ChémInformatique STRASBOURG

Chemoinformatics and molecular modeling lab.

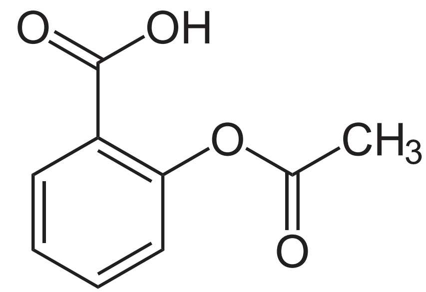
INSTITUTE OF MOLECULAR AND TRANSLATIONAL MEDICINE

XXVII Symposium on Bioinformatics and Computer-Aided Drug Discovery

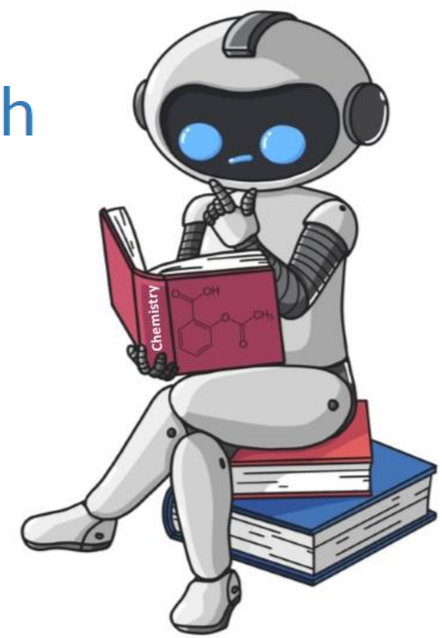
МНОГОВАРИАНТНОЕ ОБУЧЕНИЕ – НОВЫЙ МЕТОД МОДЕЛИРОВАНИЯ «СТРУКТУРА-СВОЙСТВО»

Dr. Timur I. Madzhidov
Kazan Federal University, Department of Organic Chemistry

tmadzhidov@gmail.com

Bioactivity = f ()

Chembench

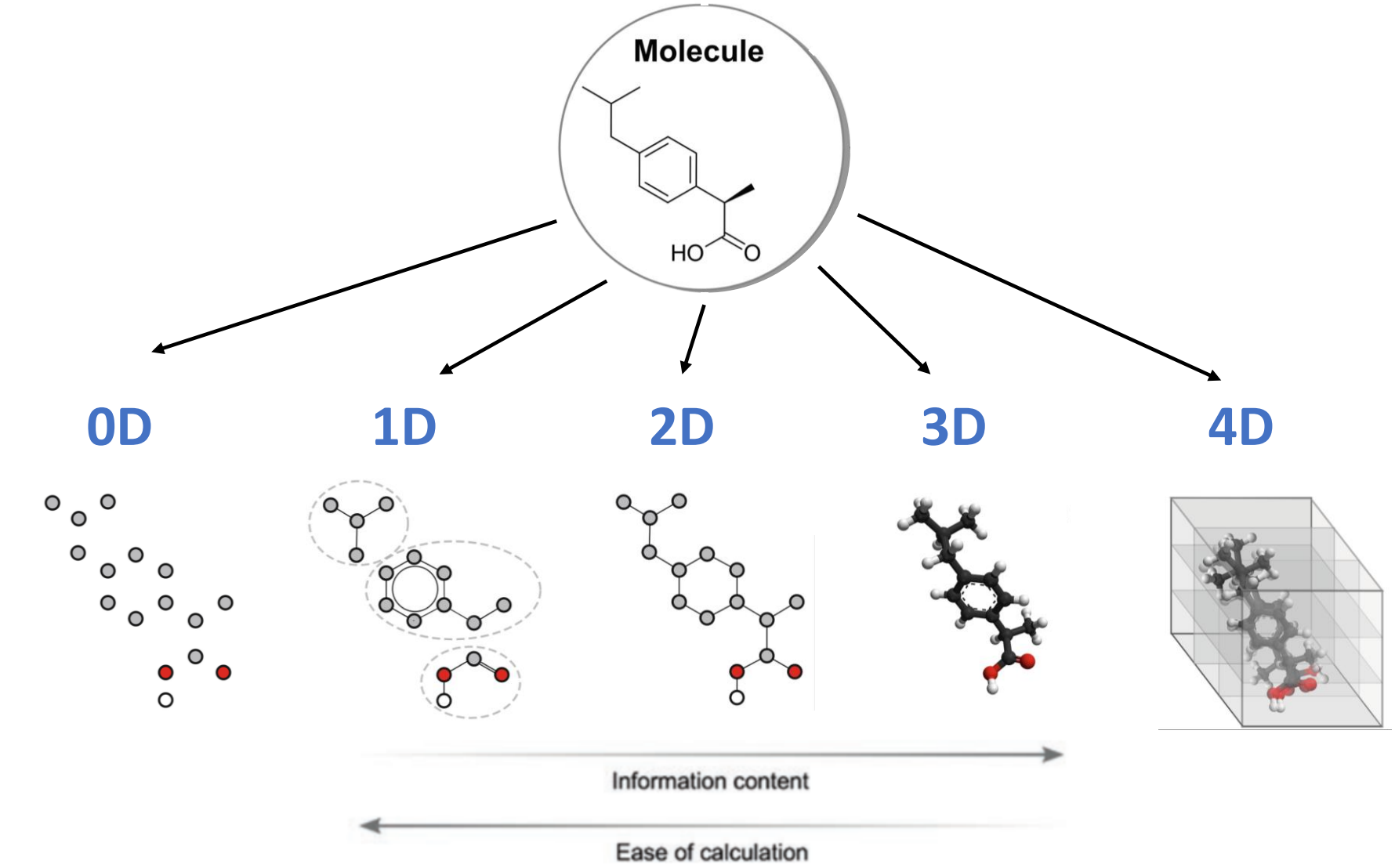


SIB Swiss Institute of Bioinformatics

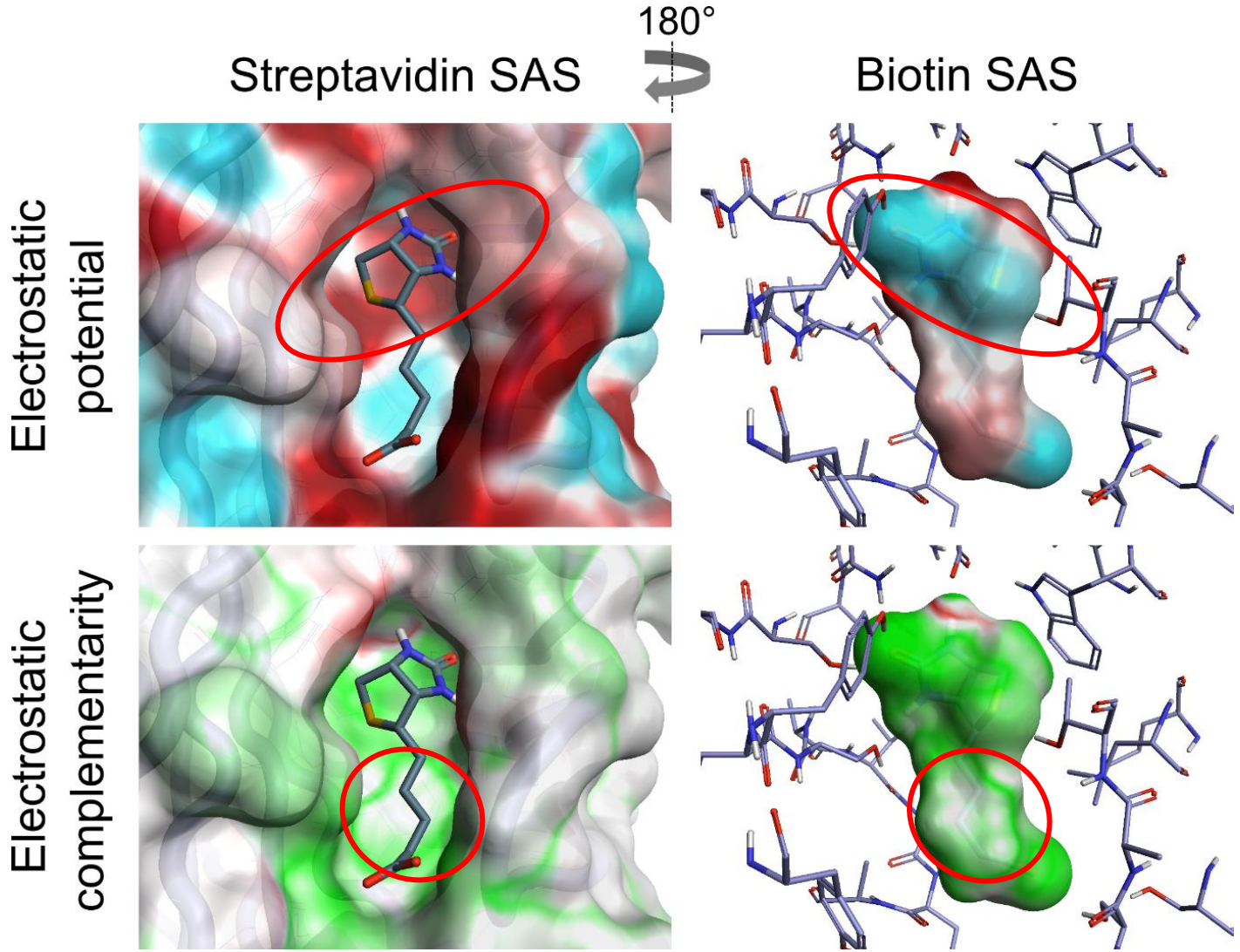
Online chemical database with modeling environment

PASS online
It is easy to use

Molecule Representations

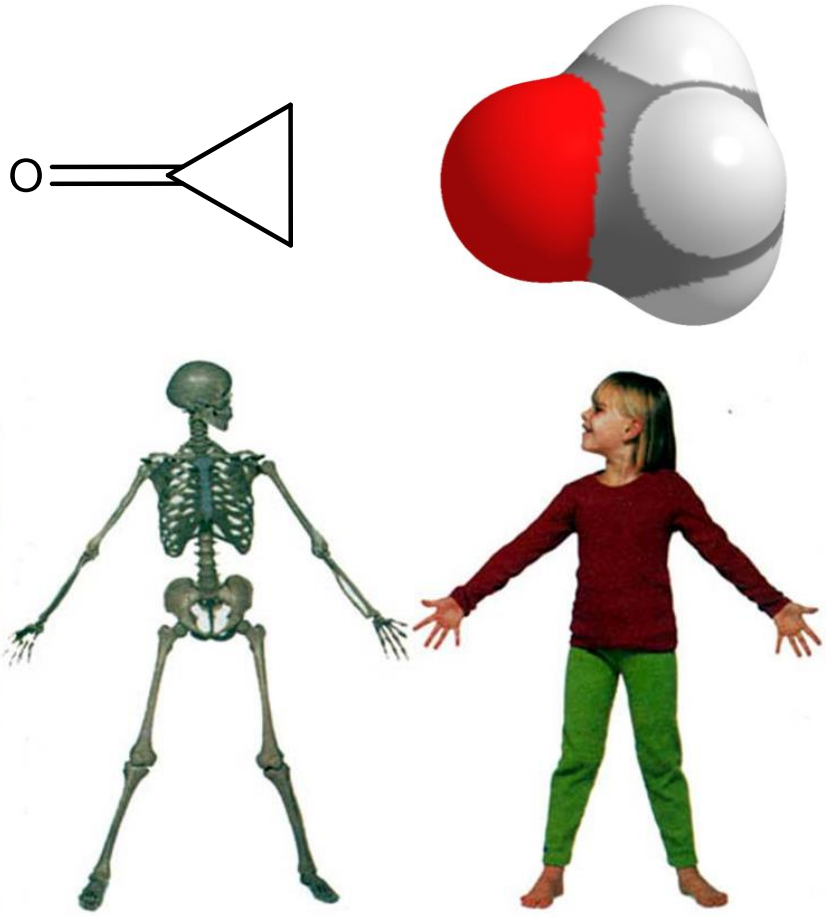


Binding


180°
Streptavidin SAS
Biotin SAS
Electrostatic potential
Electrostatic complementarity

