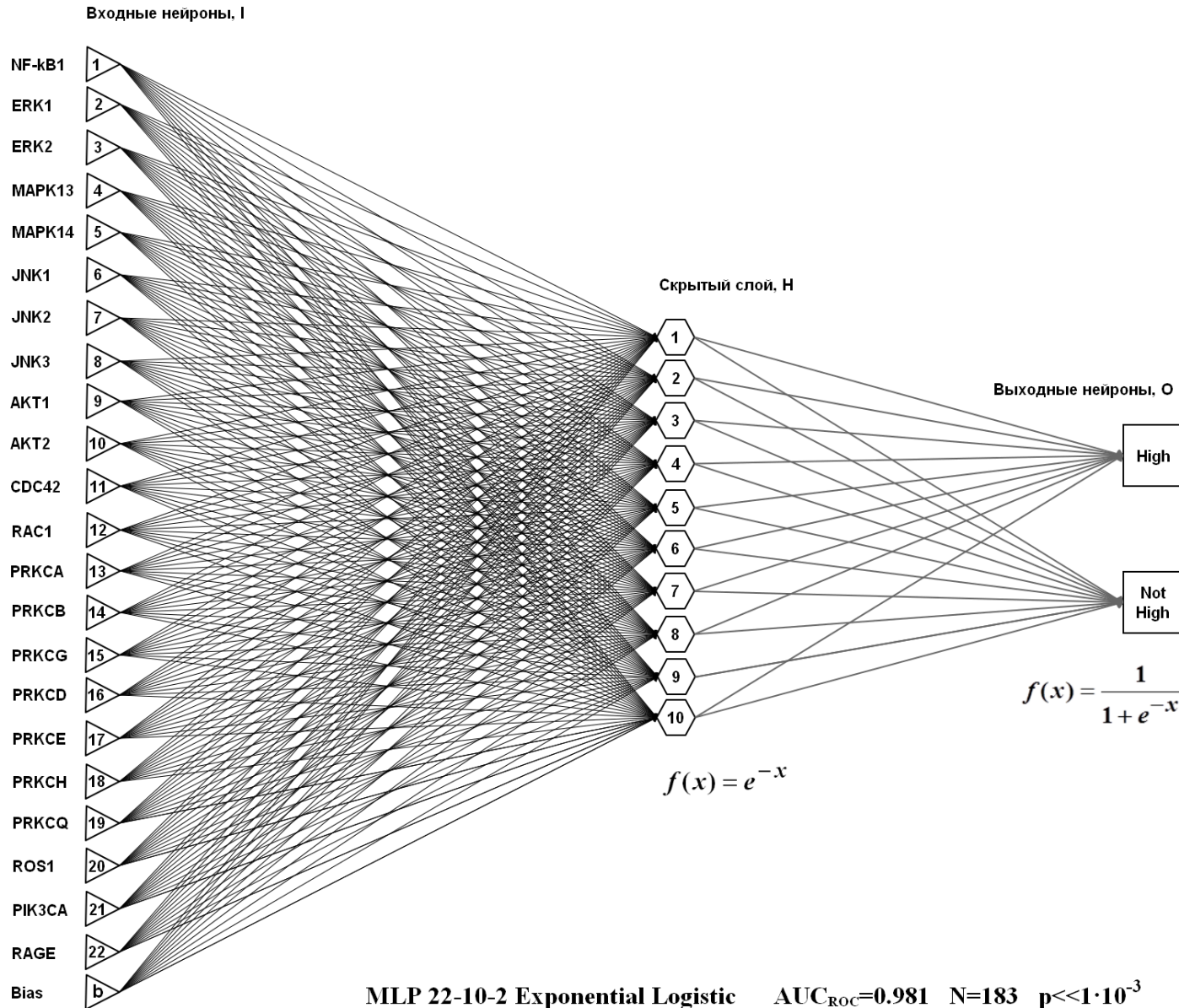
Neural network training in progress...
Cycle=40:
Classification rate: train=85.034, test=75

