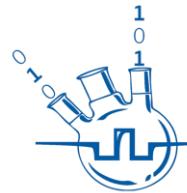




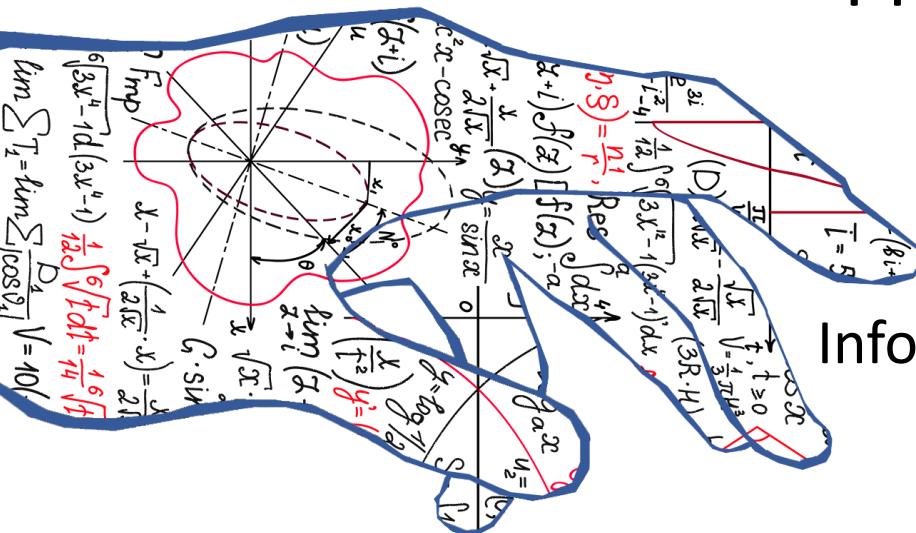
ITMO UNIVERSITY



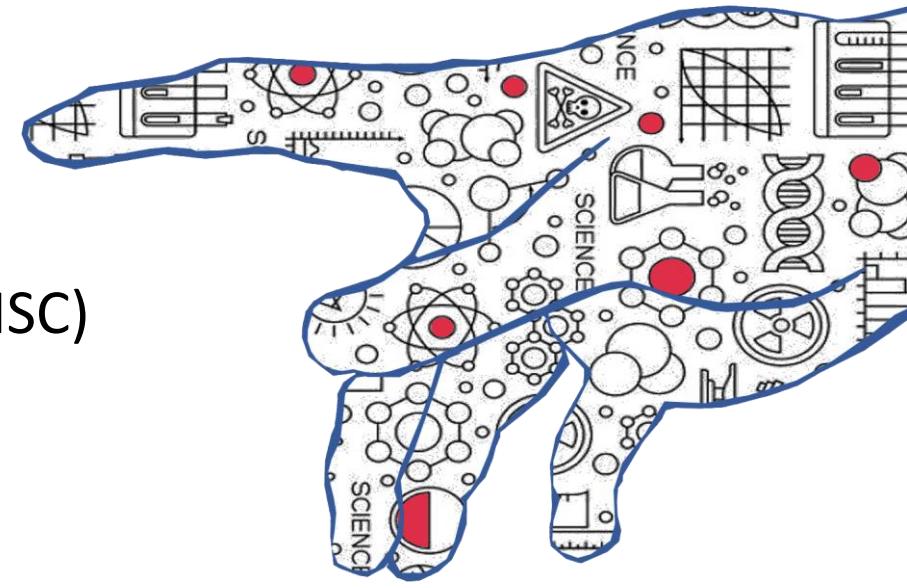
INFOCHEMISTRY SCIENTIFIC CENTER

Cheminformatics: Basic Concepts and Areas of Application

Prof. Dr. Sergey Shityakov

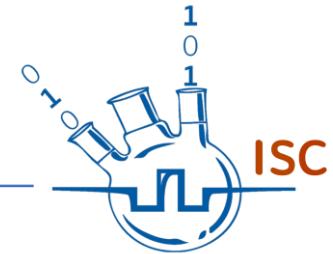


Infochemistry Scientific Center (ISC)
ITMO University
Saint-Petersburg, 2021





Chemoinformatics lab



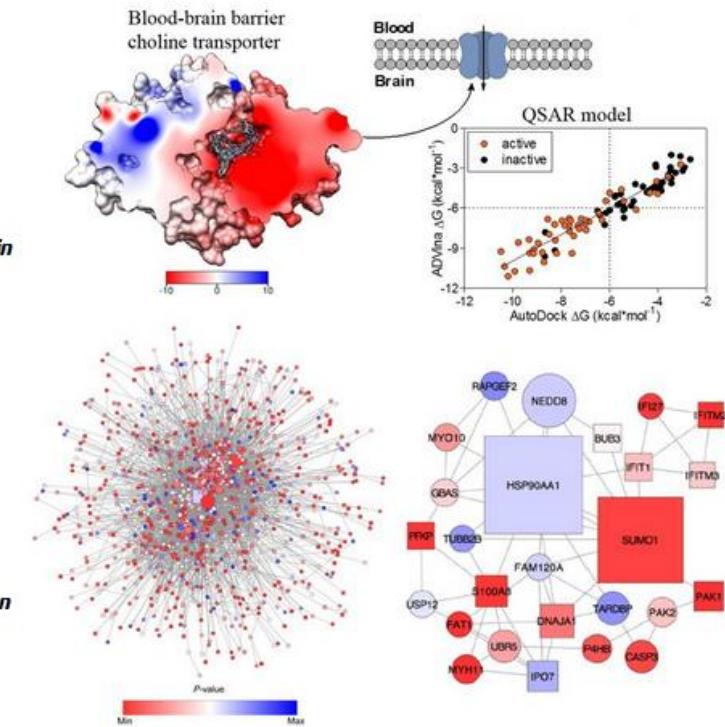
Chemoinformatics

Shityakov Sergey Vasilievich



- Development of virtual screening methods
- Creation of virtual libraries
- Search for quantitative structure-property relationships (QSAR) of chemical compounds
- Experimental validation of QSAR models