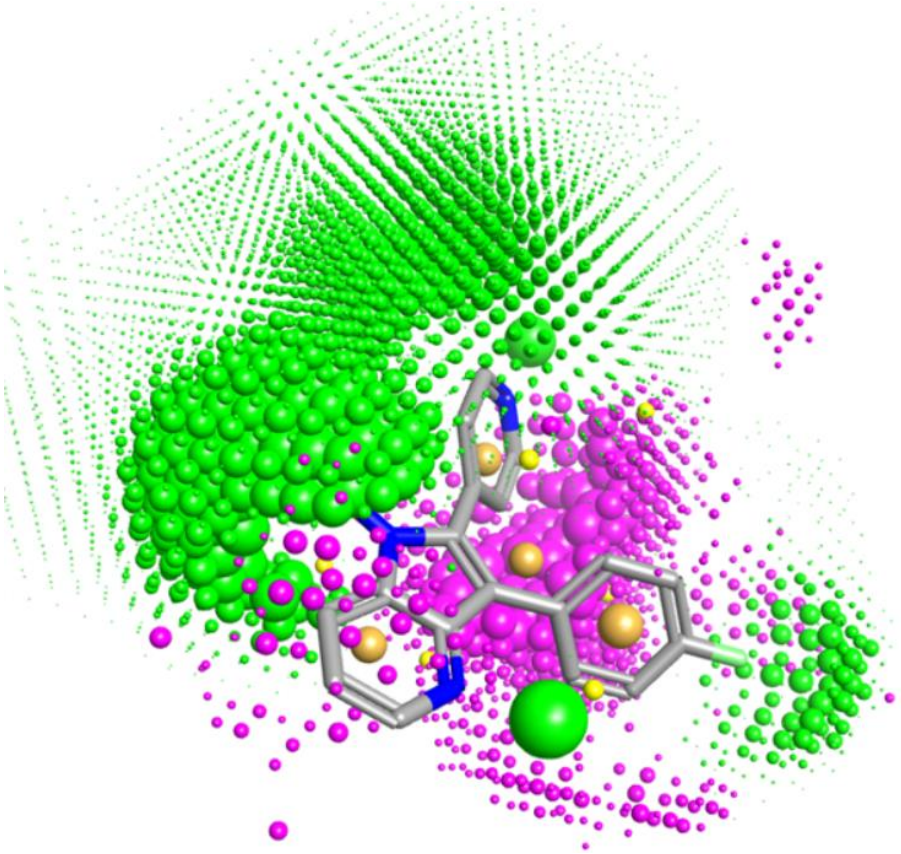
Graph is not enough

Molecular graph reflects only skeleton of the molecule
But real interactions between molecules go through their surface!

