Volgograd State Medical University

Research Center of Innovative Medicines

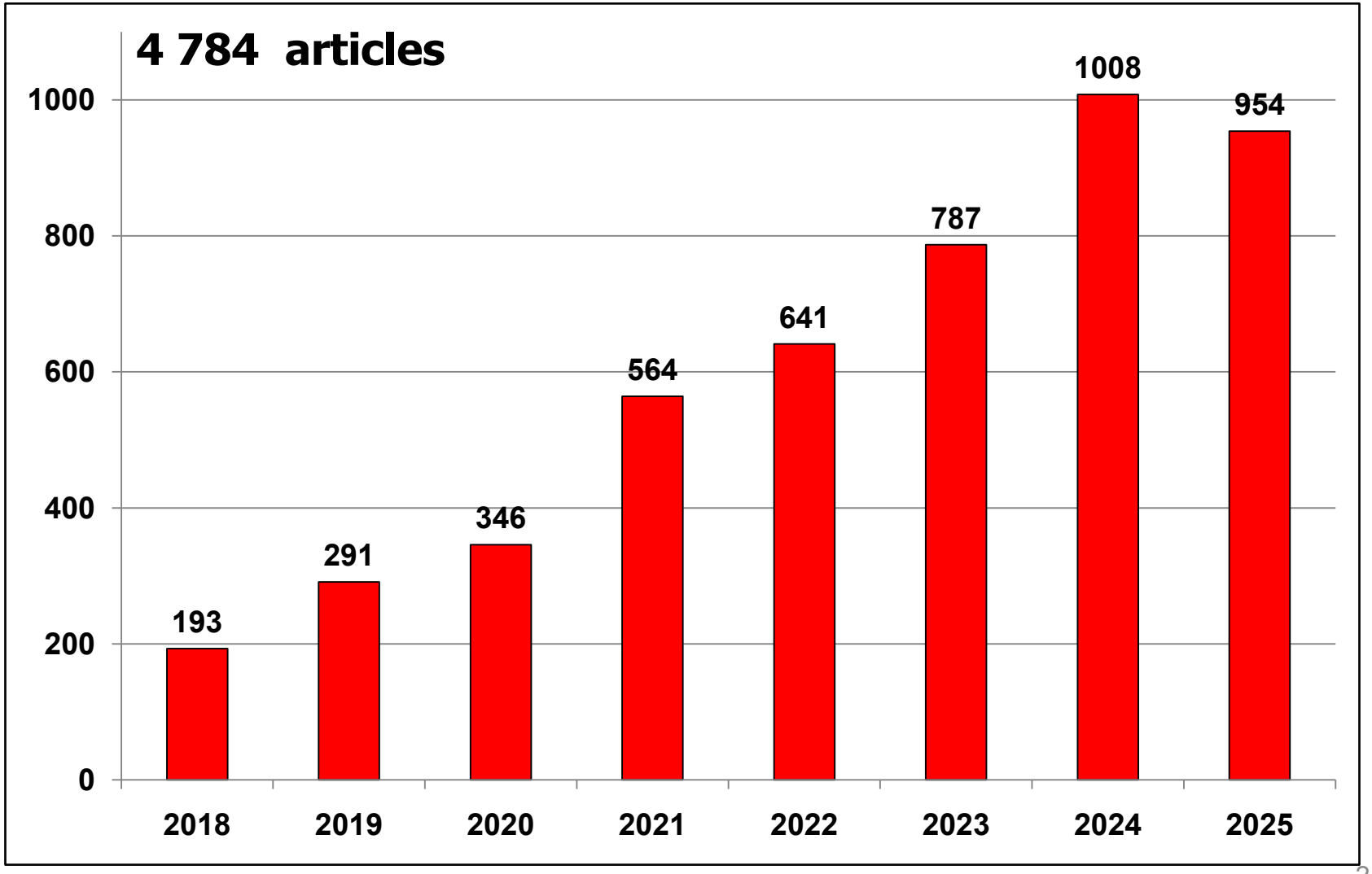
Laboratory for Information Technology in Pharmacology and Computer Modeling of Drugs

FULLY-CONNECTED CONVOLUTIONAL NEURAL NETWORKS BASED ON MULTIPLE DOCKING: A NEW MACHINE LEARNING METHOD FOR SEARCHING BIOLOGICAL ACTIVE COMPOUNDS

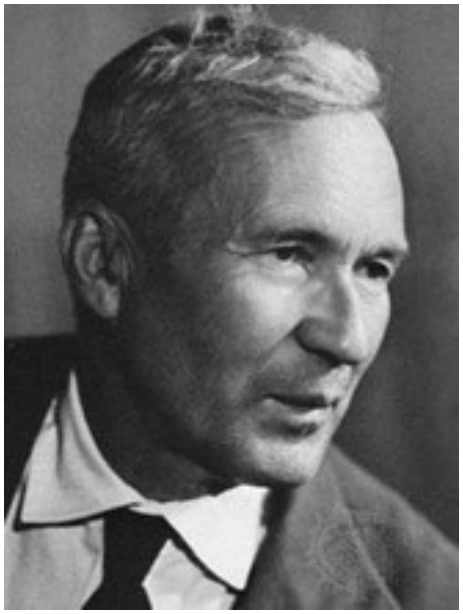
Vassiliev Pavel Mikhailovich

Publications in PubMed — Drug Design

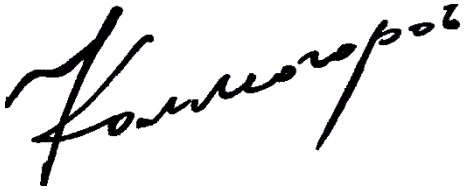
("machine learning" OR "artificial neural network" OR "artificial intelligence") AND ("drug discovery" OR "drug design")



Kolmogorov's theorem



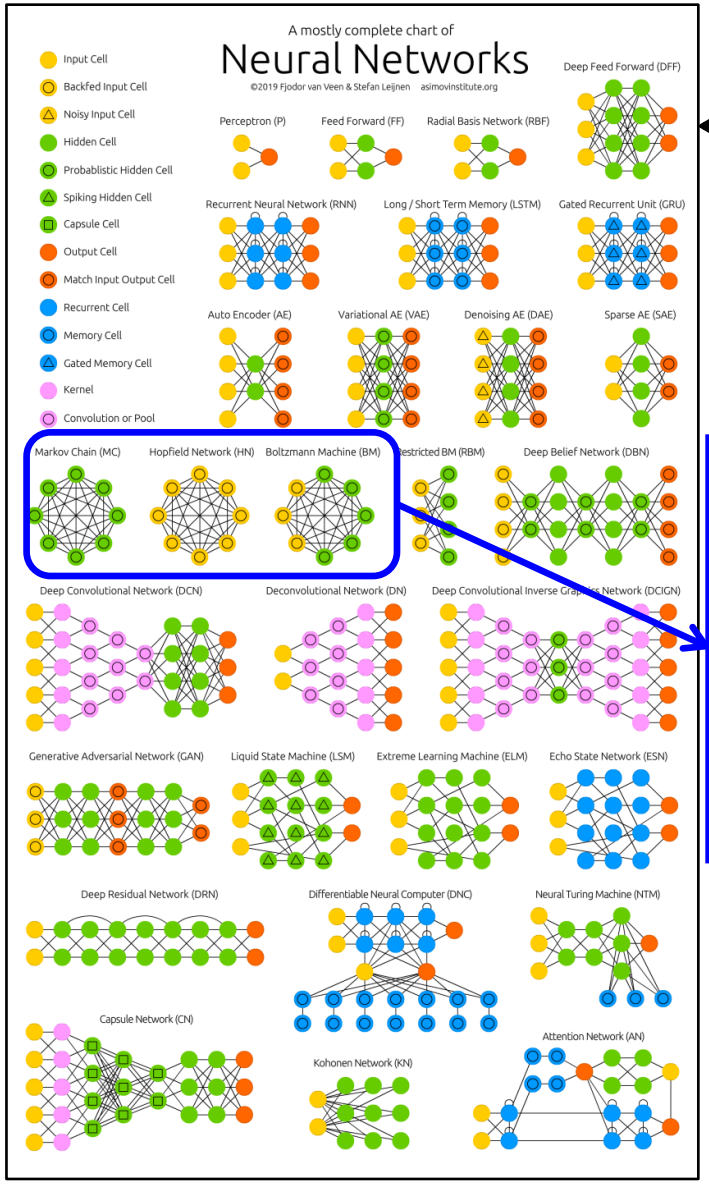
Any continuous function can be represented as a superposition of continuous functions of one or two arguments