- 1) Sergey Shityakov, M.D., Ph.D. (*Principal Investigator*)
- 2) Aqeel A. Hussein, Ph.D. (*Visiting Scientist, Computational Chemistry, University of Southampton, UK*)
- 3) Tilemachos Kafetzis, Dr. dent. cand. (*In silico approach to analyze adsorption of antibacterial peptides on hydroxyapatite nanosurfaces for more effective antibacterial treatment applied in oral surgery*)
- 4) Patrick Eirich, Dr. med. cand. (*ITC measurements and thermodynamical stability of VSG protein isolated from Trypanosoma brucei*)
- 5) Benedikt Wilhelms, Dr. med. cand. (*Chemical characterization and modeling of cyclodextrin-formulated general anaesthetics at the blood-brain barrier*)
- 6) Anneli Fischer, Dr. med. cand. (*Pharmacokinetics of drug-human serum albumin complexes*)
- 7) Tanja Gegg, Dr. med. cand. (*In vitro cytotoxicity and mutagenicity of graphene and silica nanoparticles*)



ITsMOre than a
UNIVERSITY



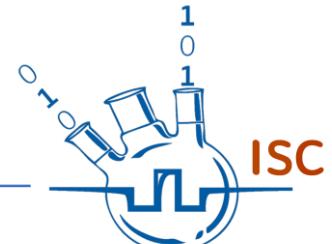
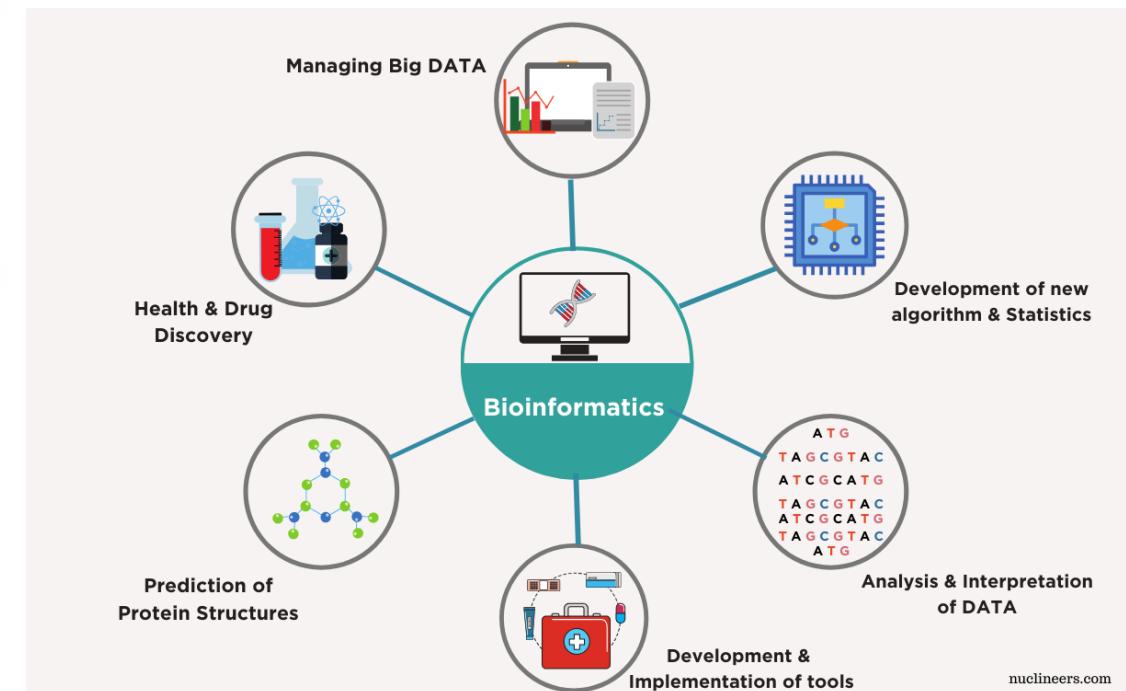
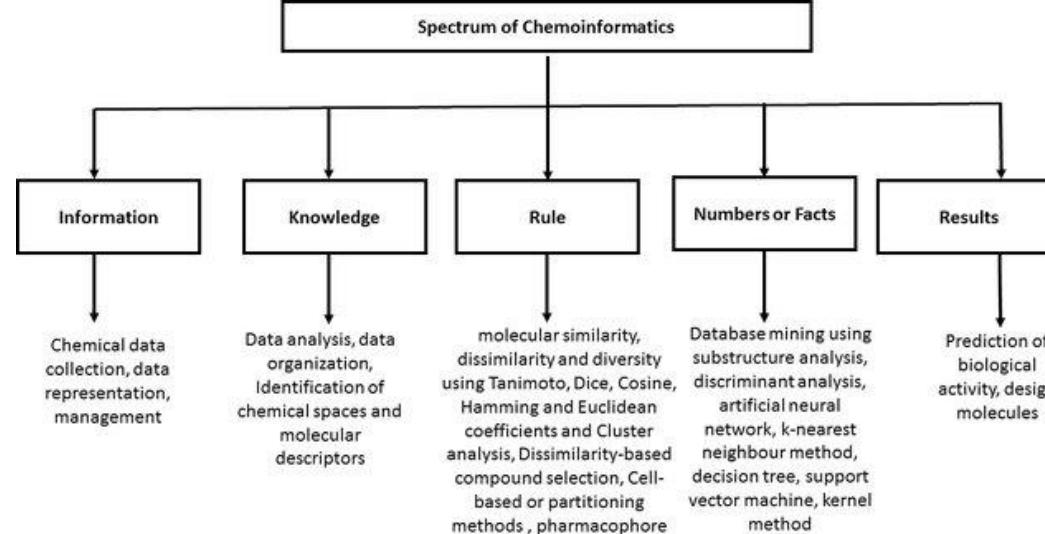
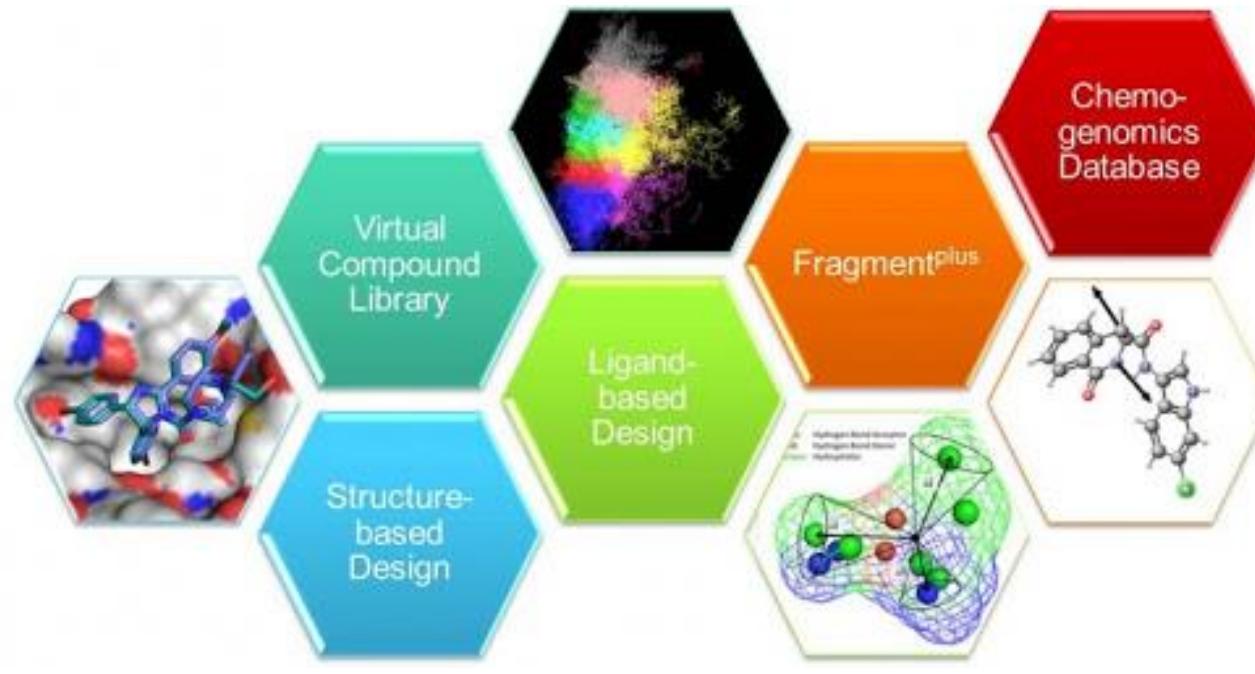
Chemoinformatics - definition