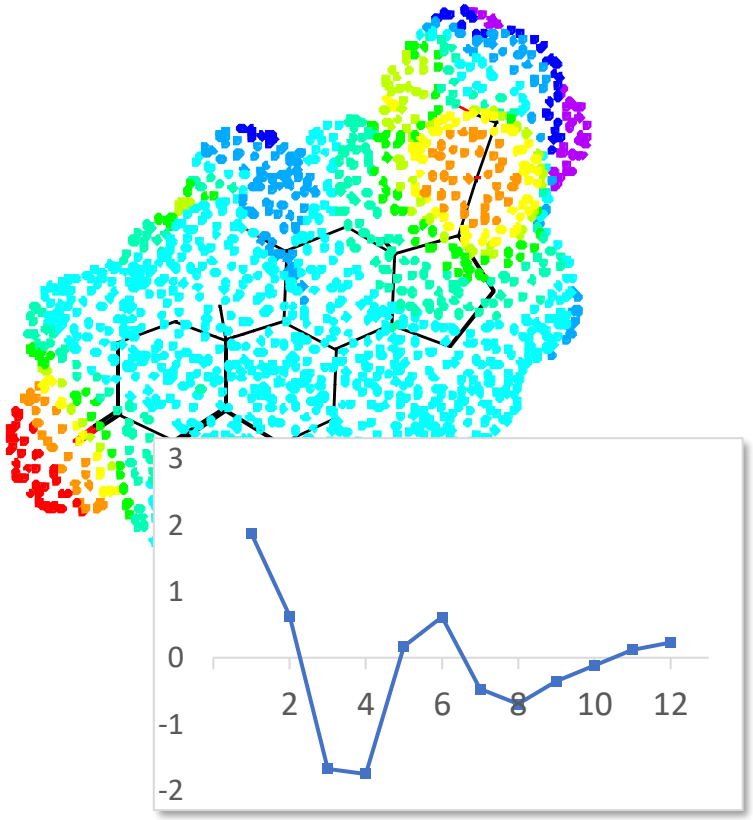


3D-QSAR

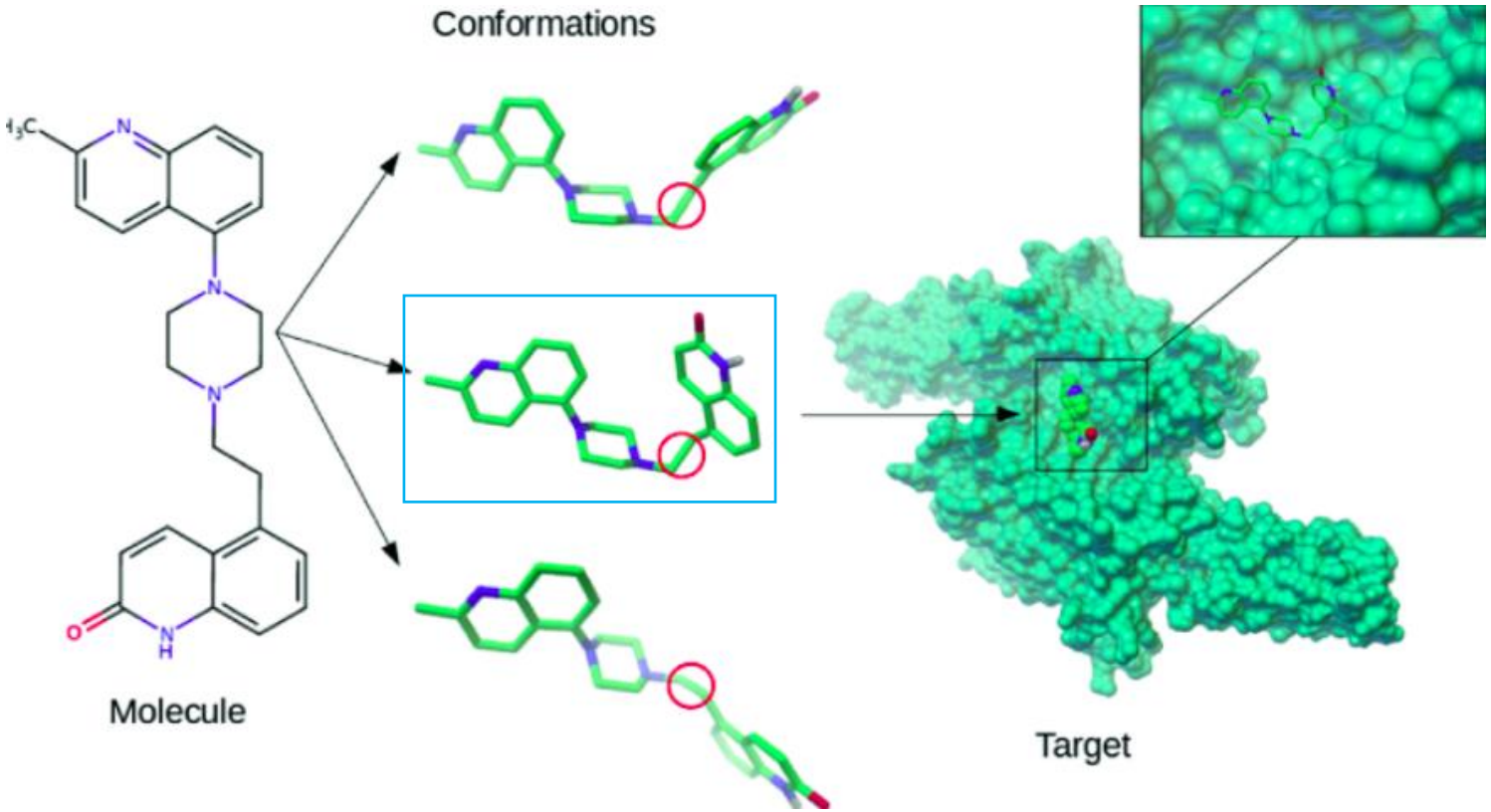
Alignment dependent 3D-QSAR



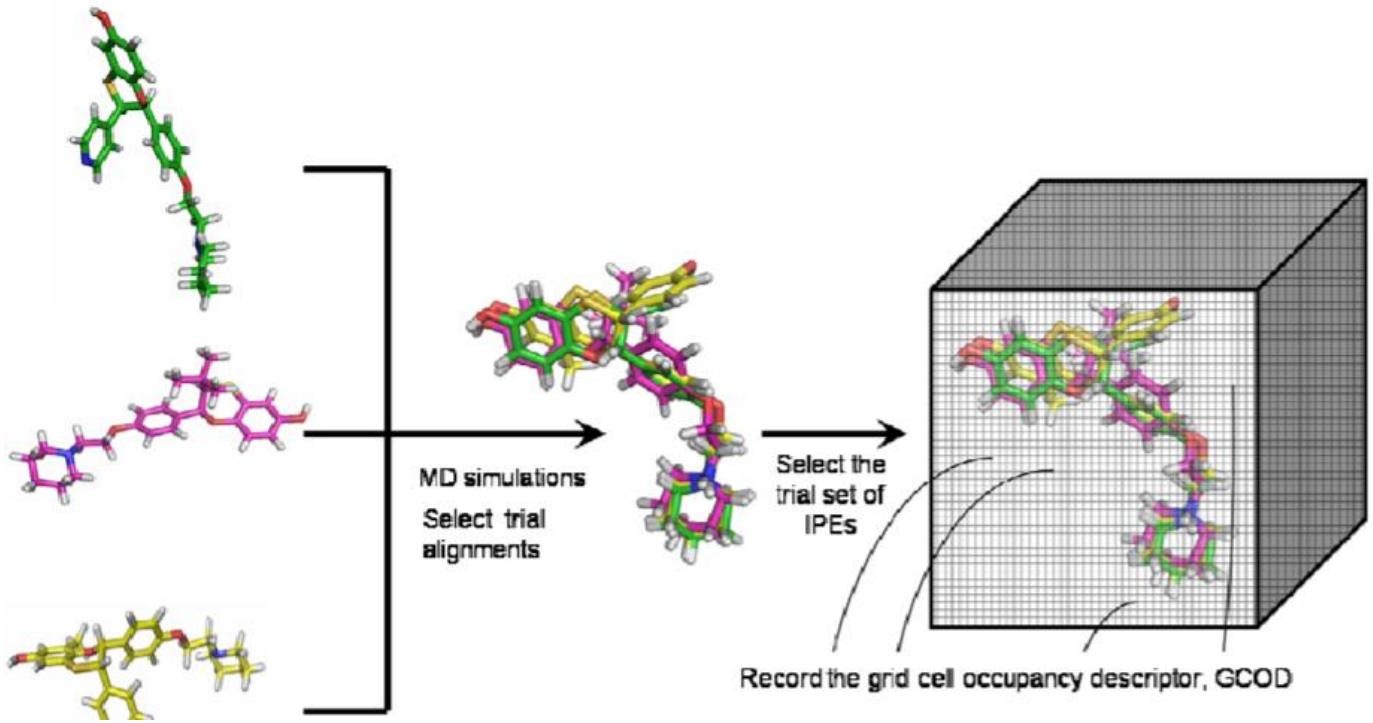
Alignment independent 3D-QSAR



Bioactive conformation selection



4D-QSAR



Andrade, C. H., Pasqualoto, K. F., Ferreira, E. I., & Hopfinger, A. J. (2010). 4D-QSAR: perspectives in drug design. *Molecules*, *15*(5), 3281-3294.

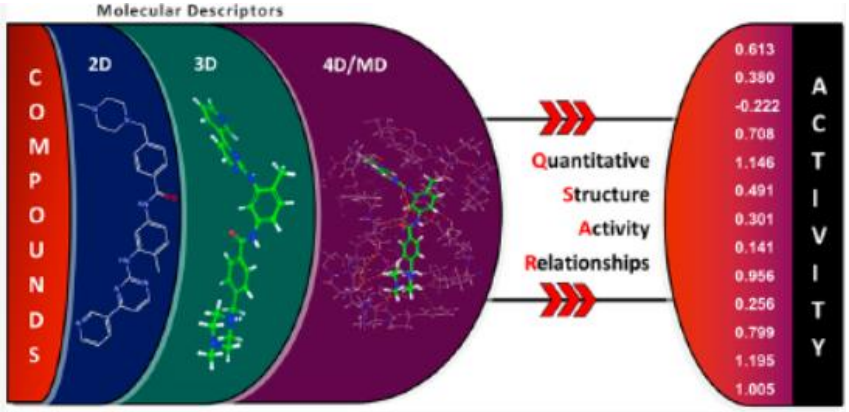
Molecule Representations

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

Benchmarking 2D/3D/MD-QSAR Models for Imatinib Derivatives: How Far Can We Predict?

Phyo Phyo Kyaw Zin, Alexandre Borrel, and Denis Fourches*



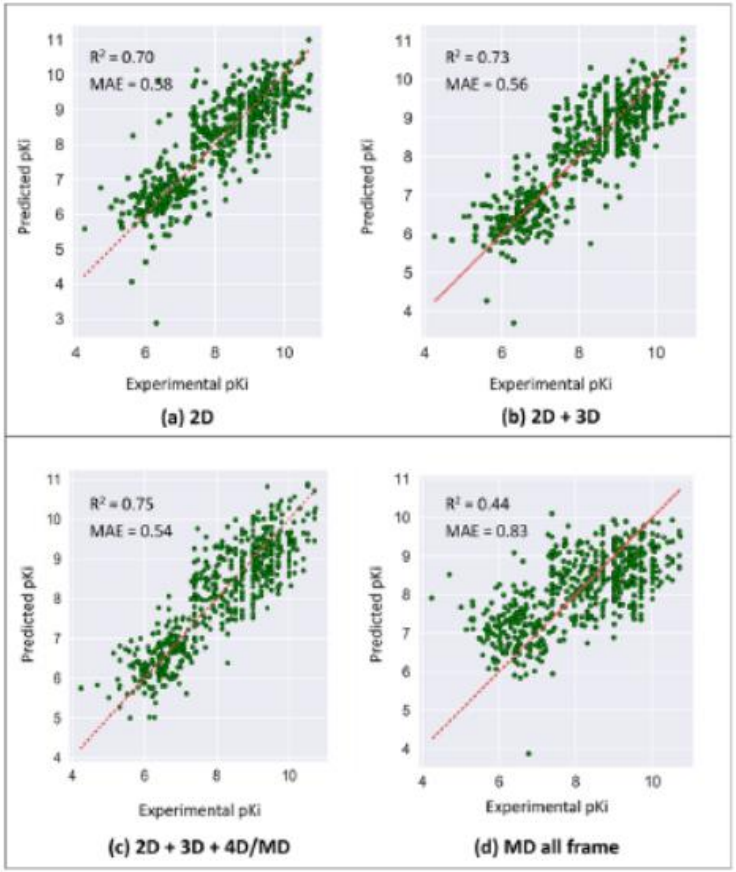
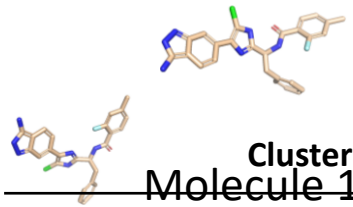
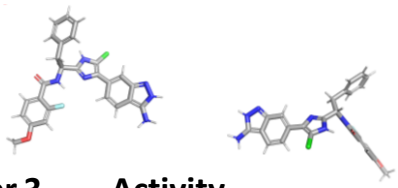
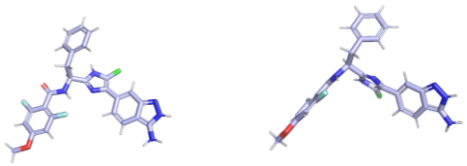


Figure 15. Native 10-fold cross-validation predictions of pKi-reported Imatinib analogues by DNN-based regression models using descriptors of (A) 2D, (B) 2D+3D, (C) 2D+3D+4D/MD, and (D) MD all frame. Red dotted line represents an ideal fit between experimental pKi and predicted pKi.

Clustering to multi-conformation problem solution







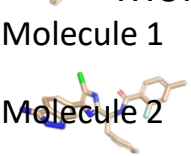
	Cluster 1	Cluster 2	Cluster 3	Activity
Molecule 1	1	1	0	6.70
Molecule 2	1	0	1	6.55
Molecule 3	1	1	0	7.19