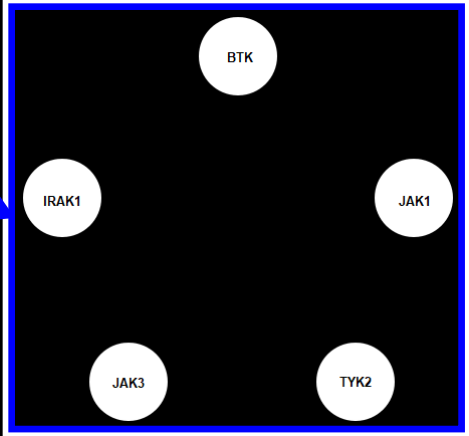


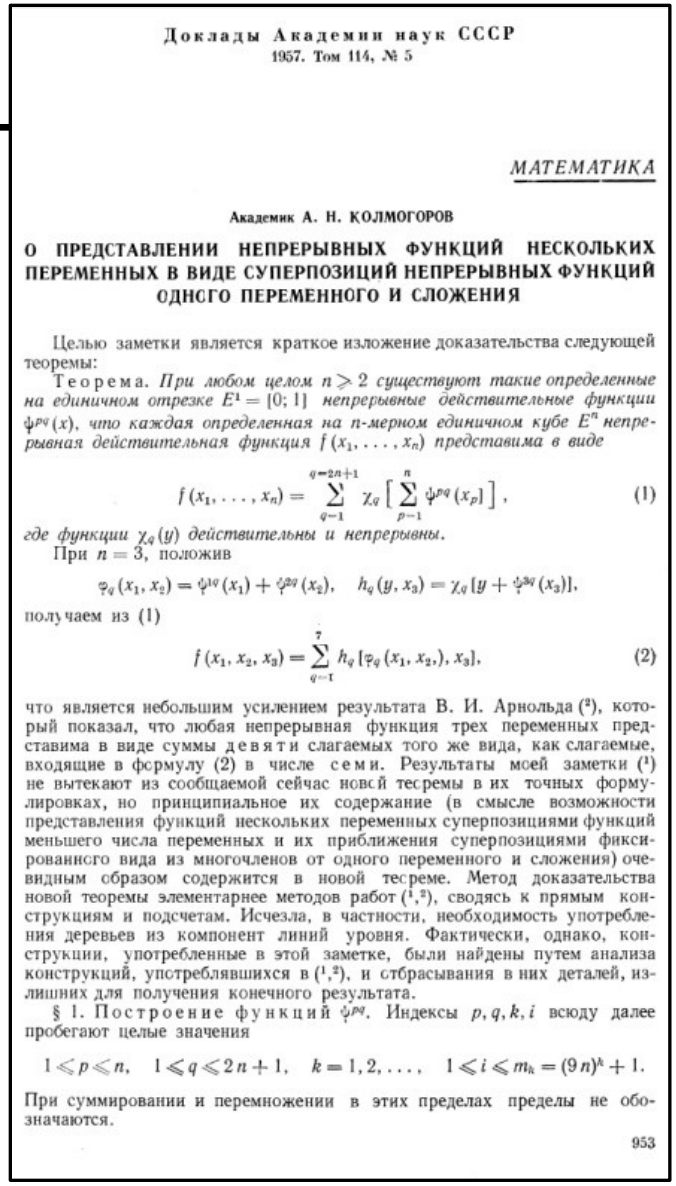
Andrei Nikolaevich Kolmogorov, 1957

Kolmogorov A. N. Proceedings of the USSR Academy of Sciences, 1957; 114(5):953-956.

The Neural Network Zoo







Доклады Академии наук СССР
1957. Том 114, № 5

МАТЕМАТИКА

Академик А. Н. КОЛМОГОРОВ

О ПРЕДСТАВЛЕНИИ НЕПРЕРЫВНЫХ ФУНКЦИЙ НЕСКОЛЬКИХ ПЕРЕМЕННЫХ В ВИДЕ СУПЕРПОЗИЦИЙ НЕПРЕРЫВНЫХ ФУНКЦИЙ ОДНОГО ПЕРЕМЕННОГО И СЛОЖЕНИЯ

Целью заметки является краткое изложение доказательства следующей теоремы:

Теорема. *При любом целом $n \geqslant 2$ существуют такие определенные на единичном отрезке $E^1 = [0; 1]$ непрерывные действительные функции $\psi^{pq}(x)$, что каждая определенная на n-мерном единичном кубе E^n непрерывная действительная функция $f(x_1, \ldots, x_n)$ представима в виде*

$$f(x_1, \ldots, x_n) = \sum_{q=1}^{q=2n+1} \chi_q \left[\sum_{p=1}^{n} \psi^{pq}(x_p) \right], \qquad (1)$$

где функции $\chi_q(y)$ действительны и непрерывны.

При $n = 3$, положив

$$\varphi_q(x_1, x_2) = \psi^{1q}(x_1) + \psi^{2q}(x_2), \quad h_q(y, x_3) = \chi_q[y + \psi^{3q}(x_3)],$$

получаем из (1)

$$f(x_1, x_2, x_3) = \sum_{q=1}^{7} h_q[\varphi_q(x_1, x_2), x_3], \qquad (2)$$

что является небольшим усилением результата В. И. Арнольда [2], который показал, что любая непрерывная функция трех переменных представима в виде суммы девяти слагаемых того же вида, как слагаемые, входящие в формулу (2) в числе семи. Результаты моей заметки [1] не вытекают из сообщаемой сейчас новой теоремы в их точных формулировках, но принципиальное их содержание (в смысле возможности представления функций нескольких переменных суперпозициями функций меньшего числа переменных и их приближения суперпозициями фиксированного вида из многочленов от одного переменного и сложения) очевидным образом содержится в новой теореме. Метод доказательства новой теоремы элементарнее методов работ [1,2], сводясь к прямым конструкциям и подсчетам. Исчезла, в частности, необходимость употребления деревьев из компонент линий уровня. Фактически, однако, конструкции, употребленные в этой заметке, были найдены путем анализа конструкций, употреблявшихся в [1,2], и отбрасывания в них деталей, излишних для получения конечного результата.

§ 1. Построение функций ψ^{pq}. Индексы p, q, k, i всюду далее пробегают целые значения

$$1 \leqslant p \leqslant n, \quad 1 \leqslant q \leqslant 2n+1, \quad k = 1, 2, \ldots, \quad 1 \leqslant i \leqslant m_k = (9n)^k + 1.$$

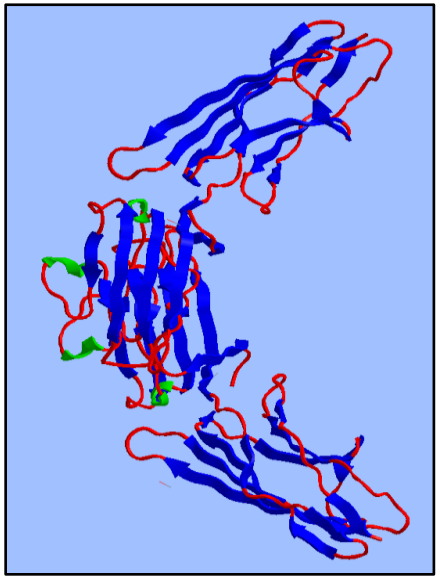
При суммировании и перемножении в этих пределах пределы не обозначаются.

953

Leijnen S., Van Veen F. Proceedings, 2020; 47(4): 9.

Multiple docking

Ligands interact with the protein throughout its entire volume



- Active concentration 10^{-10} M.

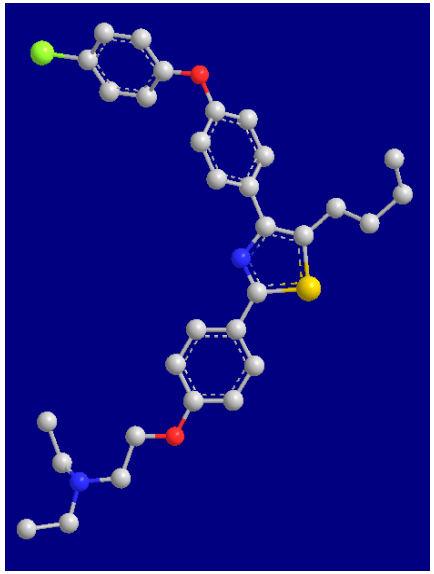
Avogadro's number ~$6 \cdot 10^{23}$.

~$6 \cdot 10^{13}$ ligand molecules interact with the target protein!

- RAGE volume 34 902 Å^3.