- Chemoinformatics is a generic term that encompasses the design, creation, organization, management, retrieval, analysis, dissemination, visualization, and use of chemical information - G. Paris, 1998
- Chemoinformatics is the mixing of those information resources to transform data into information and information into knowledge for the intended purpose of making better decisions faster in the area of drug lead identification and optimization - F.K. Brown, 1998
- Chemoinformatics is the application of informatics methods to solve chemical problems - J. Gasteiger, 2004
- Chemoinformatics is a field based on the representation of molecules as objects (graphs or vectors) in a chemical space - A. Varnek & I. Baskin, 2011

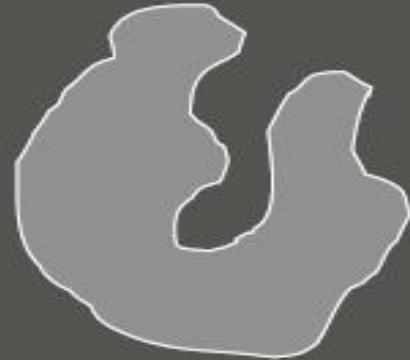
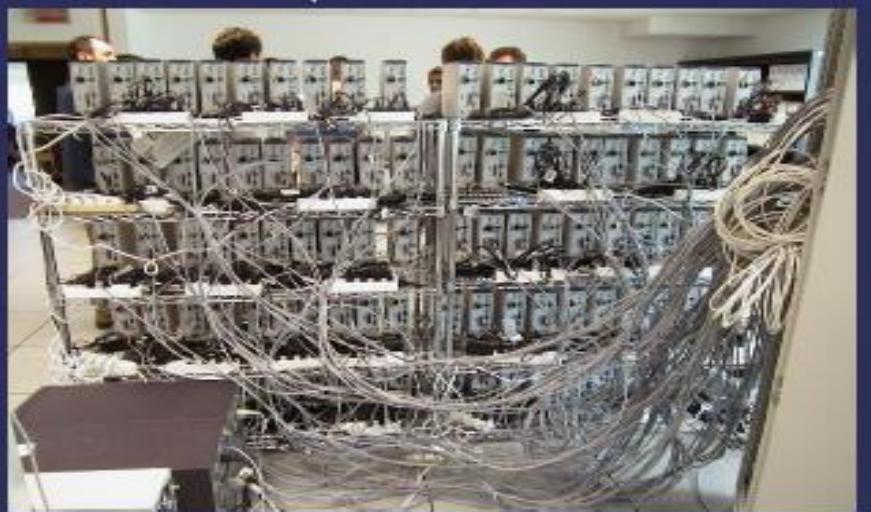