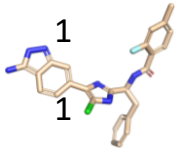
Molecule 1

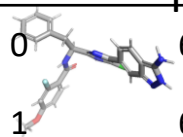
Molecule 2

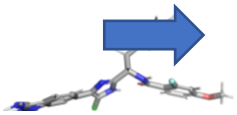
Random Forest

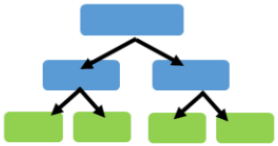
Molecule 3

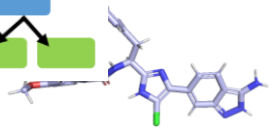










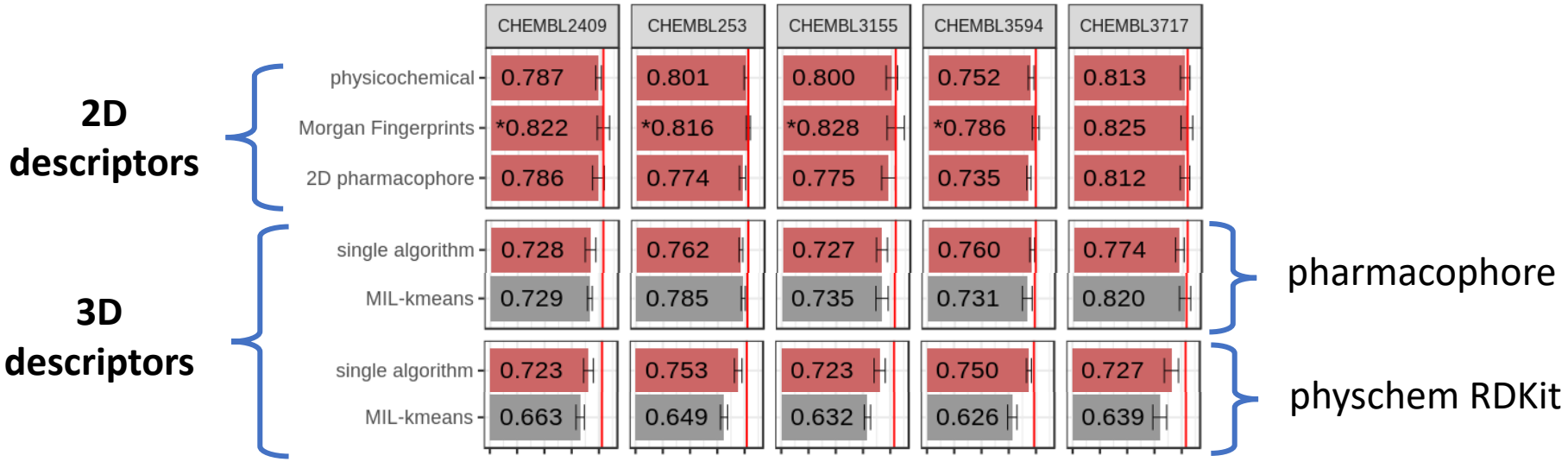


Cluster 1

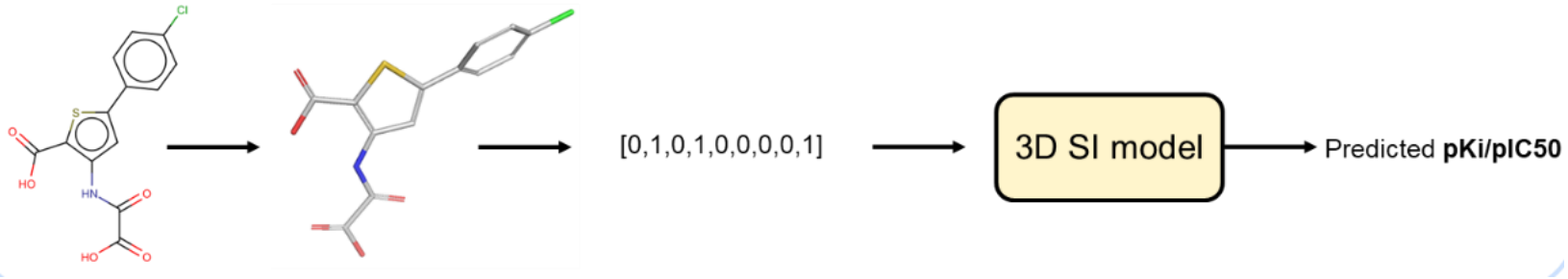
Cluster 2

Cluster 3

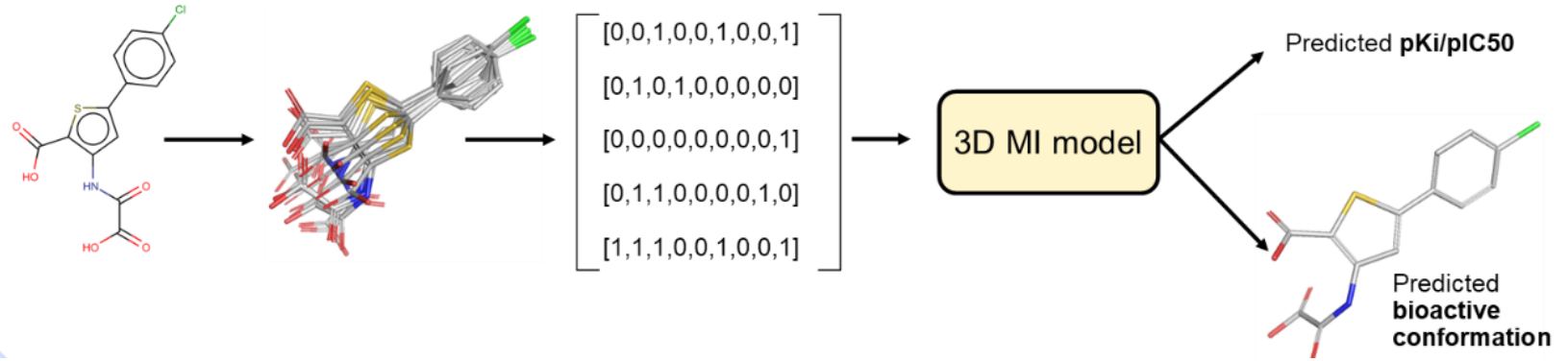
Clustering has contradictory results



Single-Instance QSAR

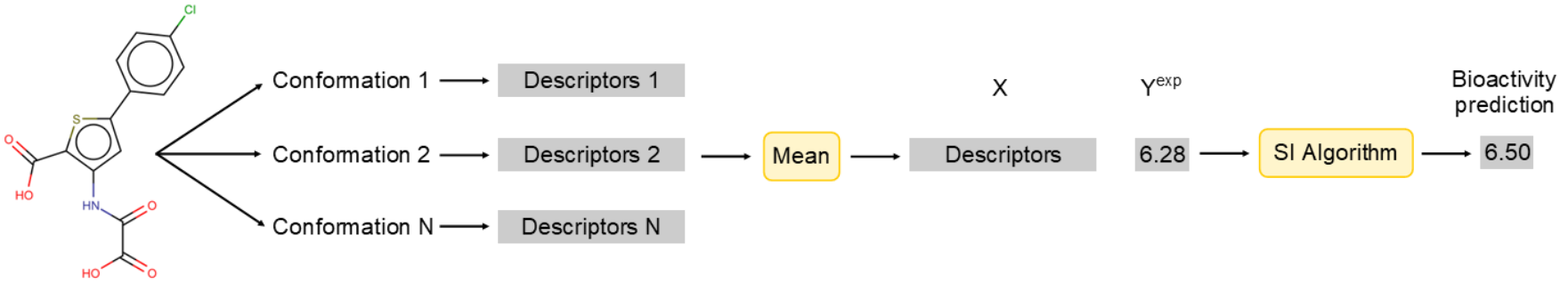


Multi-Instance QSAR

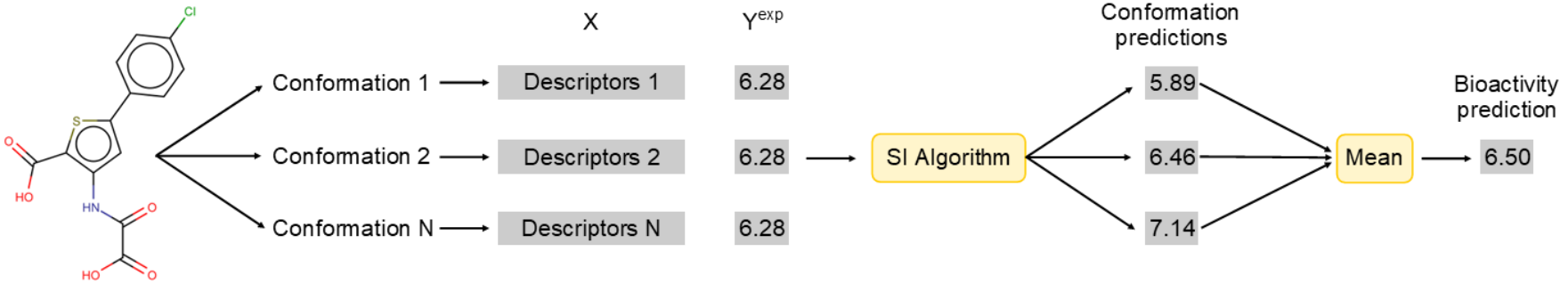


Multi-Instance Algorithms: Intuitive approaches

Bag-Wrapper = Averaging conformation **descriptors**

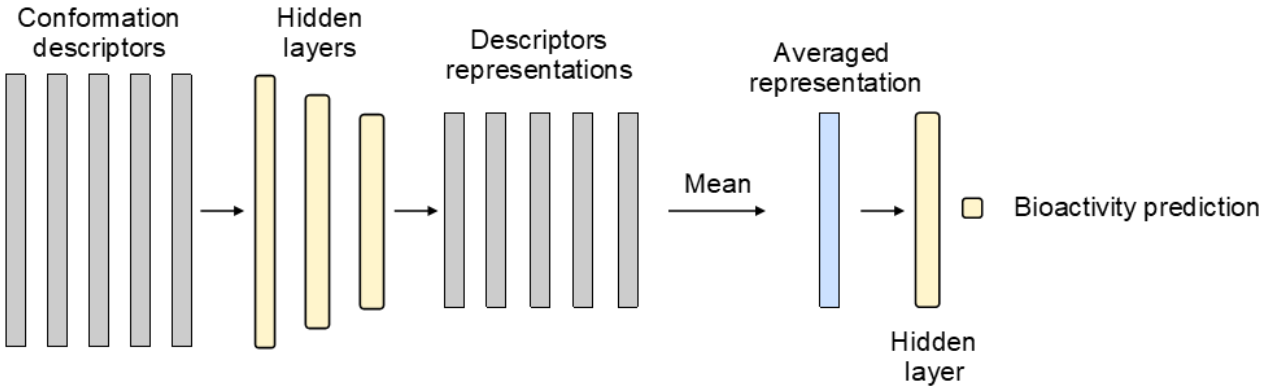


Instance-Wrapper = Averaging conformation **predictions**

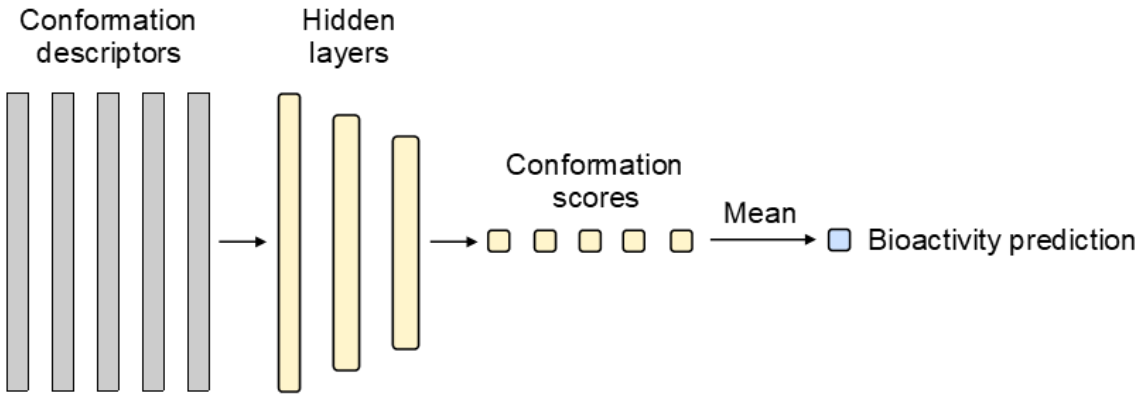


Multi-Instance Algorithms: Neural net approaches

Bag-Net = Averaging conformation **embeddings**

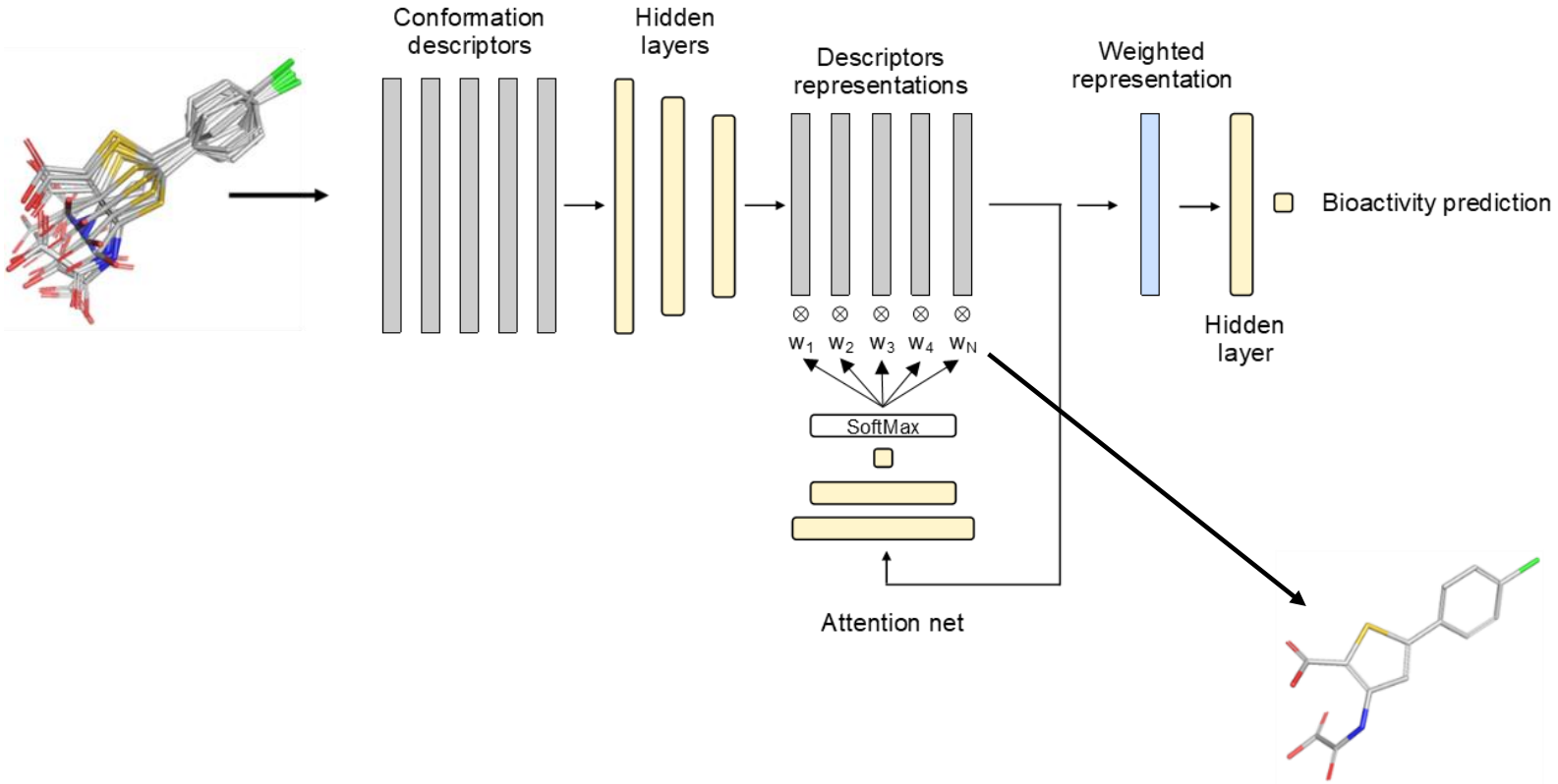