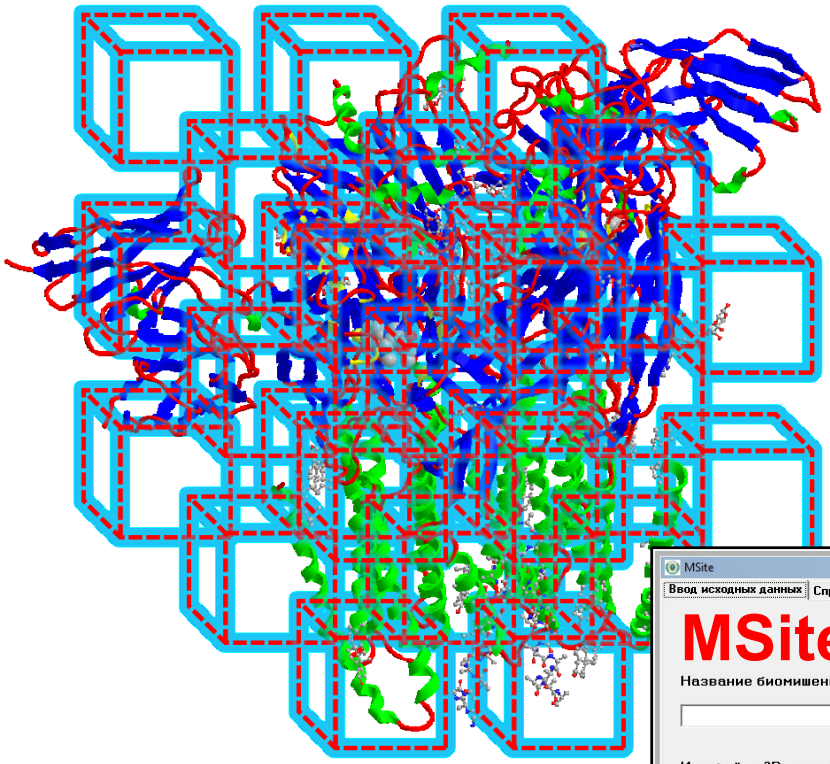
Ligand volume 1 578 Å^3.

The ligand fits into the receptor space at least 22 times!



Vassiliev P. M., et al. Volgograd Scientific Medical Journal, 2021; 3:57-61.

Formation of spaces for multiple docking



3·3·3
27 spaces

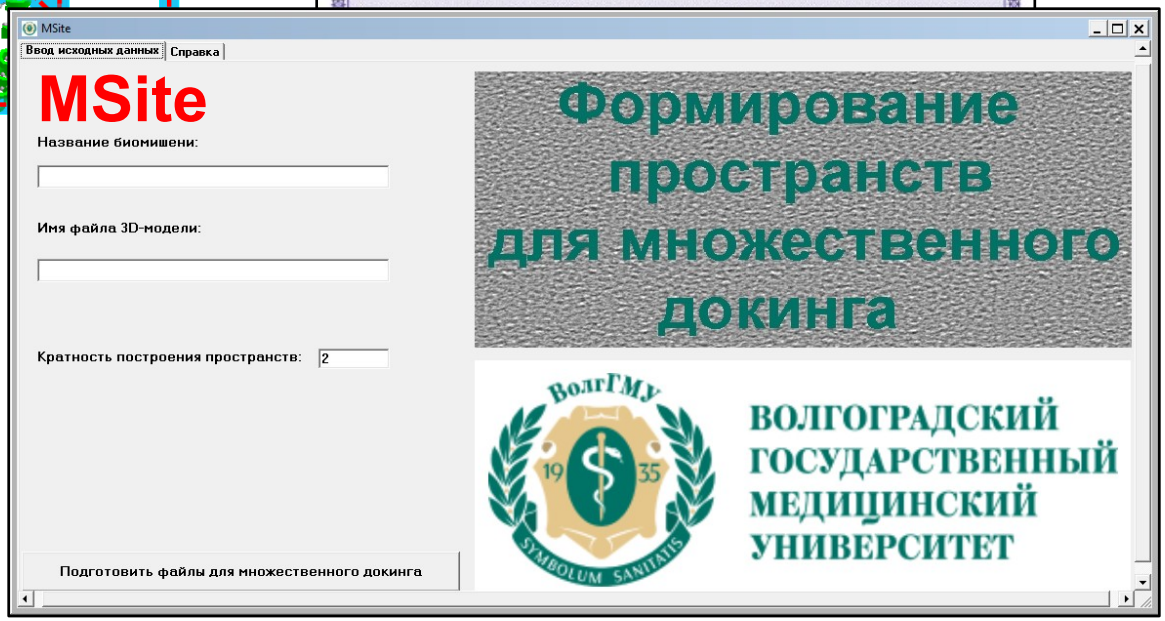

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

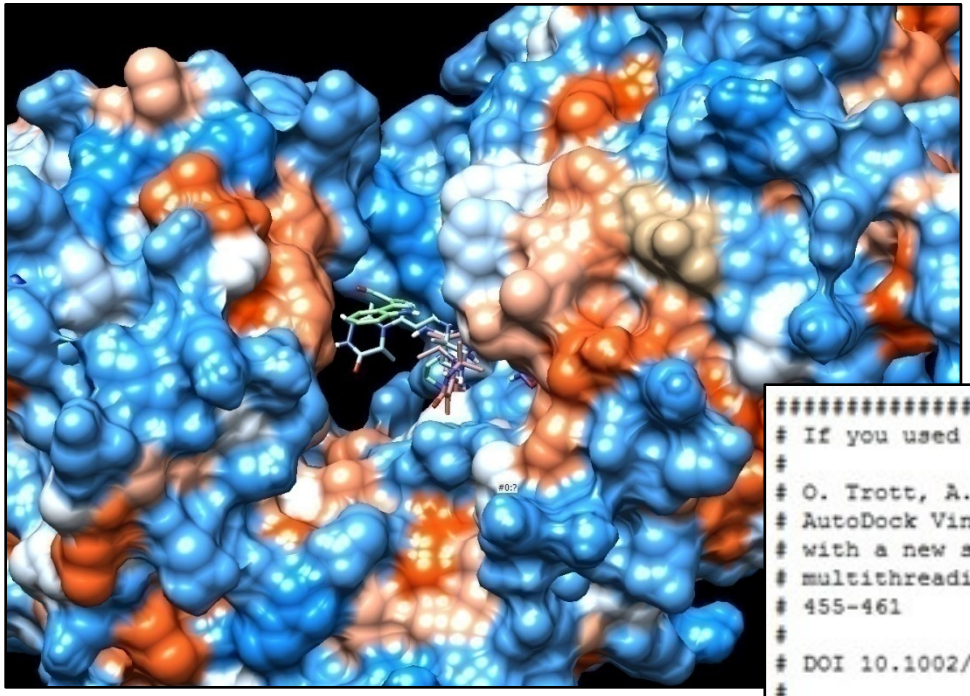
№ 2025682291

Формирование пространств для множественного докинга MSite



Vassiliev P. M., et al. Journal of Volgograd State Medical University, 2022; 19(4):88-93.

Multiple ensemble docking to relevant biotargets



Each of N ligands
to each of M 3D models
to each of 27 spaces
5 times in 10 conformations