Chemoinformatics and bioinformatics

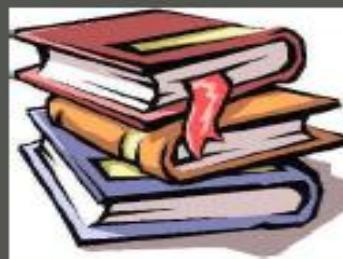



IT'S MORE than a
UNIVERSITY

computations



Target Protein



Large libraries
of molecules

Virtual
Screening

Small Library of selected hits



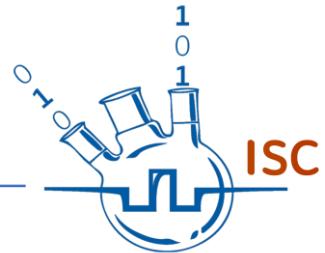
experiment



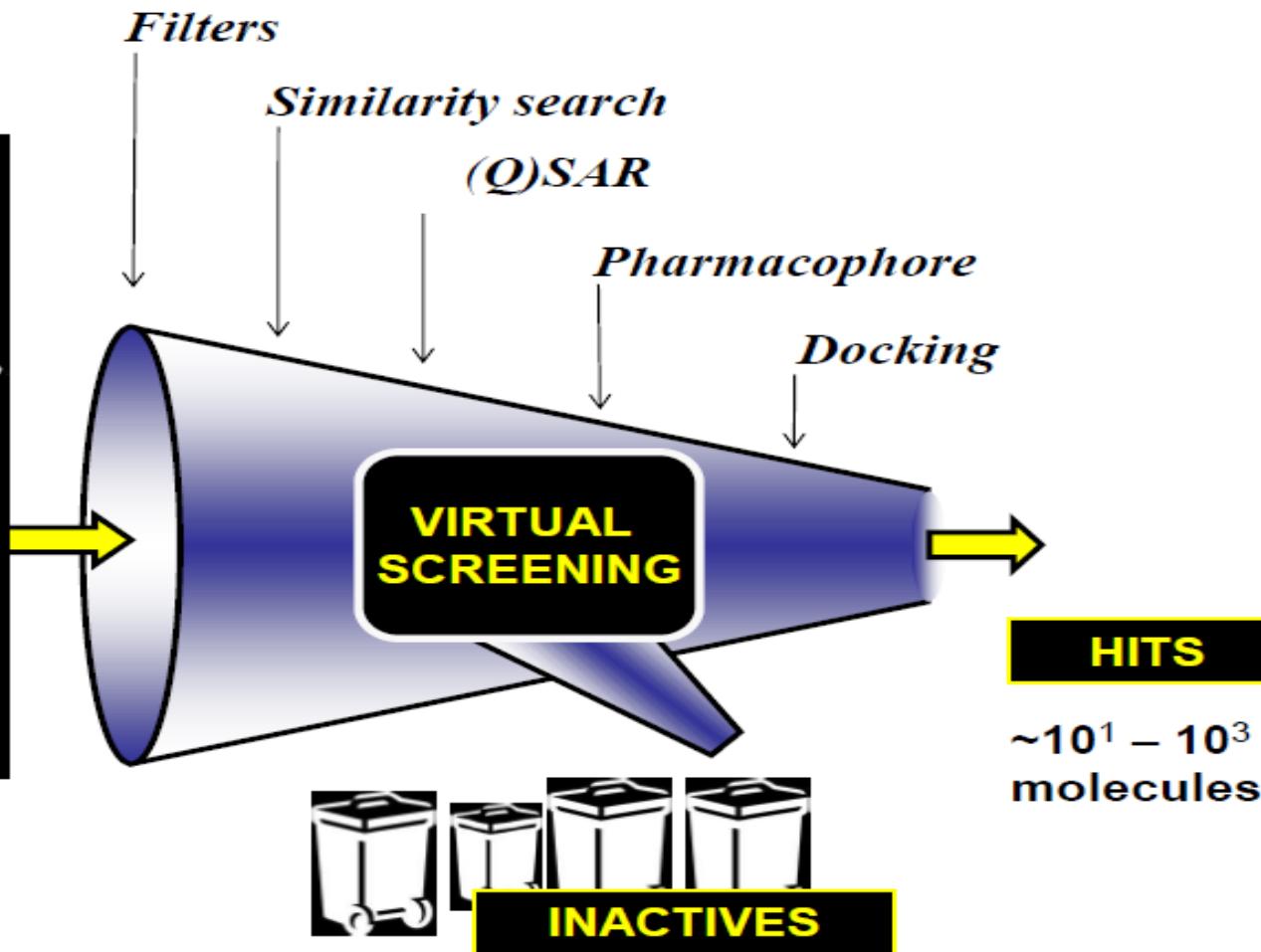
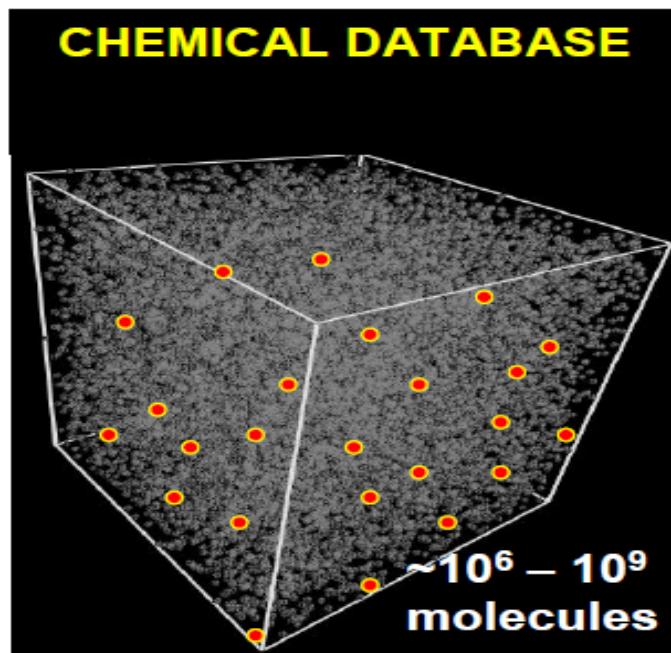
High Throughout Screening



Chemoinformatics paradigm

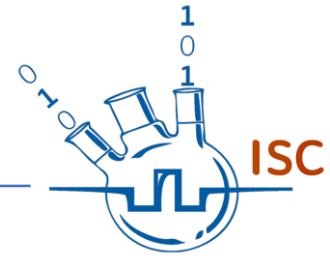


Chemoinformatics: Virtual screening “funnel”