Next
Finish

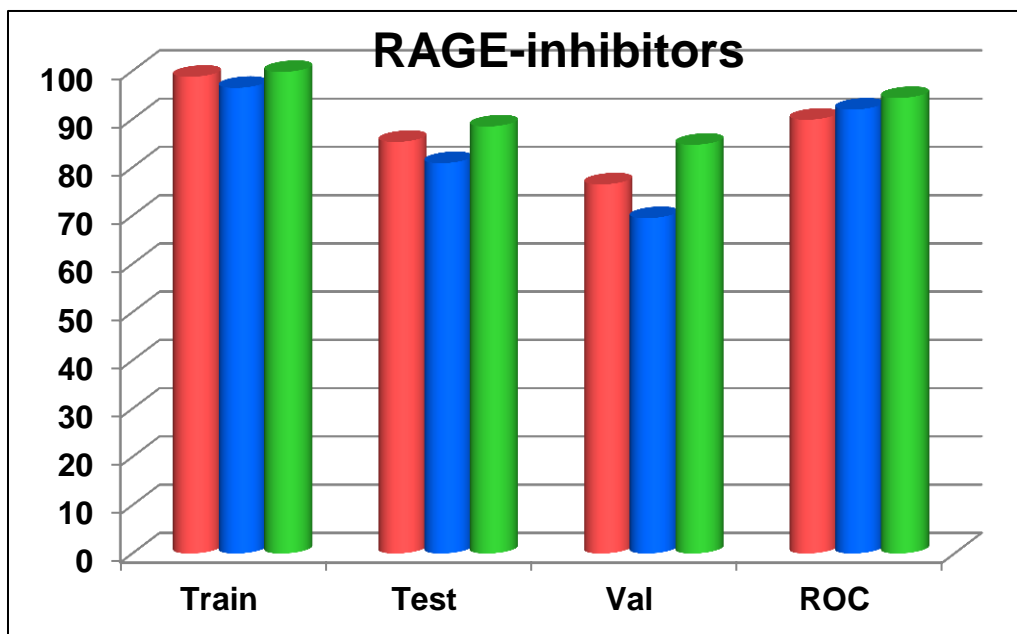
Sample
 Train
 Test
 Missing

An example of a neural network architecture

RAGE-inhibition from docking energies for 34 biotargets

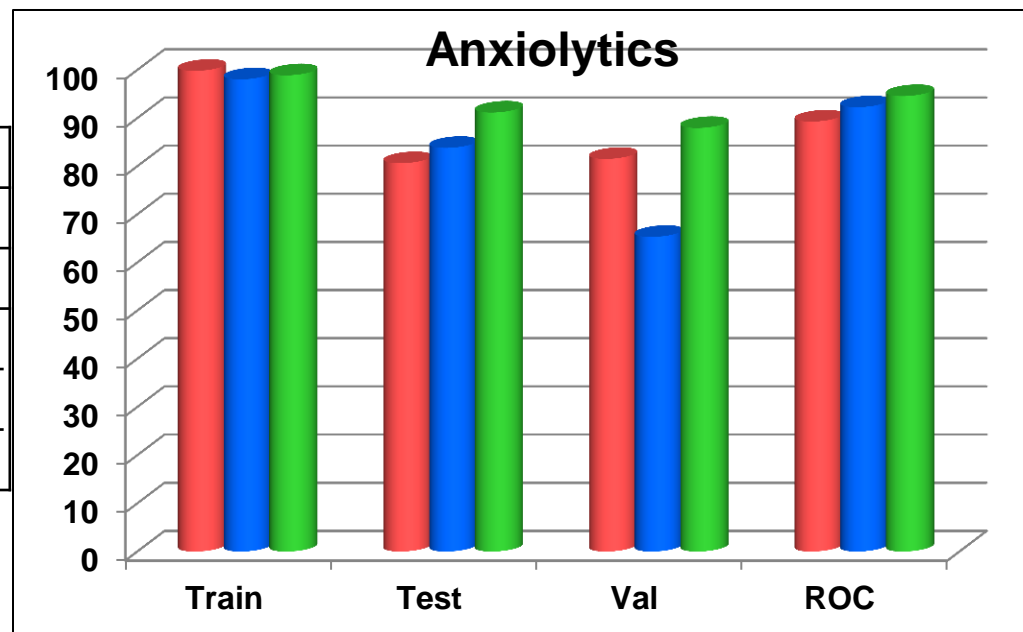


Accuracy of neural network ensembles

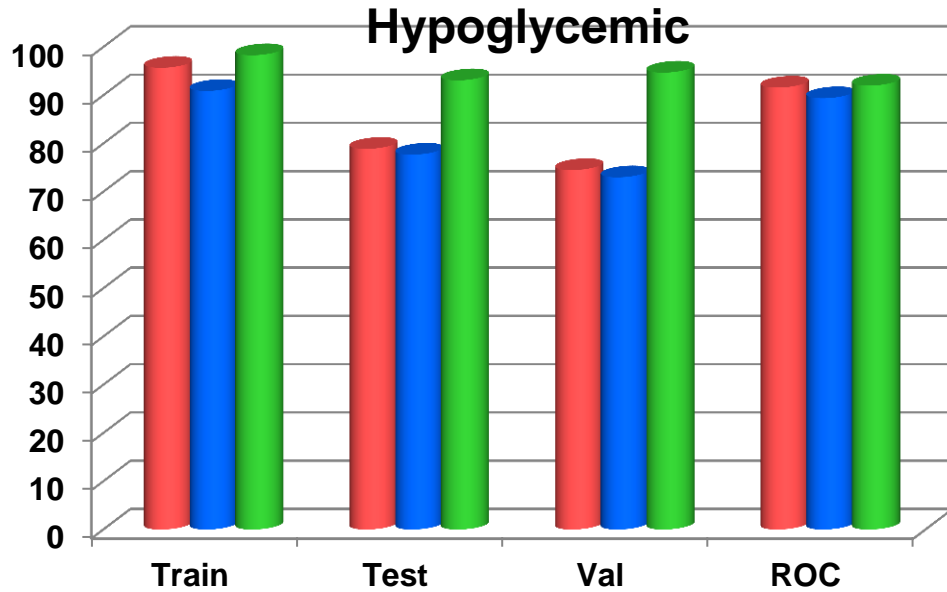


Activity level	Accuracy, %			
	Train	Test	Val	ROC
RAGE-inhibitory activity				
High	98.8	85.3	76.5	89.9
High or Moderate	96.5	80.9	69.5	92.0
Active	99.8	88.5	84.7	94.4