$$\Delta E_{lmi} = \min(\Delta E_{lmijk})$$

AutoDock Vina

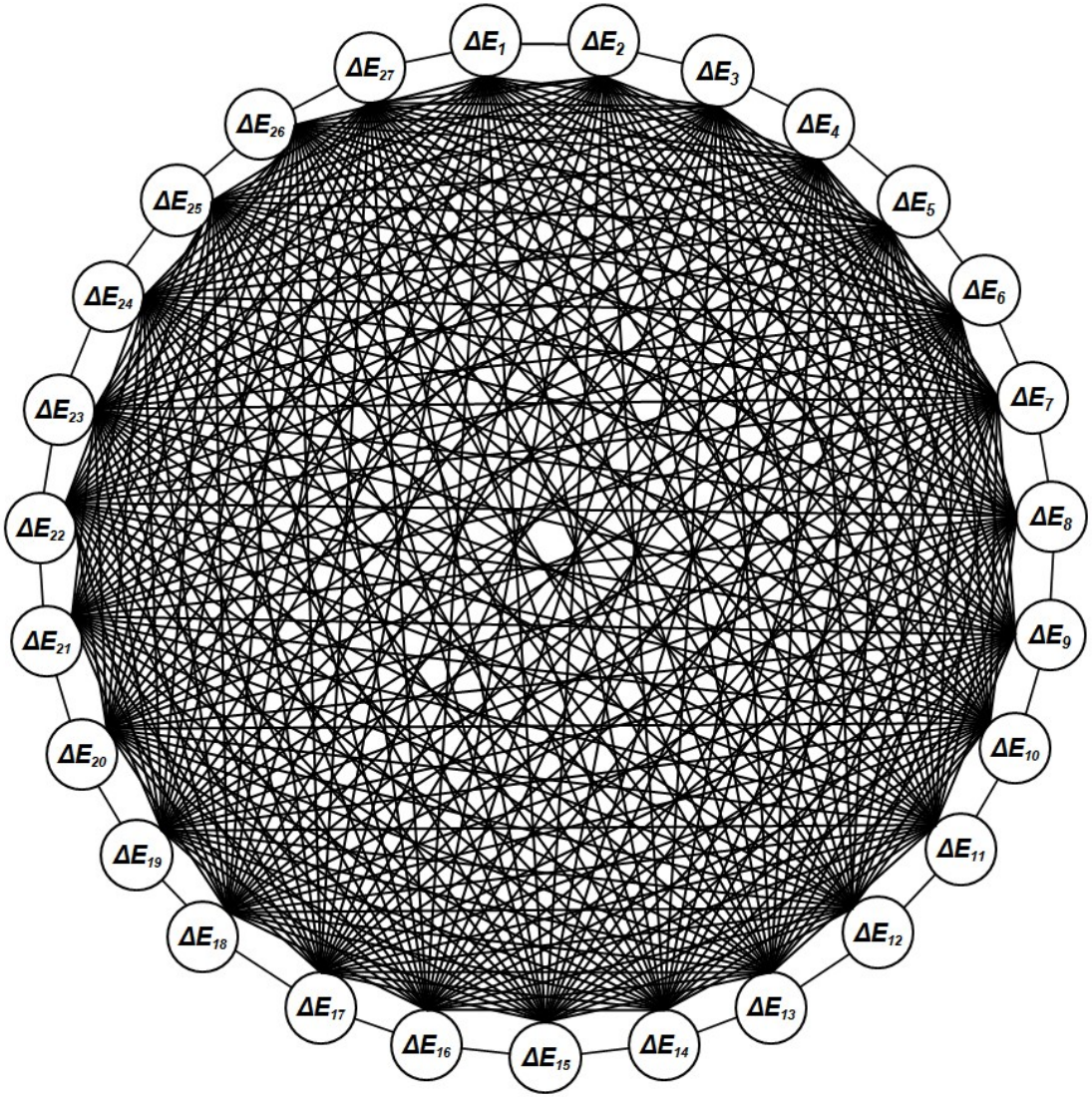
```
#################################################################
# If you used AutoDock Vina in your work, please cite:          #
#                                                               #
# O. Trott, A. J. Olson,                                        #
# AutoDock Vina: improving the speed and accuracy of docking    #
# with a new scoring function, efficient optimization and       #
# multithreading, Journal of Computational Chemistry 31 (2010)  #
# 455-461                                                       #
#                                                               #
# DOI 10.1002/jcc.21334                                         #
#                                                               #
# Please see http://vina.scripps.edu for more information.      #
#################################################################

Output will be ..\..\Ligands\Flavopiridol-PM7_out.pdbqt
Detected 12 CPUs
WARNING: at low exhaustiveness, it may be impossible to utilize all CPUs
Reading input ... done.
Setting up the scoring function ... done.
Analyzing the binding site ... done.
Using random seed: 58910256
Performing search ...
0%   10   20   30   40   50   60   70   80   90   100%
|----|----|----|----|----|----|----|----|----|----|
**********
```

Total processed
N · M · 27 · 5 · 10
~ 1350 · N · M
docking energies

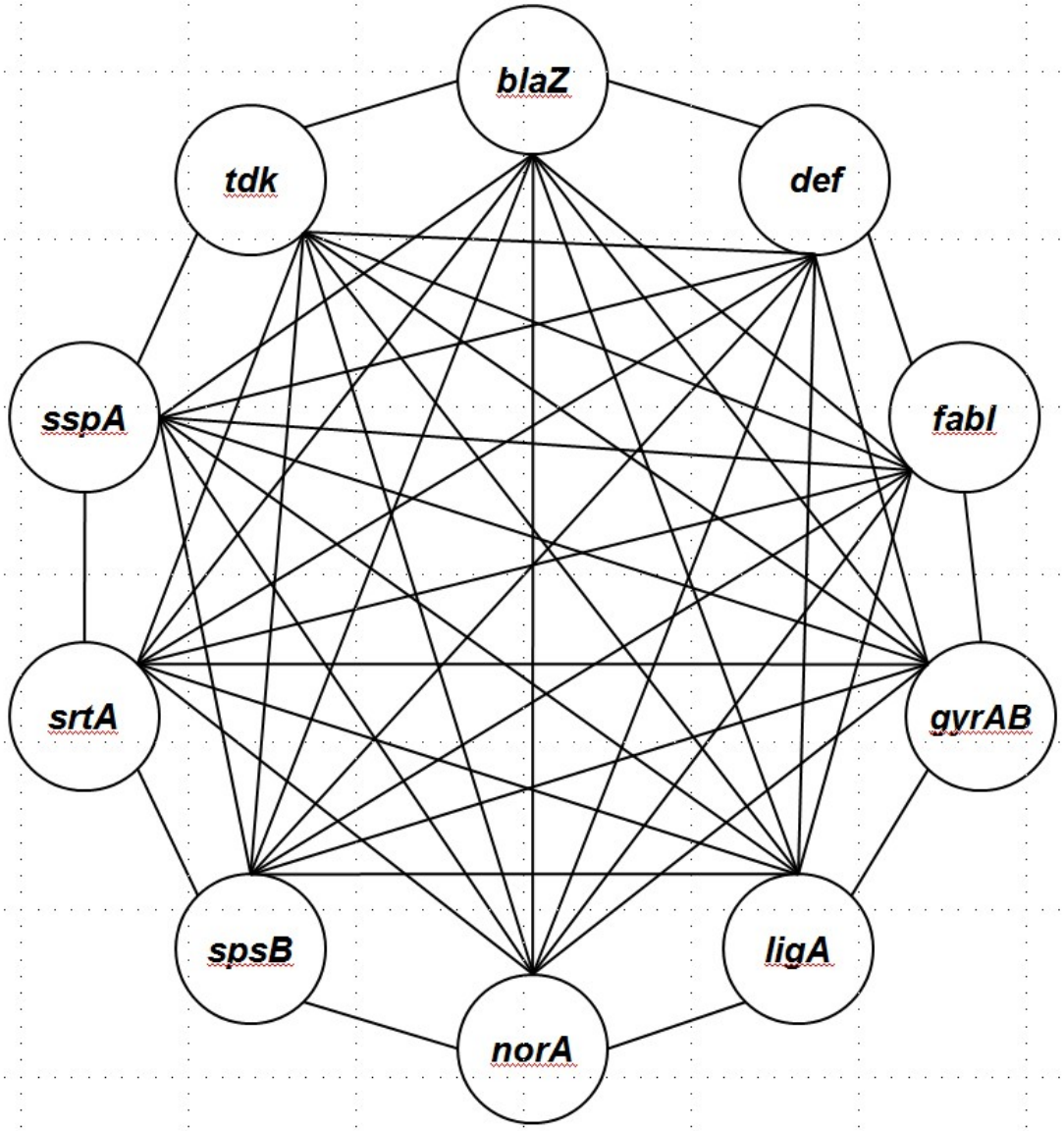
Perfilev M. A., Vassiliev P. M., et al. Russian Chemical Bulletin, 2023; 72(4):1075-1082.

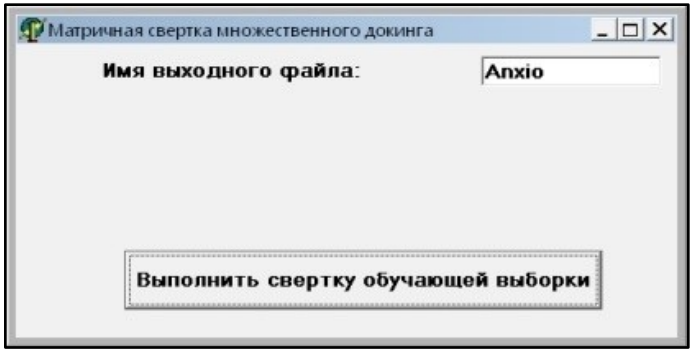
Fully-connected convolutional neural network

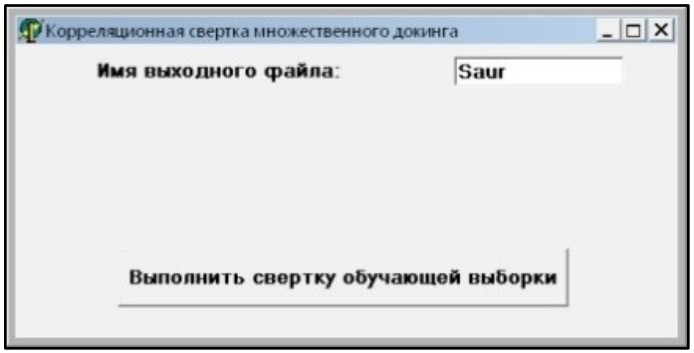


Vassiliev P. M., et al. Journal of Ural Medical Academic Science, 2025; 22(3):279-293.