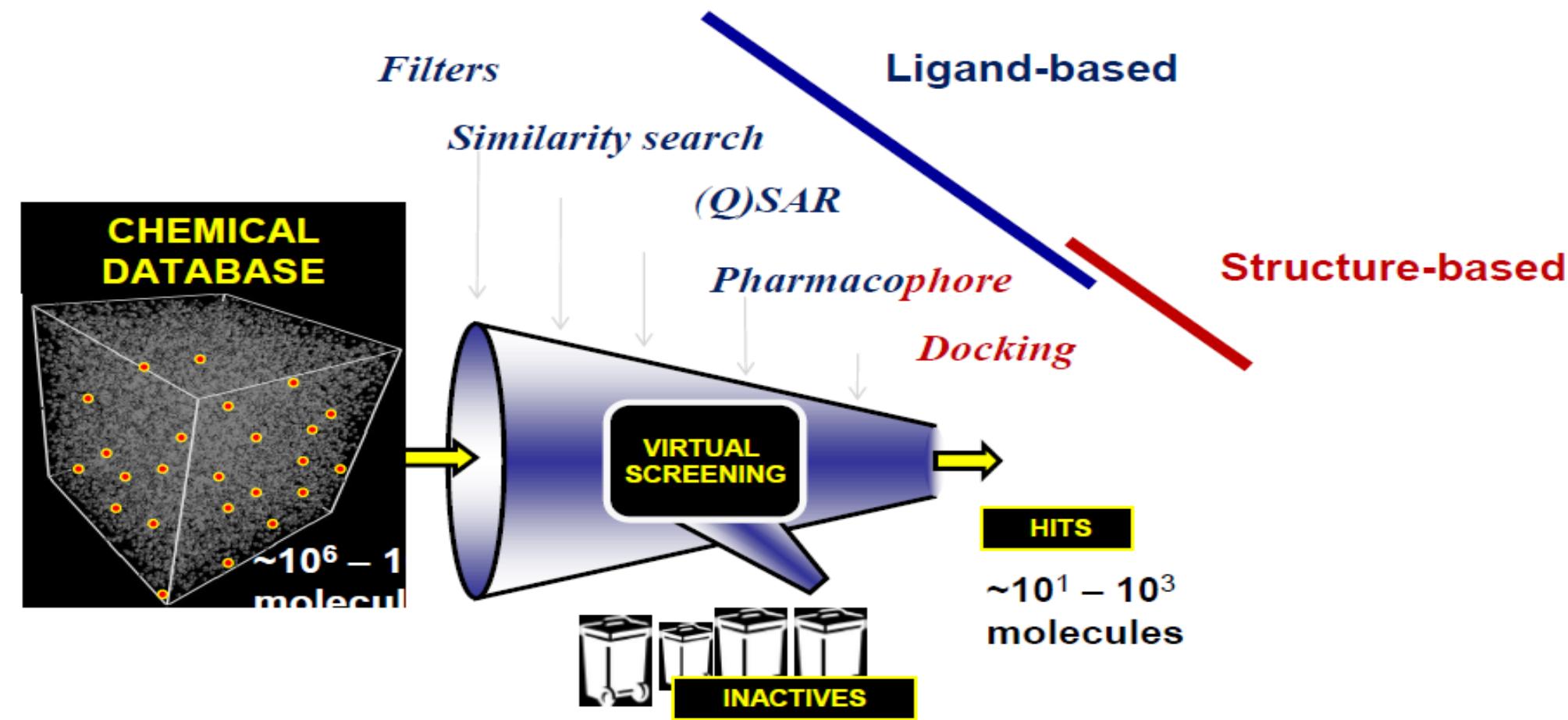




Chemoinformatics paradigm

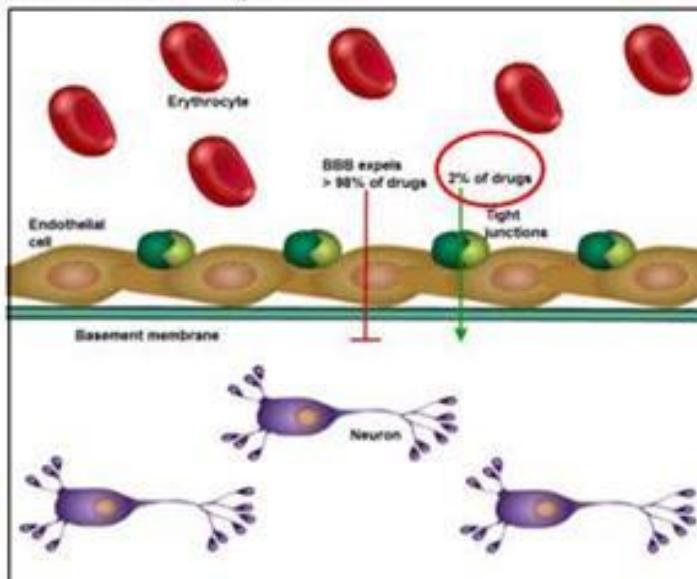


Chemoinformatics: Virtual screening “funnel”



Research interests: medicinal chemistry and rational drug design

Schematic representation of the BBB permeability for different drug-like chemical compounds



Rational drug design to detect hit/lead substances via filtering chemical compound libraries available for the BBB research