Instance-net = Averaging conformation **scores**



Multi-Instance Algorithms: Kill two birds with one stone

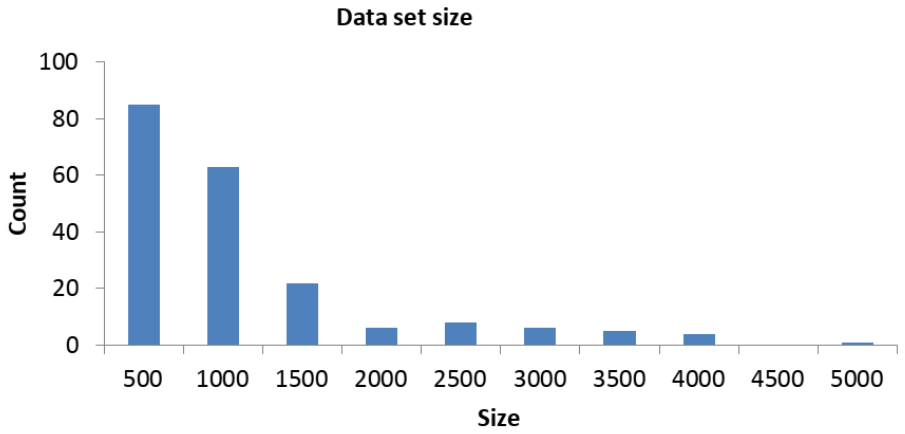
Bag-Attention Net = Weighted averaging conformation **embeddings**

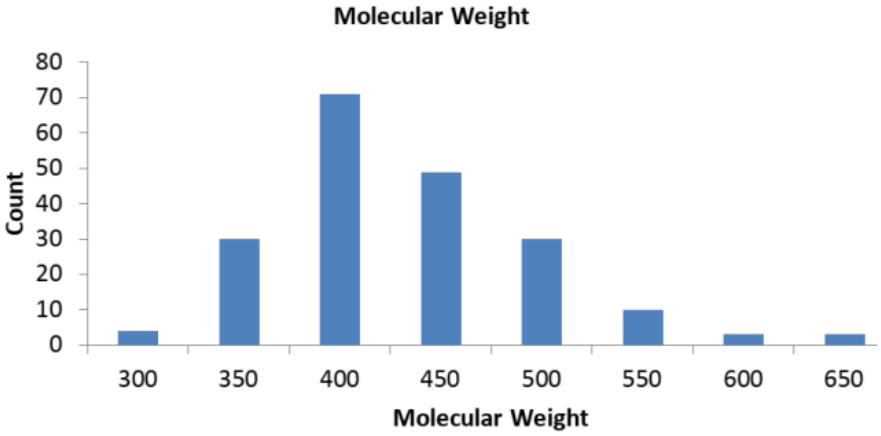


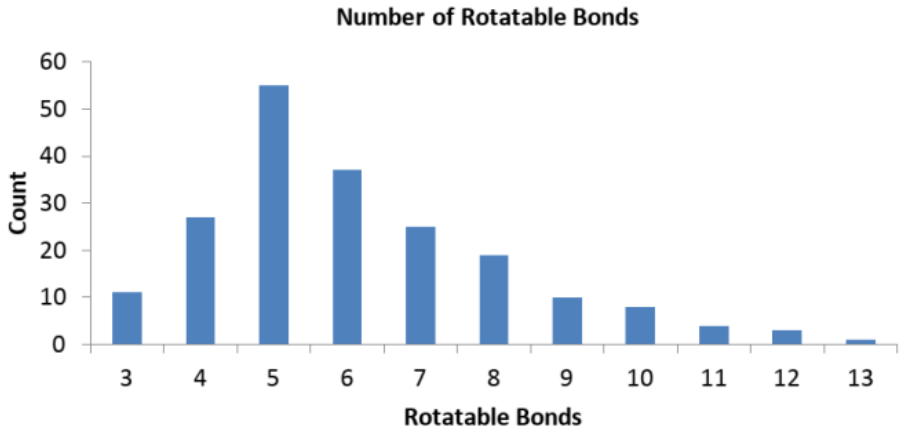
Data Sets

We extracted **175** sets from the CHEMBL23 database

CHEMBL ID	SMILES	pKi
CHEMBL15789	C=CC[C@H](NC(=O)[C@@H]1CCCN1C(=O)[C@@H](NC(=O)...	8.40
CHEMBL82121	O=C(O)CC(c1ccccc1)c1cc(I)c(O)c(I)c1	4.00
CHEMBL83092	O=C(O)CCc1cc(I)c(O)c(I)c1	4.10
CHEMBL189314	CC(C)(C)OC(=O)c1ccccc1C(=O)c1ccc(OCCCOC(=O)N[C...	6.00
CHEMBL82941	O=C(N[C@@H](Cc1cc(I)c(O)c(I)c1)C(=O)O)OCCCOc1c...	6.10
...	...	...
CHEMBL3661086	COc1ccc2c(O[C@@H]3C[C@H]4C(=O)N[C@]5(C(=O)NS(=...	10.00
CHEMBL3661087	COc1ccc2c(O[C@@H]3C[C@H]4C(=O)N[C@]5(C(=O)NS(=...	9.52



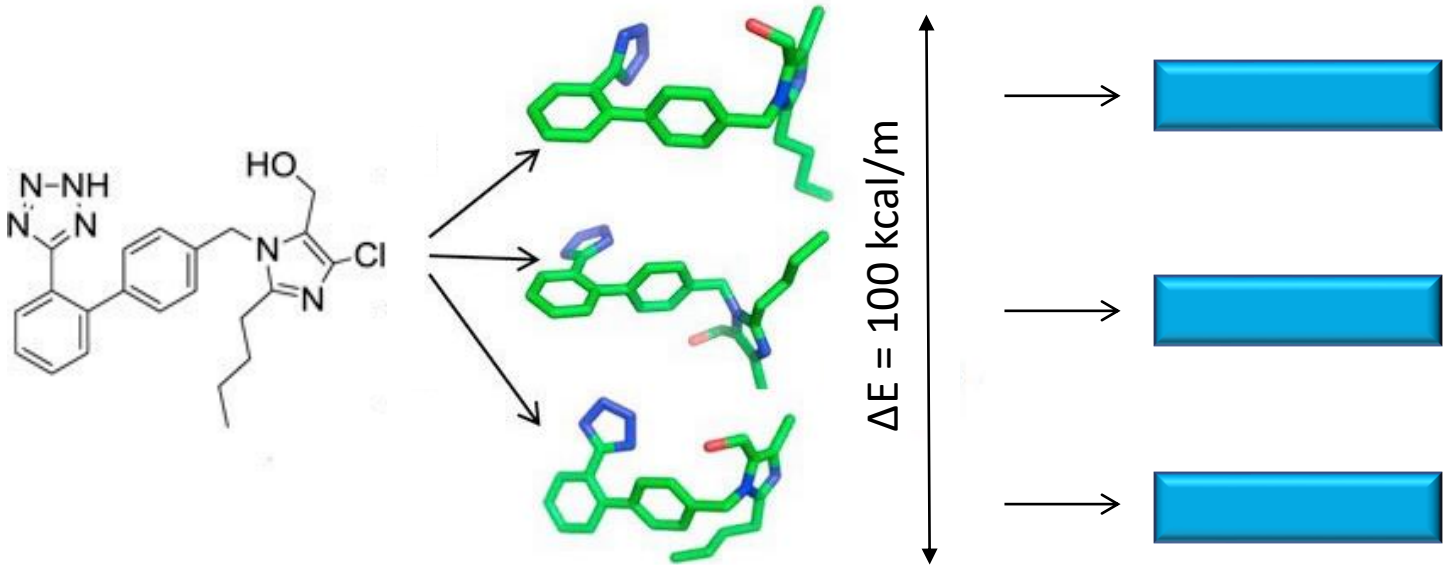




Conformation Generation

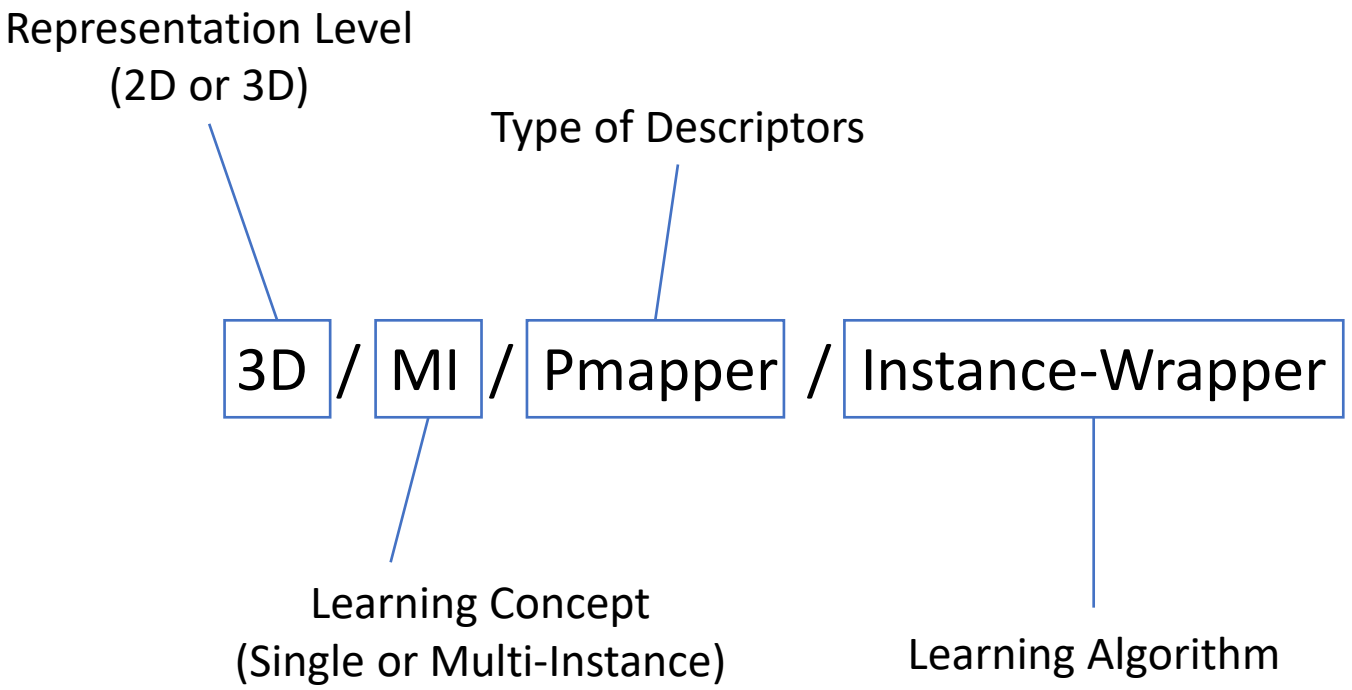
Generation of
100 conformations
per molecule
RDKit