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

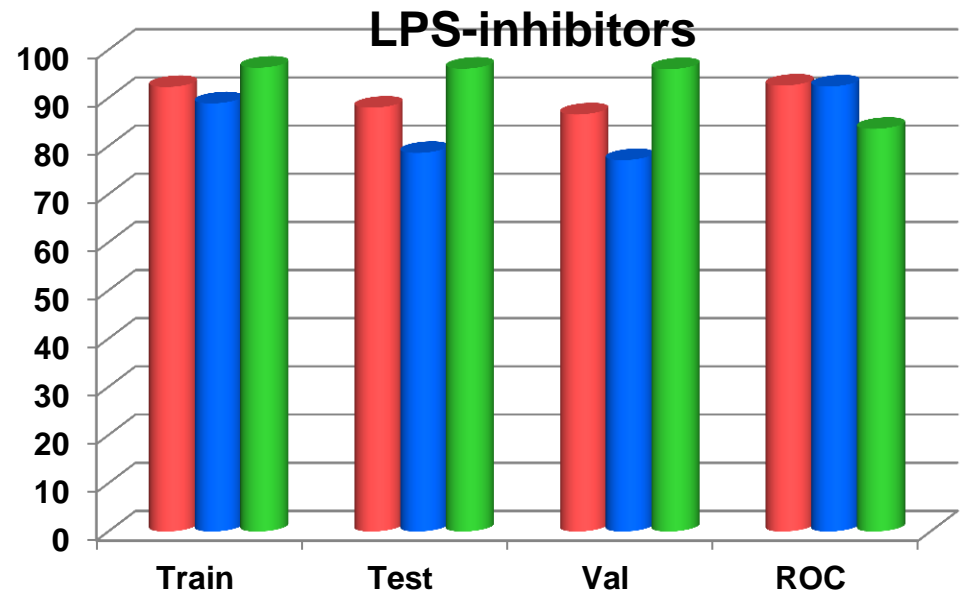


Accuracy of neural network ensembles



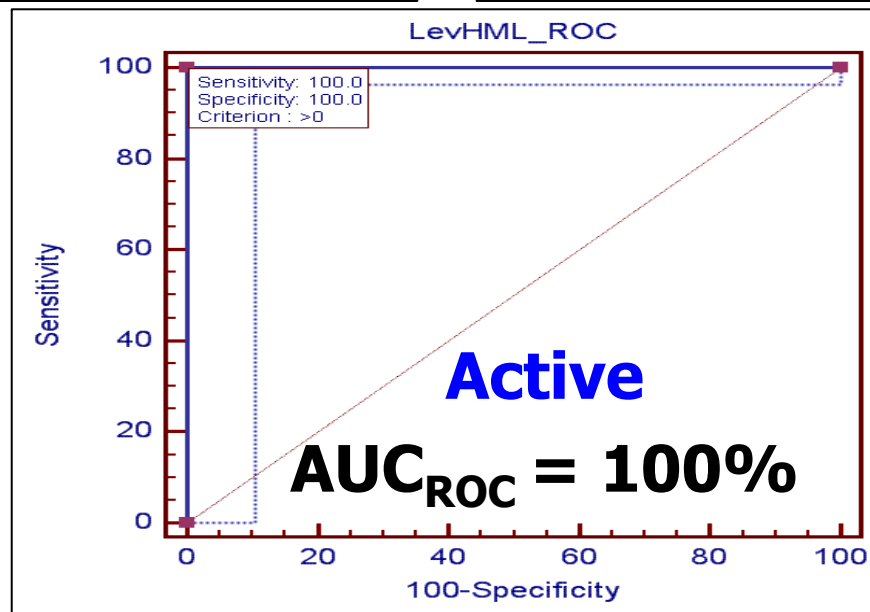
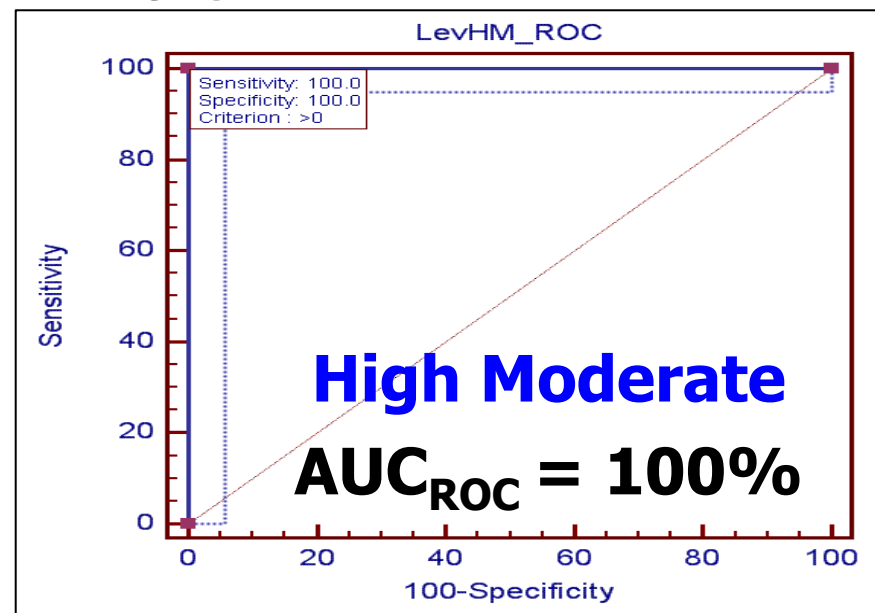