Shityakov *et al.*, OA Anesthetics, 2013

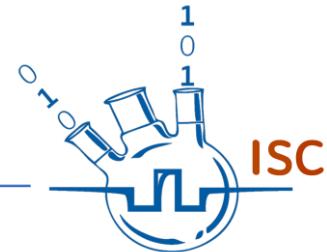
Universitätsklinikum Würzburg



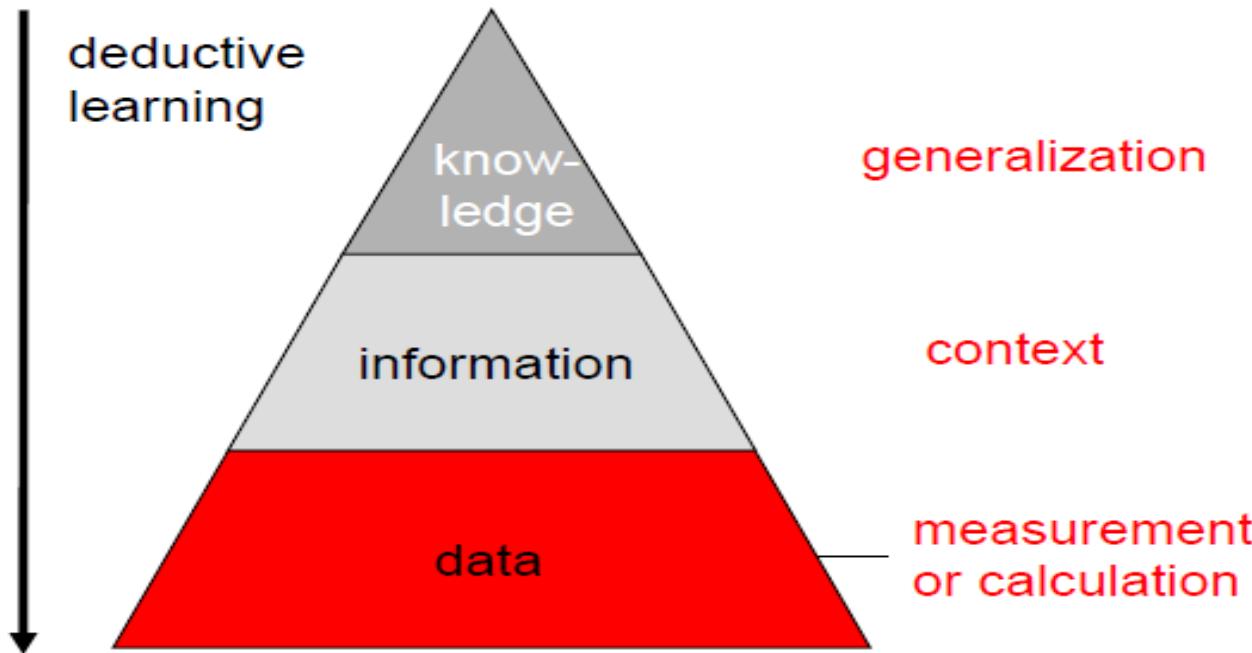
ITSMORE than a
UNIVERSITY



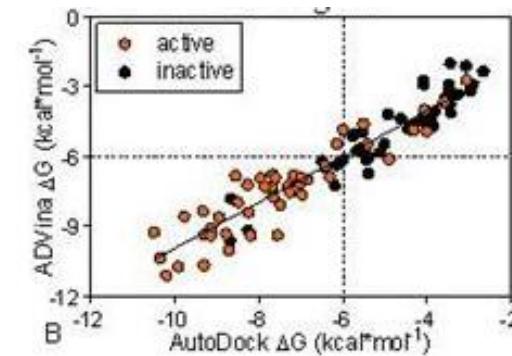
Chemoinformatics



Chemoinformatics: From Data to Knowledge



inductive
learning



ΔG – Gibbs free energy of binding (binding affinity)

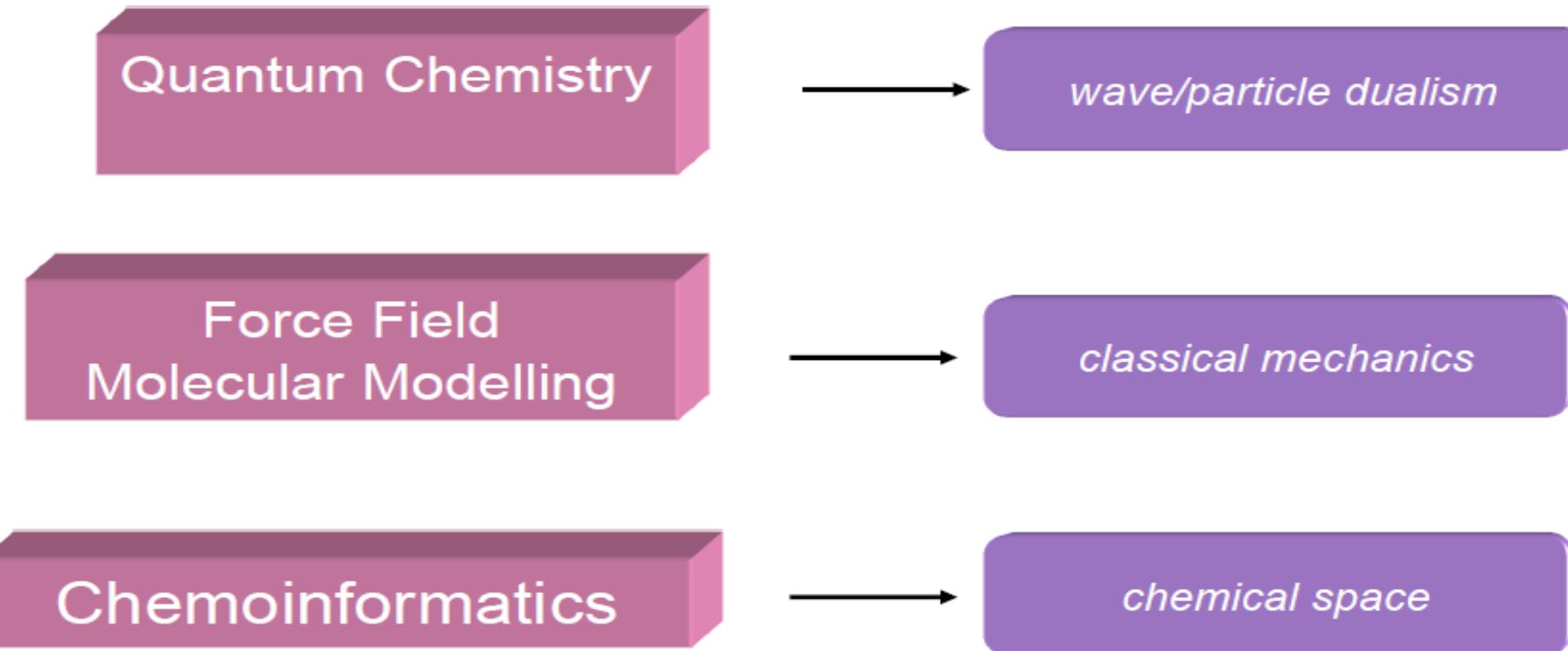
Shityakov *et al.*, Bioinf., 2014

Chemoinformatics learns from experimental data !