Calculation of
Pmapper* 3D descriptors



*Kutlushina, A., Khakimova, A., Madzhidov, T., & Polishchuk, P. (2018). Ligand-based pharmacophore modeling using novel 3D pharmacophore signatures. *Molecules*, *23*(12), 3094.

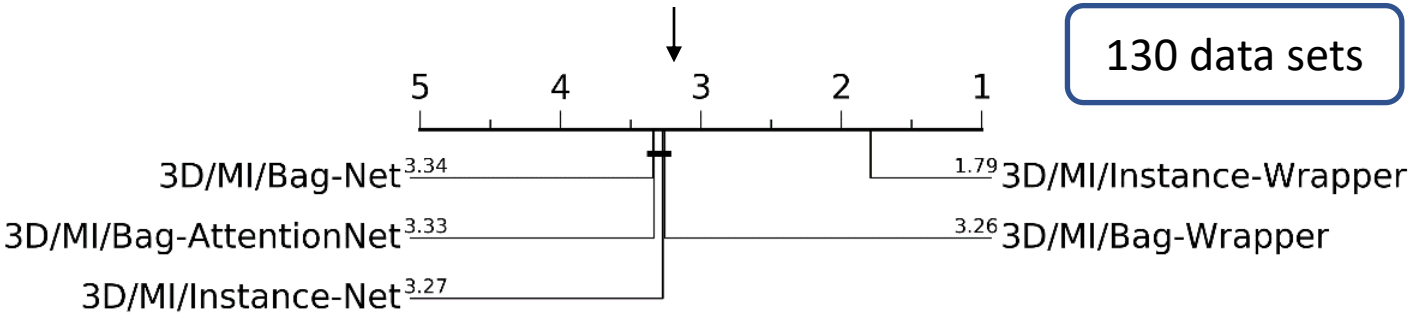
Nomenclature



Benchmarking Multi-Instance Algorithms

	3D/MI/Bag-AttentionNet	3D/MI/Instance-Wrapper	3D/MI/Bag-Wrapper	3D/MI/Bag-Net	3D/MI/Instance-Net
~~CHEMBL1855~~	~~0.248~~	~~0.272~~	~~0.322~~	~~0.261~~	~~0.267~~
CHEMBL2034	0.624	0.664	0.645	0.644	0.638
CHEMBL261	0.313	0.404	0.326	0.322	0.321
CHEMBL322	0.348	0.432	0.357	0.377	0.375
CHEMBL2094122	0.254	0.310	0.420	0.422	0.430
...	...	...	...	...	...
~~CHEMBL238~~	~~0.395~~	~~0.397~~	~~0.367~~	~~0.374~~	~~0.373~~
CHEMBL210	0.535	0.636	0.546	0.548	0.558
CHEMBL3571	0.357	0.420	0.359	0.383	0.379
CHEMBL222	0.418	0.437	0.391	0.400	0.395
CHEMBL1899	0.255	0.660	0.578	0.584	0.582

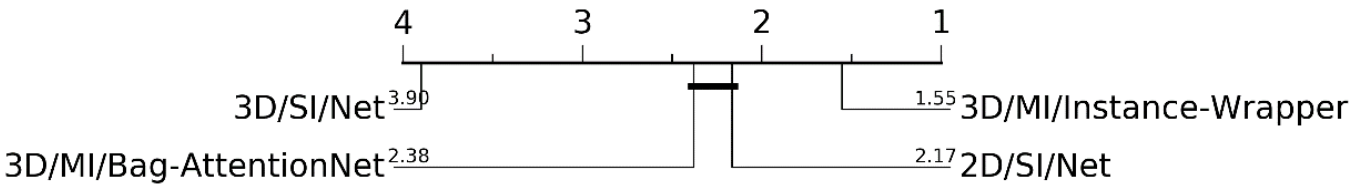
130 data sets



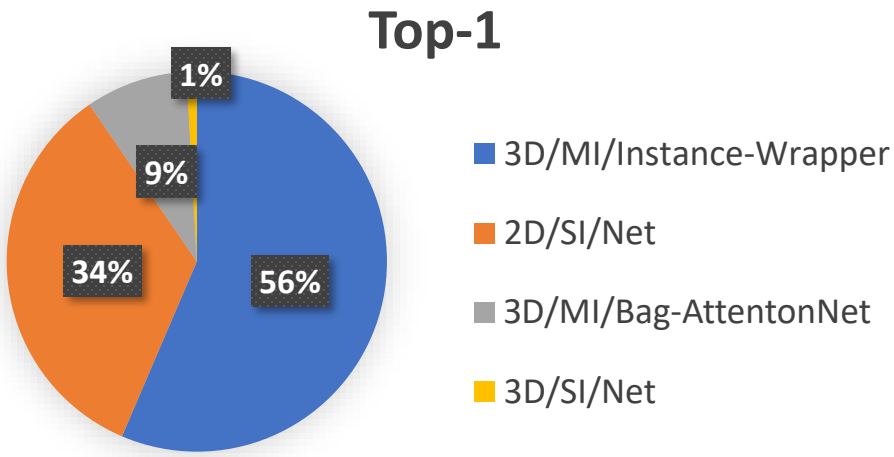
Groups of models that are not significantly different (at a confidence level of 0.05) are connected by the thick line

Comparison of 2D and 3D models

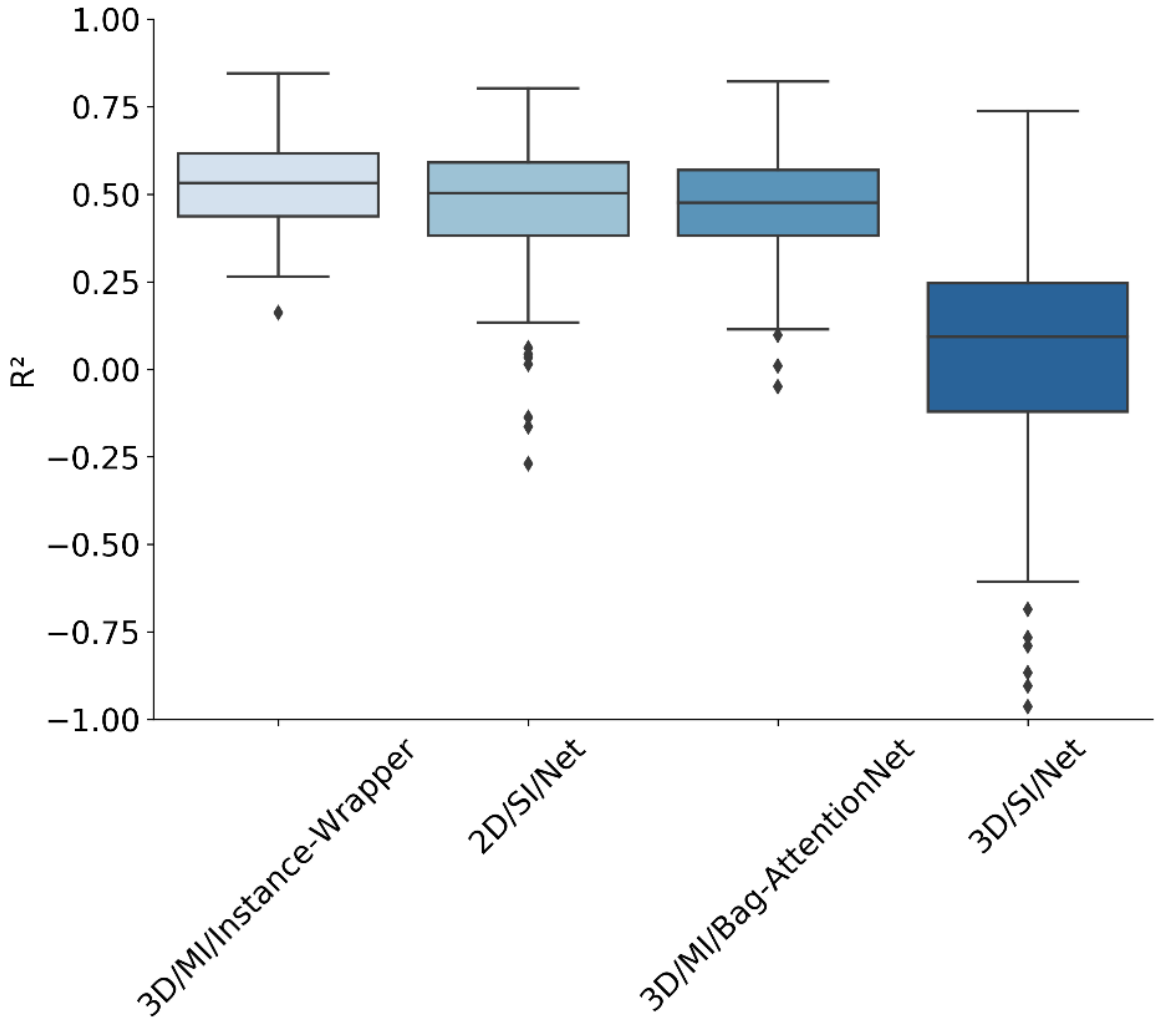
Model	Mean	Median
3D/MI/Instance-Wrapper	0.530 ± 0.123	0.532
2D/SI/Net	0.474 ± 0.189	0.503
3D/MI/Bag-Attention Net	0.474 ± 0.157	0.476
3D/SI/Net	0.027 ± 0.374	0.092