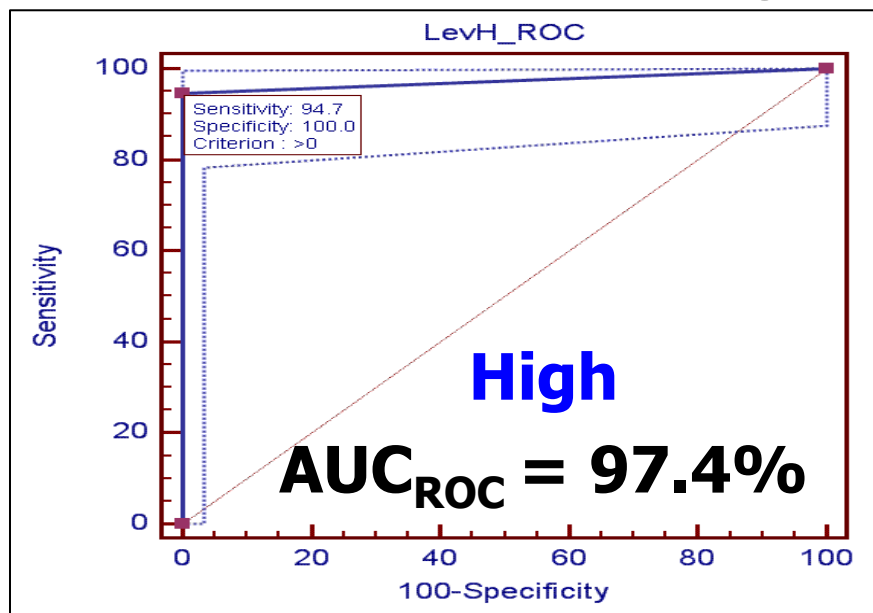
Activity level	Accuracy, %			
	Train	Test	Val	ROC
Hypoglycemic activity				
High	95.7	78.9	74.5	91.6
High or Moderate	90.9	77.7	73.0	89.4
Active	98.3	93.1	94.7	92.0