ITMO more than a
UNIVERSITY

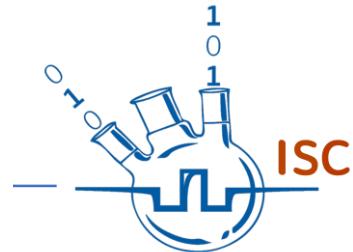


Basic concepts





Molecular model



Quantum Chemistry

electrons and nuclei

Force Field Molecular Modelling

atoms and bonds

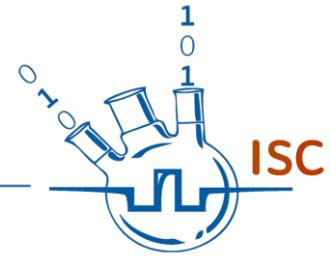
Chemoinformatics

- *molecular graph*
- *descriptor vector*

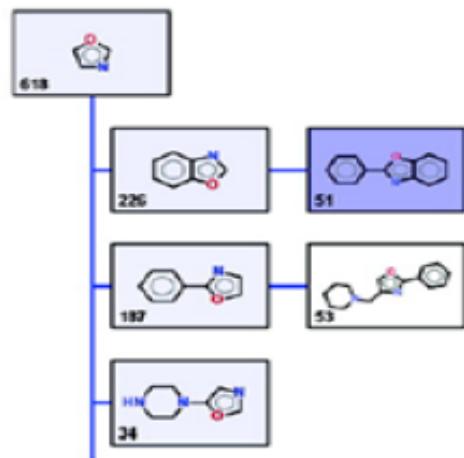
Chemoinformatics is a field based on the representation of molecules as objects (graphs or vectors) in a chemical space



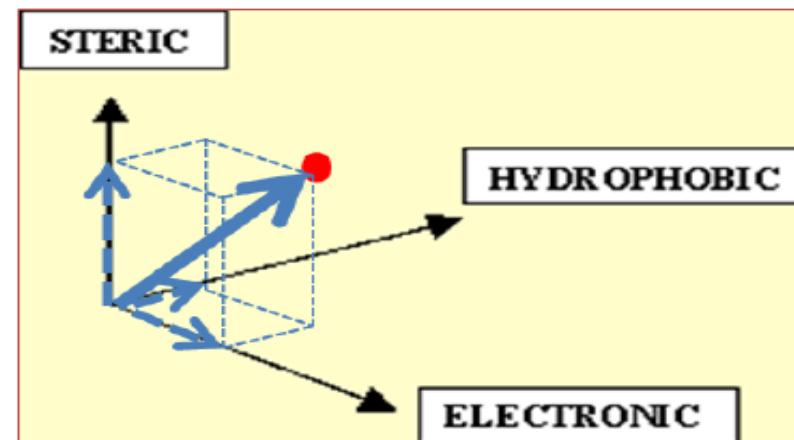
Chemical scape and descriptors



Chemical Space representations



graphs-based

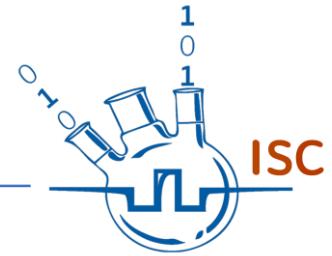


descriptors -based

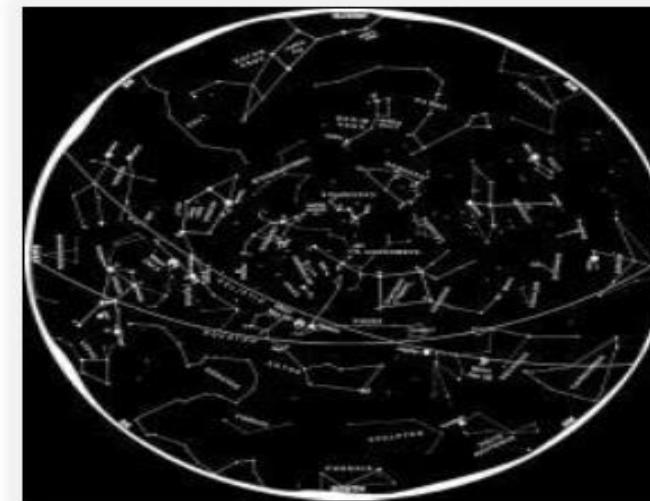
SPACE = objects + metric



Chemical space

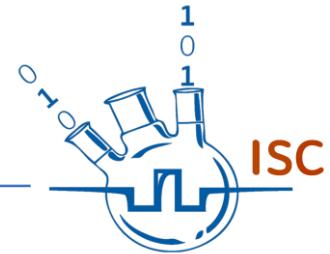


Graph-based chemical space

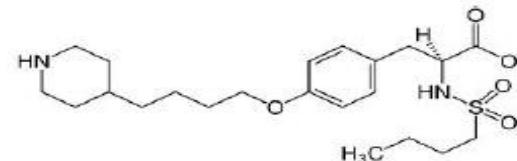
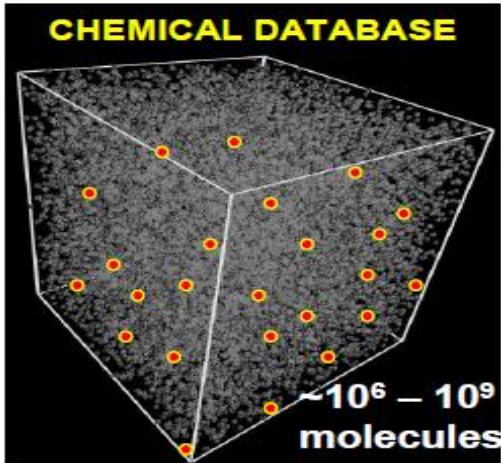




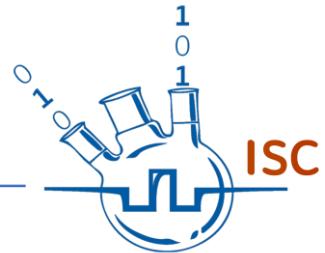
Chemical space



Virtual screening : finding the needle in the haystack

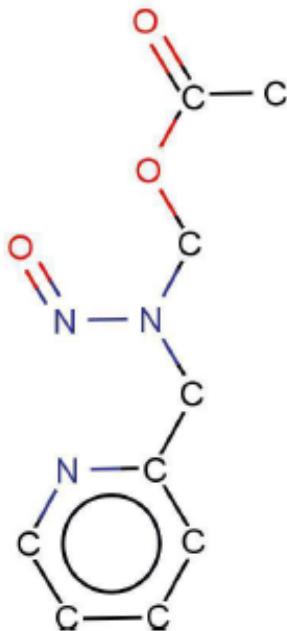


Molecular descriptors



ensemble of topological, electronic, geometry parameters calculated directly from molecular structure

Molecular graph



- Topological indices,
- Atomic charges,
- Inductive descriptors,
- Substructural fragments,
- Molecular volume and surface, ...