Modular fully-connected convolutional neural network







Matrix convolution

Minimum docking energy for one biotarget

$$\Delta E_{li} = \min_{j=1}^{5} \left(\min_{k=1}^{10} \left(\Delta E_{lijk} \right) \right), \; l = 1..L, \; i = 1...27$$

Matrix of distances between docking energies

$$\mathbf{D_l} = \{ D_{lij} \} = \{ | \Delta E_{li} - \Delta E_{lj} | \}, \; i, j = 1...27, \; i \neq j, \; l = 1..L$$

Energy of fully-connected neural network

$$W_l = det \, \mathbf{D_l}, \; l = 1..L$$

L – number of compounds

Correlation convolution

Minimum docking energy for one biotarget

$$\Delta E_{lmi} = \min_{j=1}^{5} \left(\min_{k=1}^{10} \left(\Delta E_{lmijk} \right) \right), \; l = 1..L, \; m = 1..M, \; i = 1...27$$

Energy of fully-connected neural network

$$W_{lm} = \frac{1}{2} \cdot \frac{1}{27} \sum_{\substack{i,j=1 \\ i \neq j}}^{27} R_{mij} \cdot \Delta E_{lmi} \cdot \Delta E_{lmj}, \; l = 1..L, \; m = 1..M$$

Energy of modular fully-connected neural network

$$V_l = \frac{1}{2} \cdot \frac{1}{M} \sum_{\substack{i,j=1 \\ i \neq j}}^{M} R'_{ij} \cdot W_{li} \cdot W_{lj}, \; l = 1..M$$

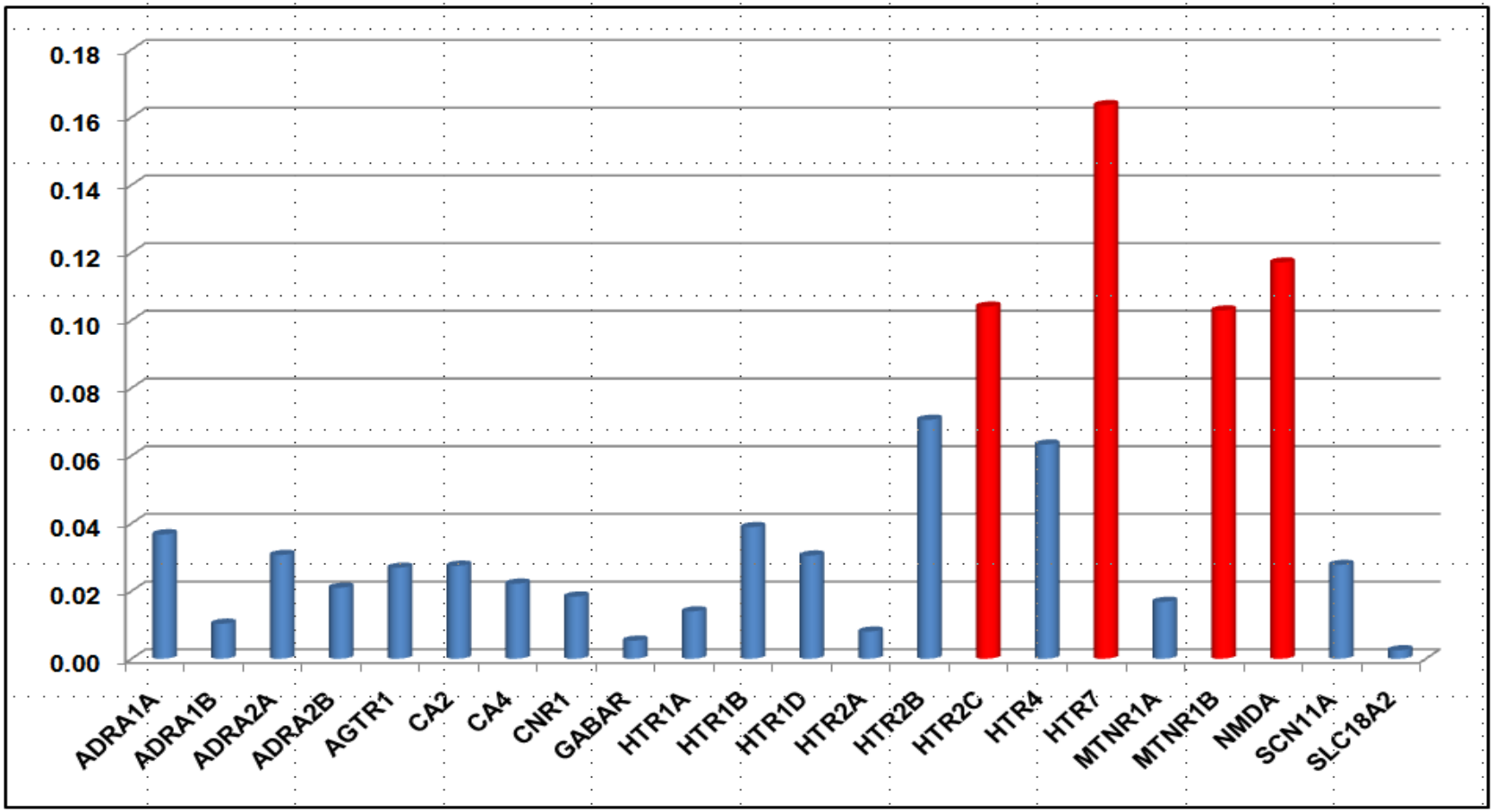
L – number of compounds; *M* – number of biotargets

Vassiliev P. M., et al. Journal of Volgograd State Medical University, 2024; 21(4):136-142.

Fully-connected convolutional neural networks

Anxiolytic activity

Matrix convolutions of W on 27 multiple docking energies
Goodman-Kruskal correlation coefficient $R_\gamma = 0.164$
Statistical significance $p < 5 \cdot 10^{-7}$



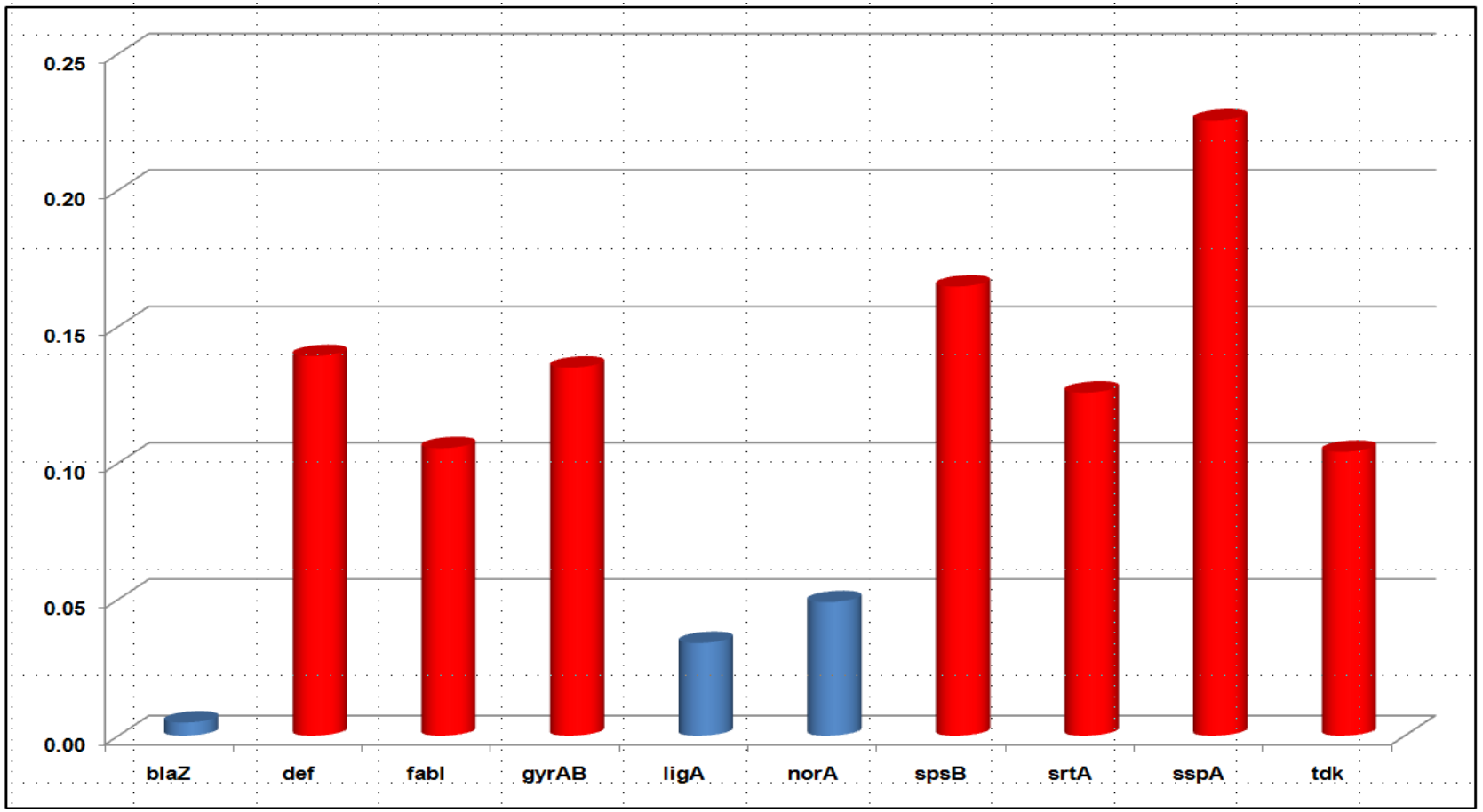
Fully-connected convolutional neural networks