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

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	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			
Oshi10	h	h	h		hm	hm	hm	hm	hml	hml	hml	hml	1	1	1					
Oshi11	h				hm	hm			hml	hml	hml						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:**

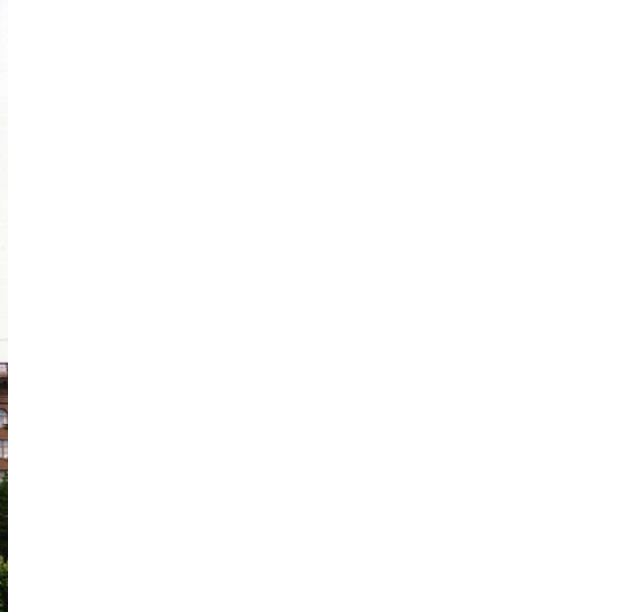
RFBR № 18-015-00499

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!