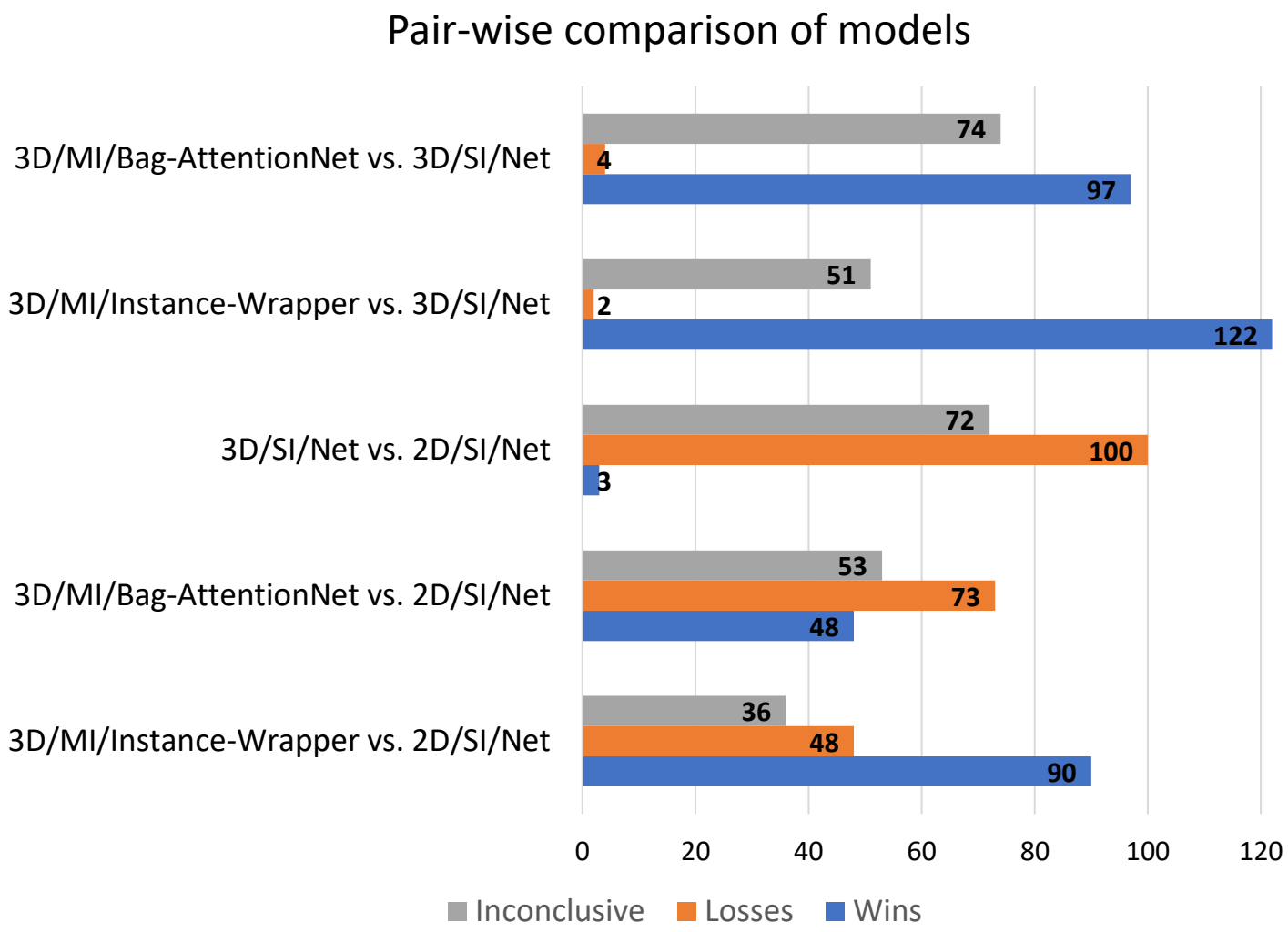
139 data sets



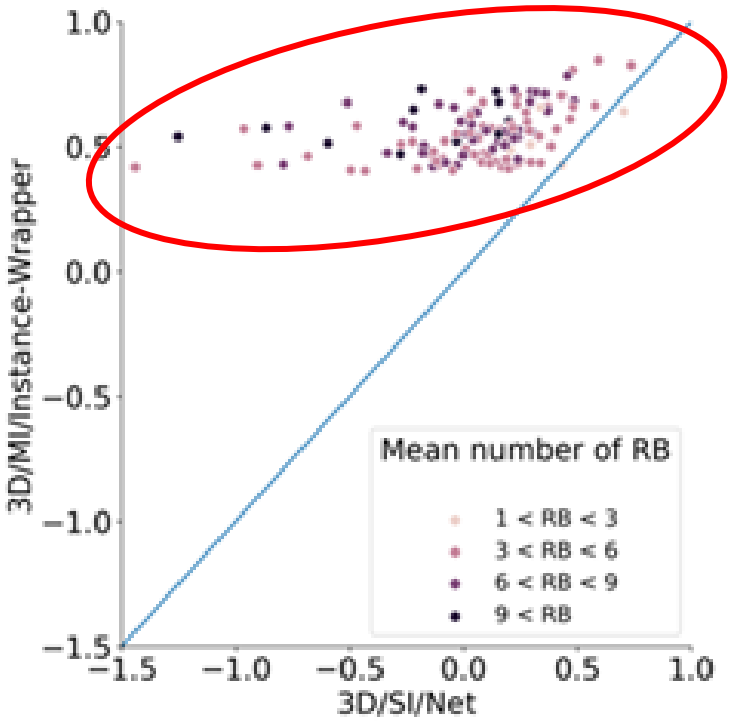
Comparison of 2D and 3D models

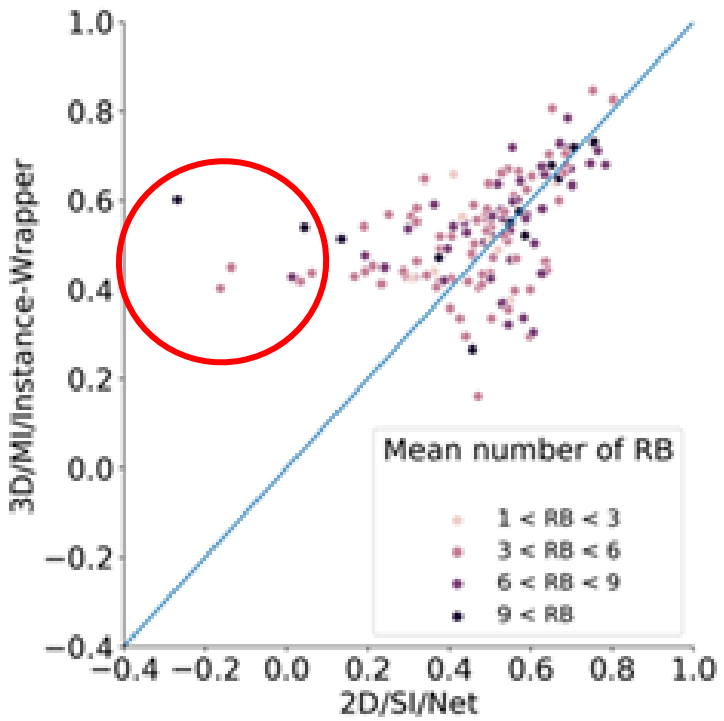


Comparison of 2D and 3D models



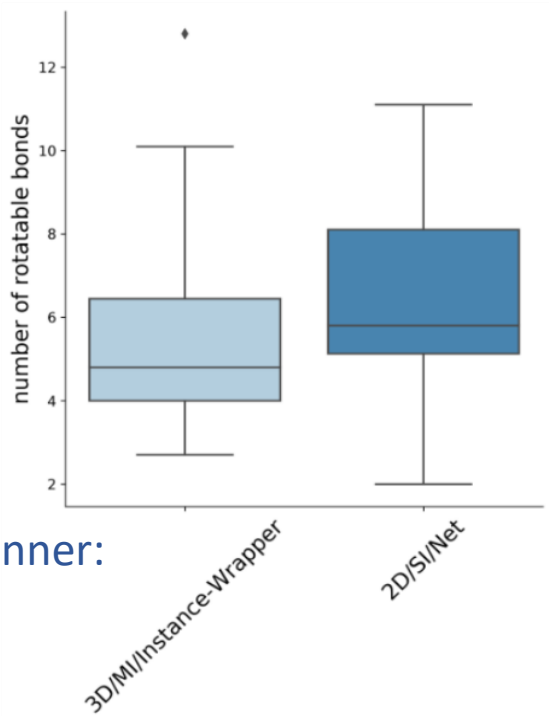
Comparison of 2D and 3D models





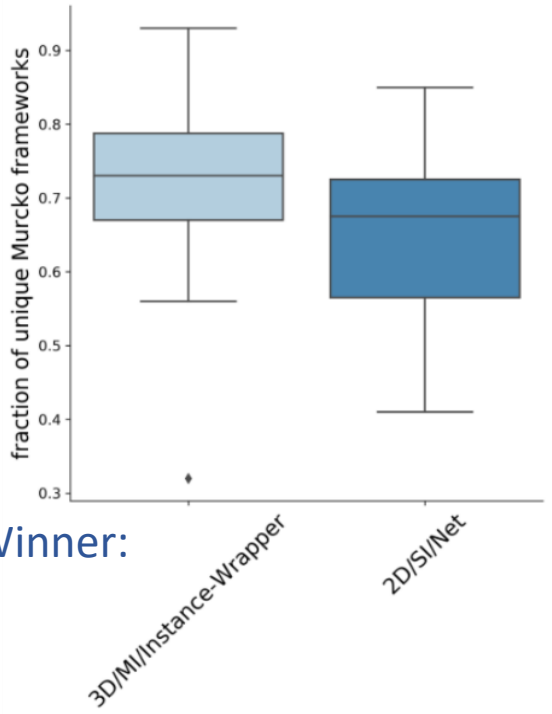
Comparison of 2D and 3D models

Number of rotatable bonds



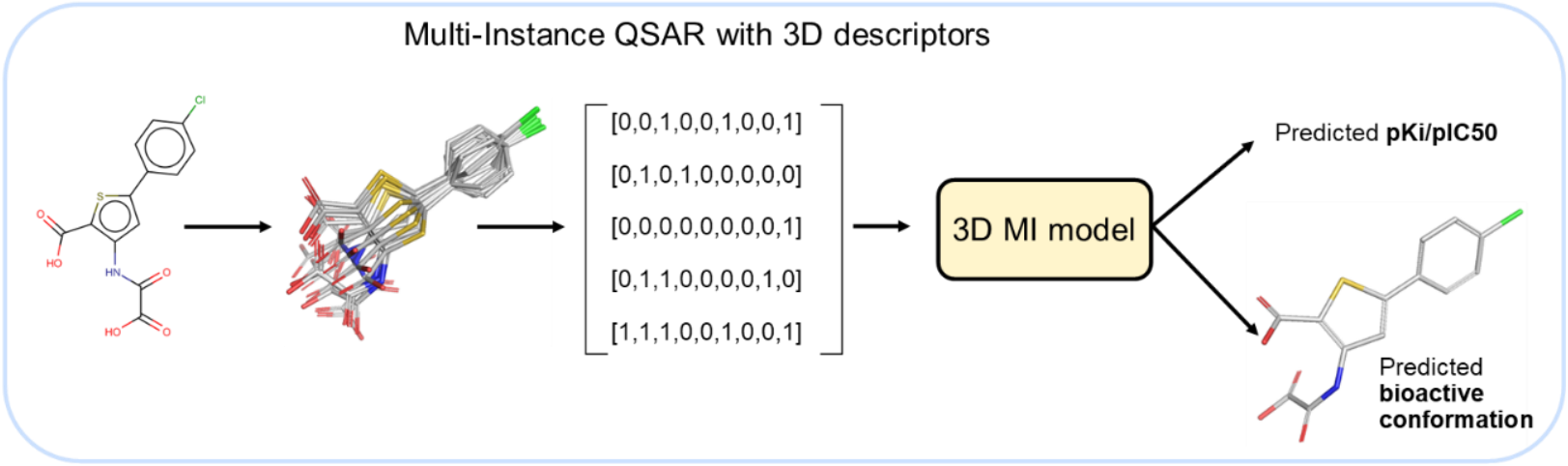
Winner:

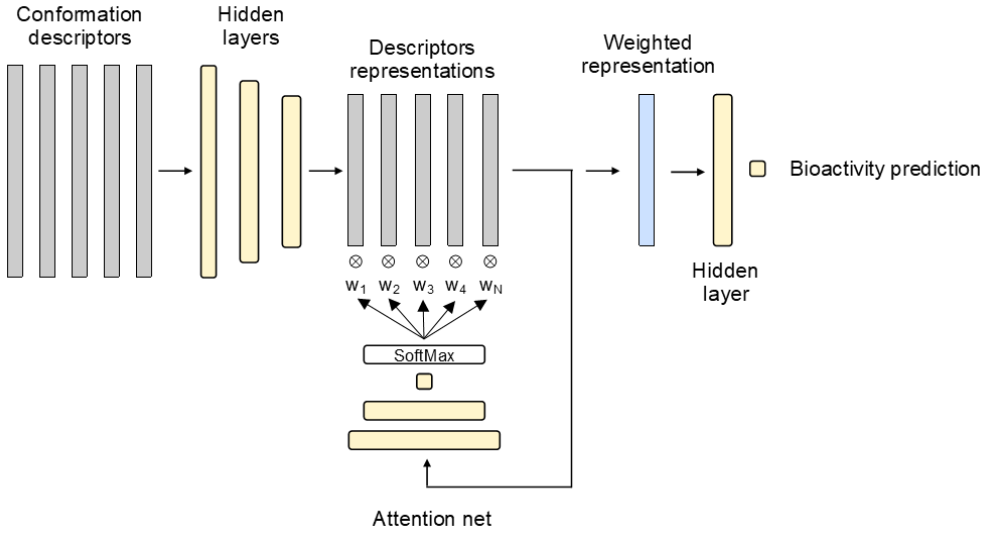
Fraction of unique Bemis-Murcko scaffolds



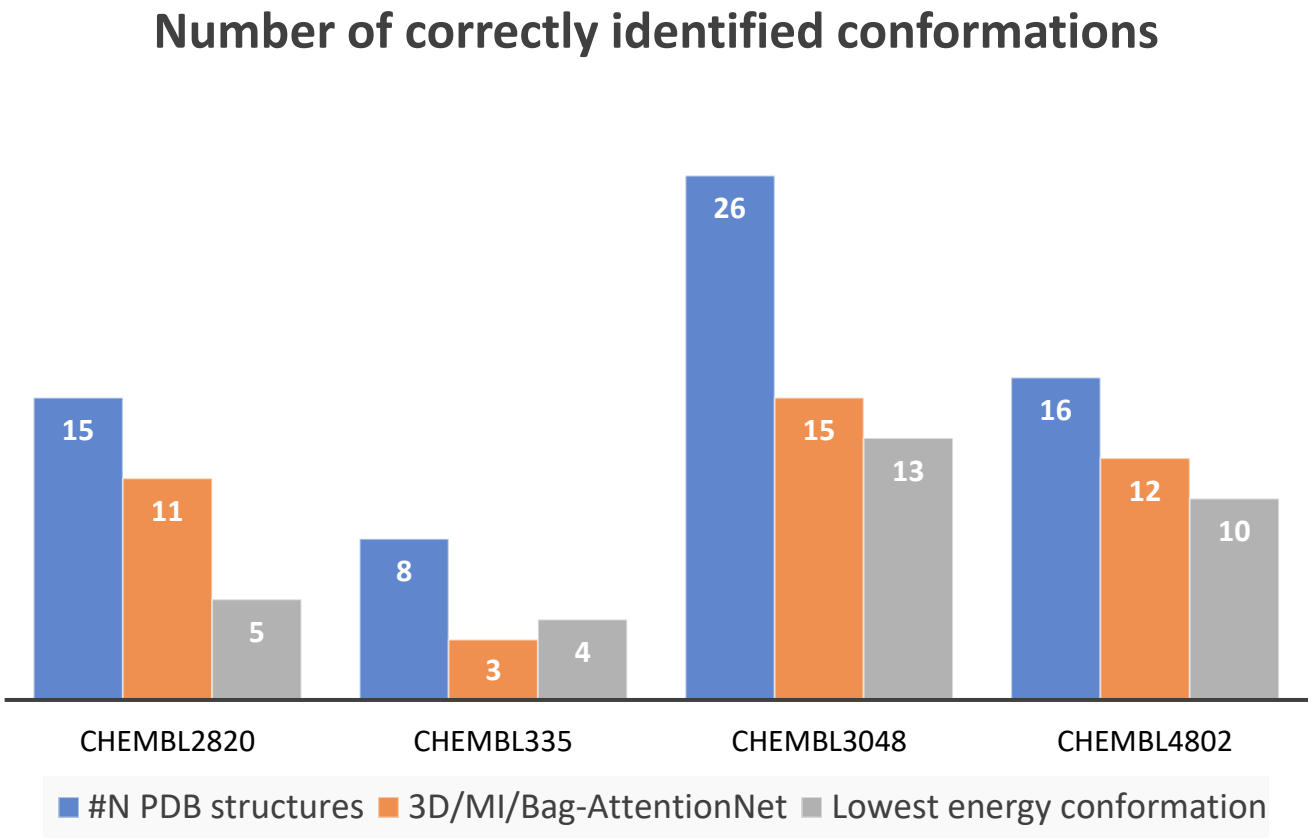
Winner:

Identification of Bioactive Conformation





Identification of Bioactive Conformation



Identification of Bioactive Conformation

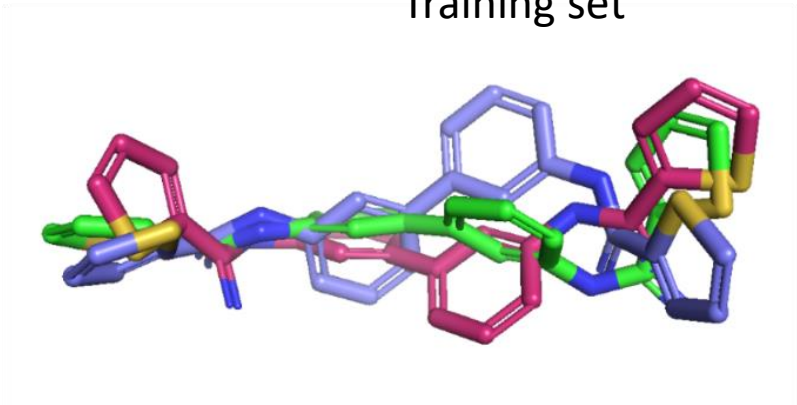
PDB conformation

Lowest-energy conformation

Conformation predicted by *3D/MI/Bag-AttentionNet* model

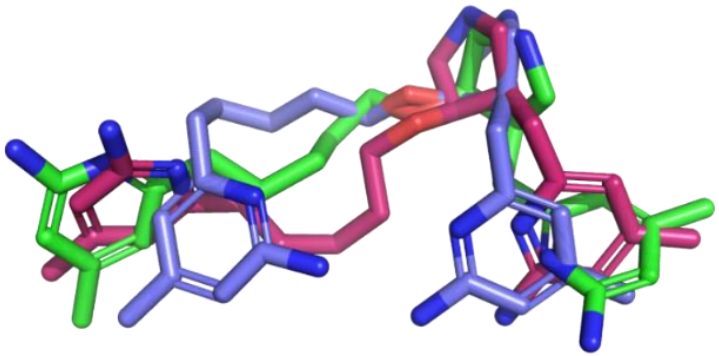
CHEMBL2820

Training set



Experimental pKi: 6.10
3D/SI/Net pKi: 6.48, RMSD = 2.42 Å
3D/MI/Bag-AttentionNet pKi: 6.31, RMSD = 1.70 Å

Test set



Experimental pKi: 7.42
3D/SI/Net pKi: 7.86, RMSD = 2.78Å
3D/MI/Bag-AttentionNet pKi: 7.41, RMSD = 1.55 Å

Conclusions

1) Multi-instance models outperform both single-instance 3D models and traditional QSAR models built on 2D descriptors in many cases
2) Multi-instance models is good alternative to 2D modeling if the latter fails
3) Multi-instance neural network with an attention mechanism can correctly identify a “bioactive” conformation close to the experimental structure of a ligand retrieved from PDB.

(1) Nikonenko, A. V.; Zankov, D. V.; Baskin, I.; Madzhidov, T. I.; Polishchuk, P. Multiple Conformer Descriptors for QSAR Modeling. *Mol. Inform.* **2021**, submitted.

(2) Zankov, D. V.; Shevelev, M. D.; Nikonenko, A. V.; Polishchuk, P. G.; Rakhimbekova, A. I.; Madzhidov, T. I. Multi-Instance Learning for Structure-Activity Modeling for Molecular Properties. In *Analysis of Images, Social Networks and Texts. AIST 2019. Communications in Computer and Information Science*; van der Aalst, W. M. P., et al., Eds.; 2020; pp 62–71. https://doi.org/10.1007/978-3-030-39575-9_7.

(3) Zankov, D.; Matveieva, M.; Nikonenko, A.; Nugmanov, R.; Varnek, A.; Polishchuk, P.; Madzhidov, T.I. QSAR modeling based on conformation ensembles using a multi-instance learning approach. *ChemRxiv, https://chemrxiv.org/articles/preprint/QSAR_Modeling_Based_on_Conformation_Ensembles_Using_a_Multi-Instance_Learning_Approach/13456277*

All algorithms and data sets are freely available at
https://github.com/cimm-kzn/3D-MIL-QSAR

Acknowledgements

Соглашение с МНиВО РФ
N 14.587.21.0049

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Программа INTER-ACTION, MSMT-5727/2018-2

Павел Полищук (Университет Оломоуца)

Александр Варнек (Университет Страсбурга)

Игорь Баскин (Технион)

Дмитрий Занков

Рамиль Нугманов

Мария Матвеева (Университет Оломоуца)

Acknowledgements

Спасибо за внимание!!!

Тимур Исмаилович Маджидов
E-mail:
Timur.Madzhidov@kpfu.ru