Antimicrobial *S. aureus* activity

Matrix convolutions of *W* on 27 multiple docking energies

Goodman-Kruskal correlation coefficient $R_\gamma = 0.226$

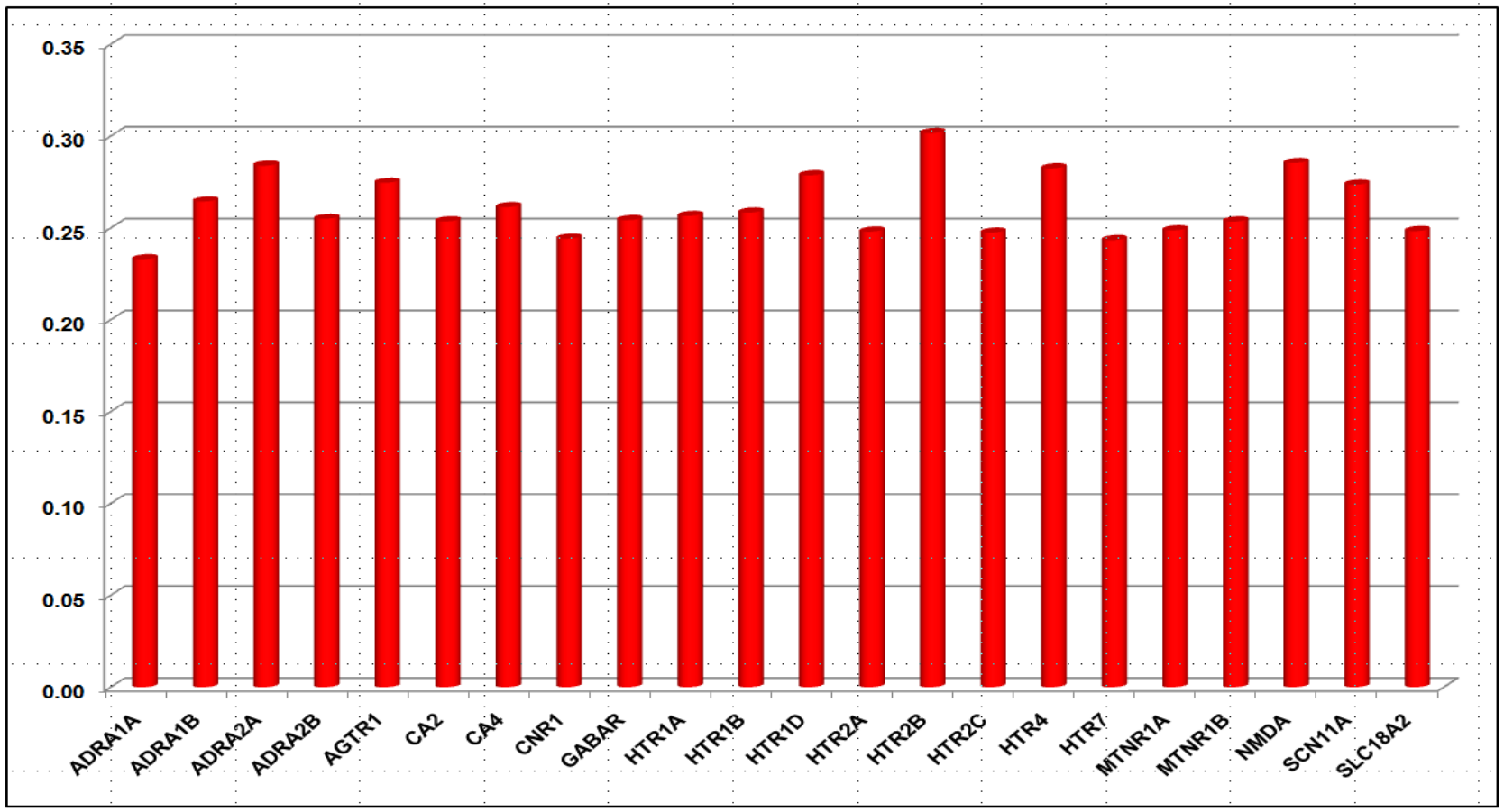
Statistical significance $p < 5 \cdot 10^{-7}$



Fully-connected convolutional neural networks

Anxiolytic activity

Correlation convolutions of W on 27 multiple docking energies
Goodman-Kruskal correlation coefficient $R_\gamma = 0.302$
Statistical significance $p < 5 \cdot 10^{-7}$



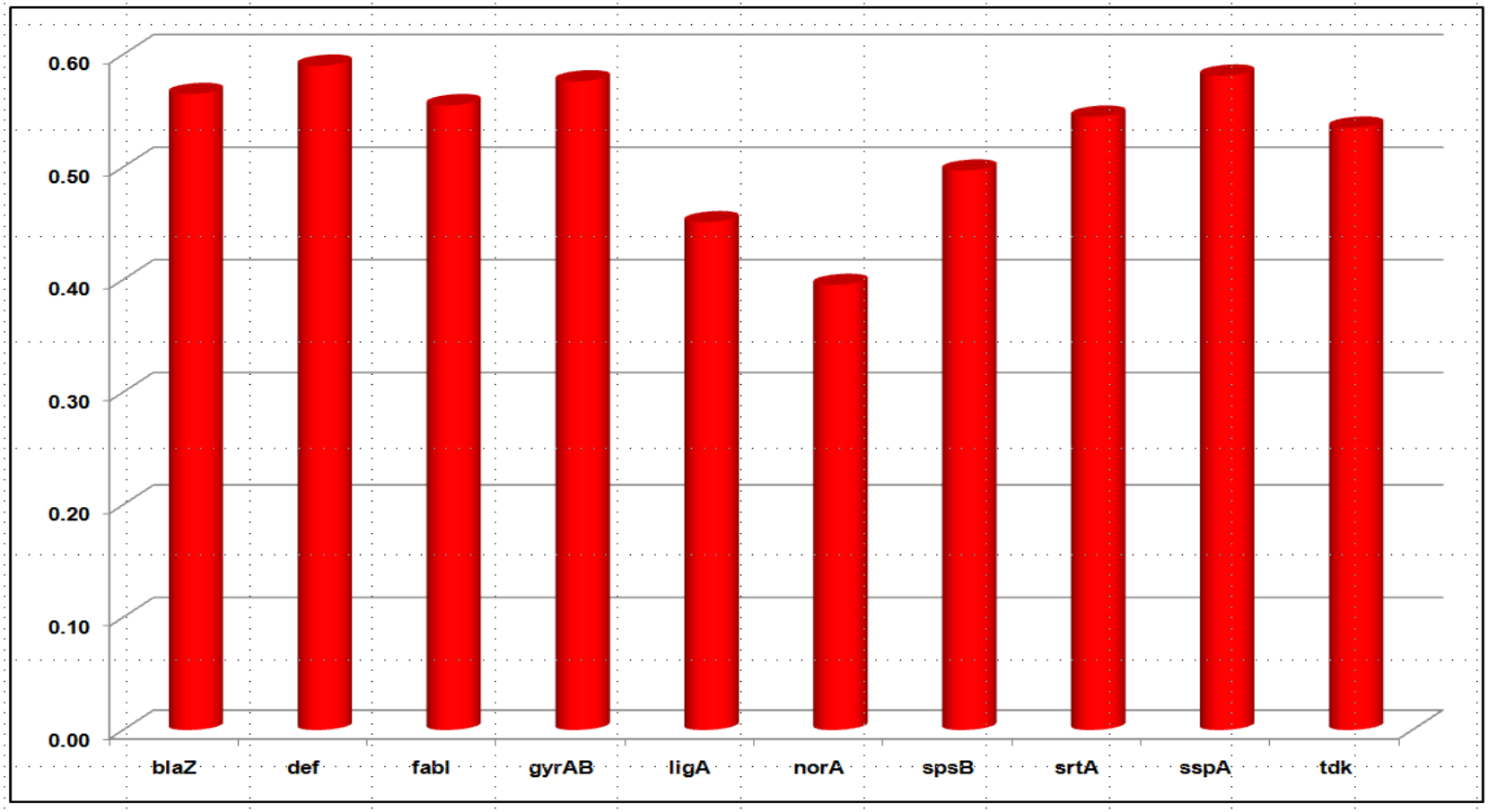
Fully-connected convolutional neural networks

Antimicrobial *S. aureus* activity

Correlation convolutions of W on 27 multiple docking energies

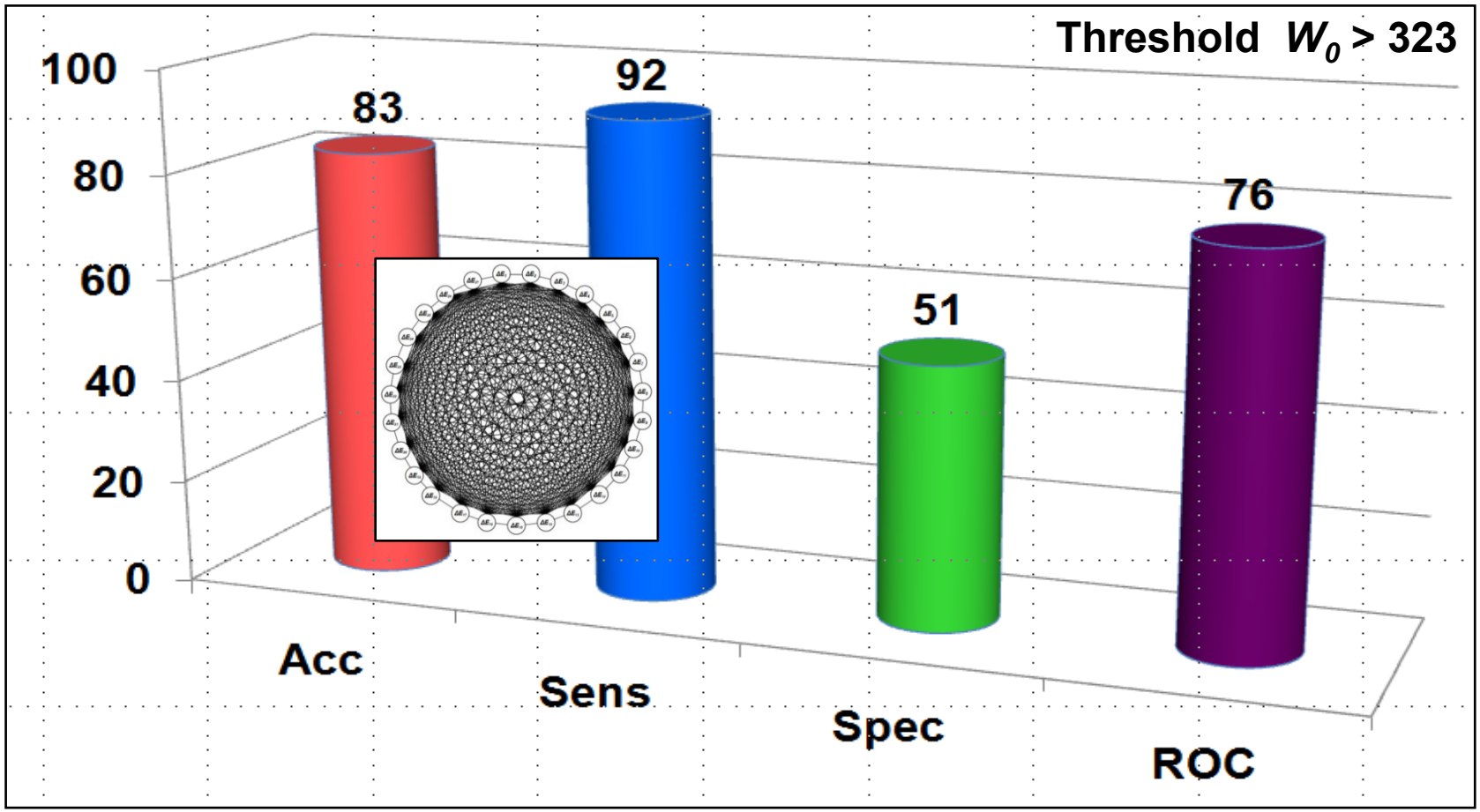
Goodman-Kruskal correlation coefficient $R_\gamma = 0.590$

Statistical significance $p < 5 \cdot 10^{-7}$



Fully-connected convolutional neural network

Antimicrobial *S. aureus* activity – def biotarget
Correlation convolutions of *W* on 27 multiple docking energies
Goodman-Kruskal correlation coefficient $R_\gamma = 0.632$
Statistical significance $p = 2.8 \cdot 10^{-12}$



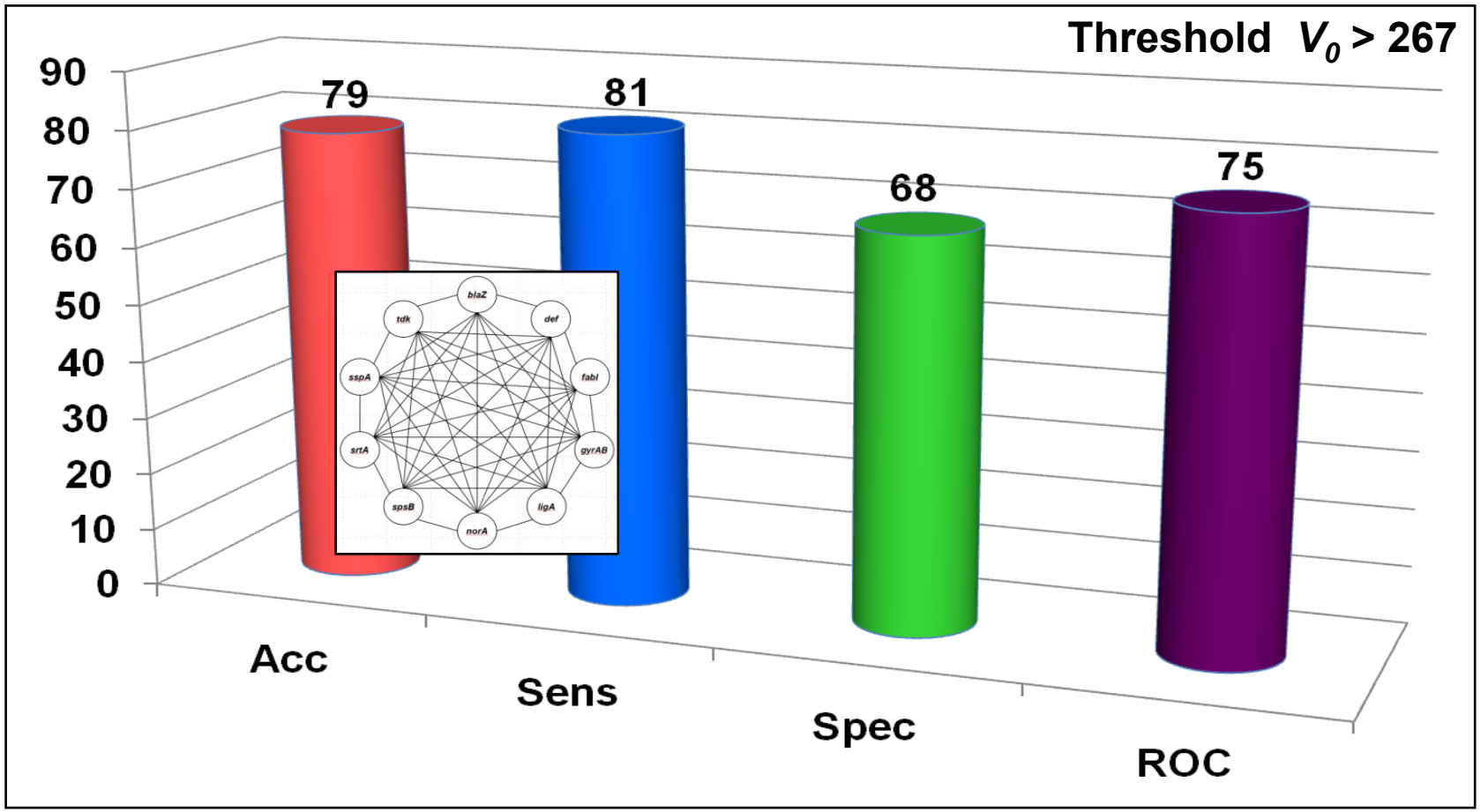
Modular fully-connected convolutional neural network

Antimicrobial *S. aureus* activity

Correlation convolution of V on 10 energies W of neural networks

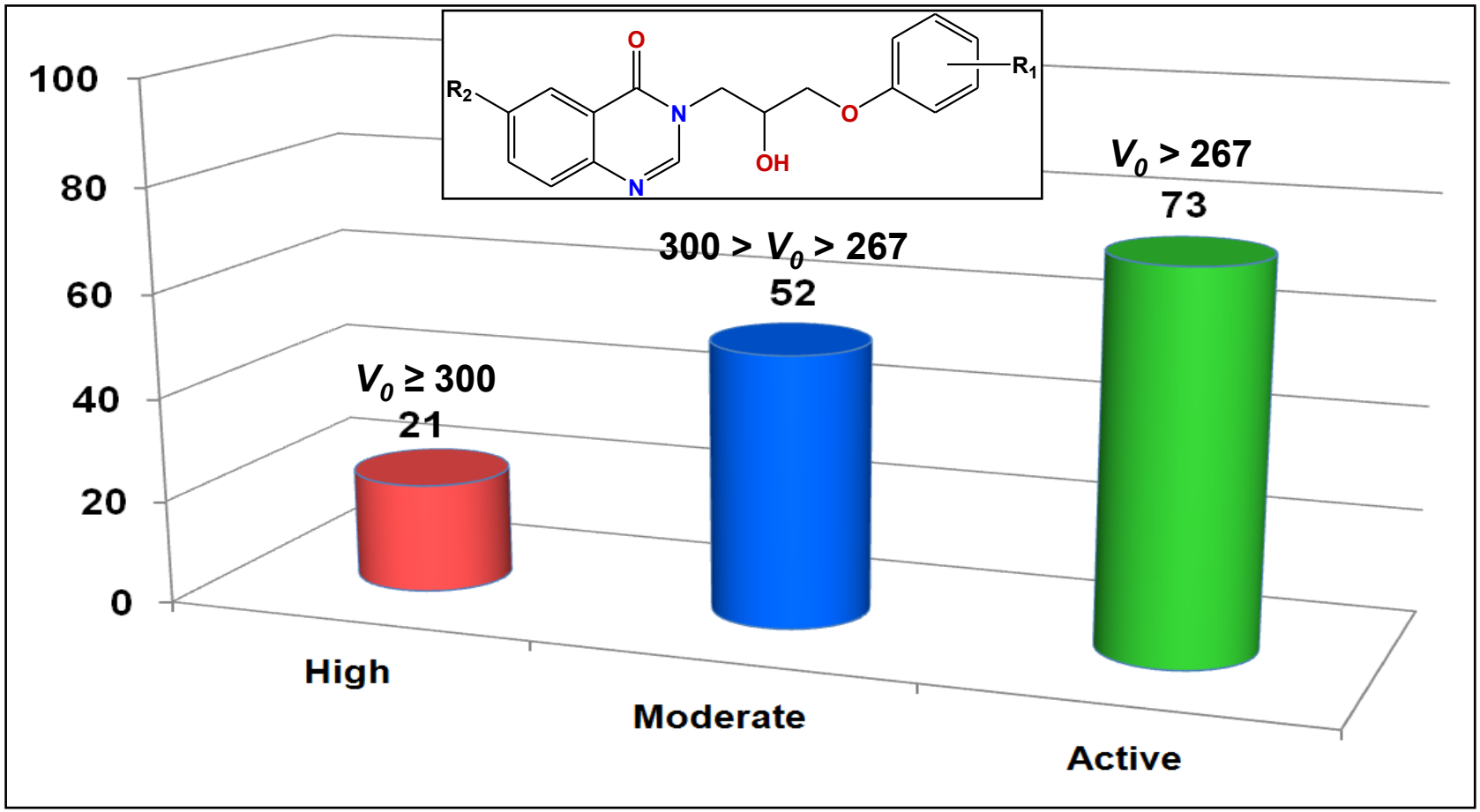
Goodman-Kruskal correlation coefficient $R_Y = 0.805$

Statistical significance $p = 1.4 \cdot 10^{-9}$



Virtual screening

Antimicrobial *S. aureus* activity
71 quinazoline-4(3*H*)-one derivatives
52 compounds with expressed activity
Enrichment coefficient *EnrCoef* = 1.37



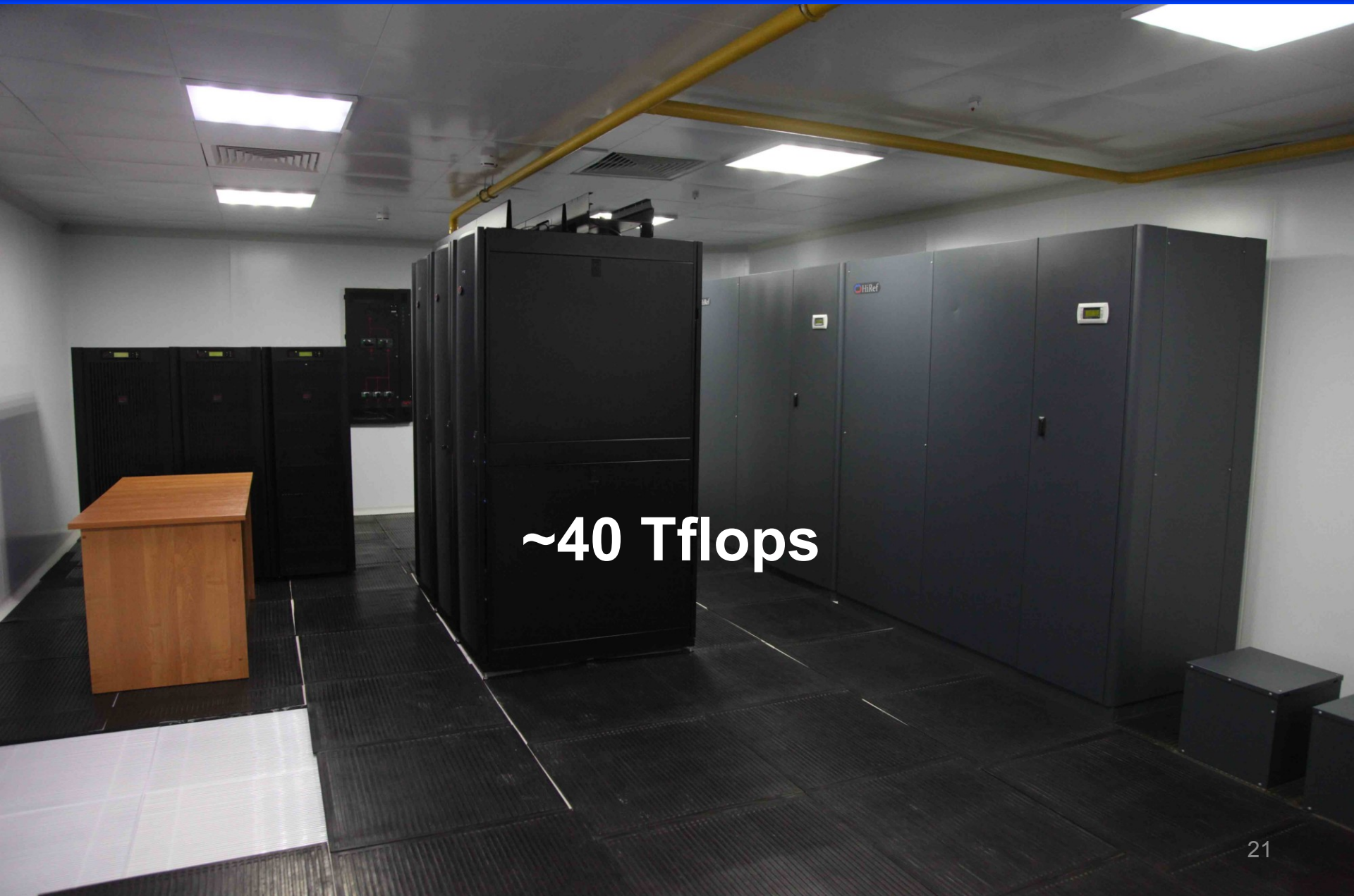
Conclusions

A novel architecture of modular multi-target fully-connected convolutional neural network based on various types of convolution of energy spectra of multiple docking was developed.

The effectiveness of this new machine learning method to in silico search for biological active compounds was demonstrated.

Research Center of Innovative Medicines

Laboratory for Information Technology in Pharmacology and Computer Modeling of Drugs



Creative team

pvassiliev@mail.ru

P.M. Vassiliev

A.V. Golubeva

M.A. Perfilev

A.N. Kochetkov

Acknowledgments

This study was supported in the framework of the state assignment of the Ministry of Health of the Russian Federation No. 23022400009-9

Thank You for Your attention!