Descriptor vector

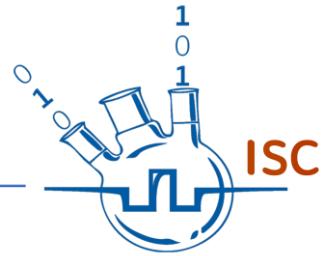
Descriptors
D ₁
D ₂
...
D _i
...



> 5000 types of descriptors are reported

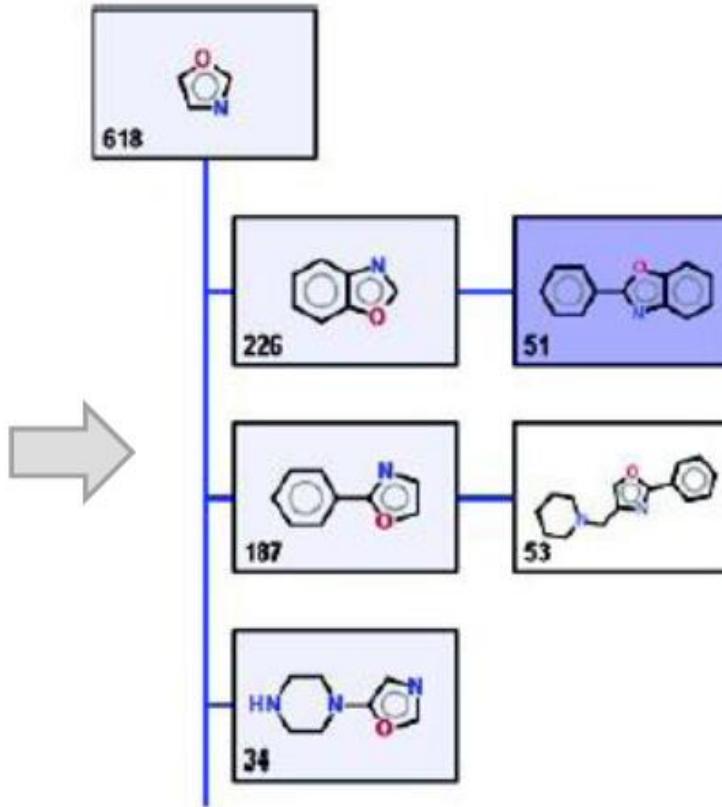
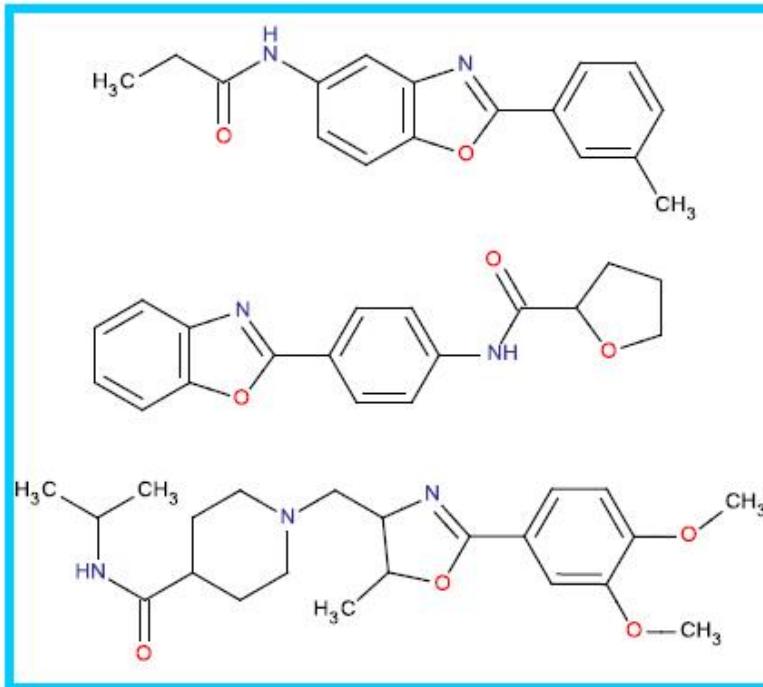


Scaffold tree



Oxazole ring

Database



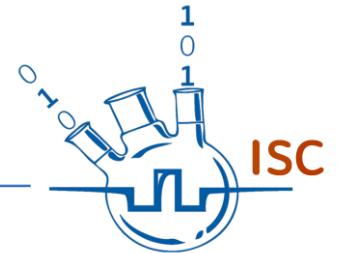
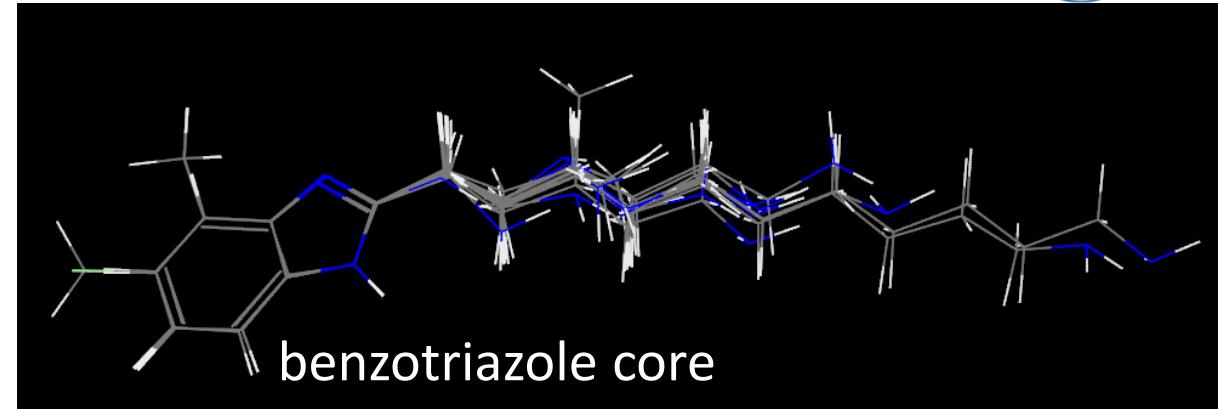
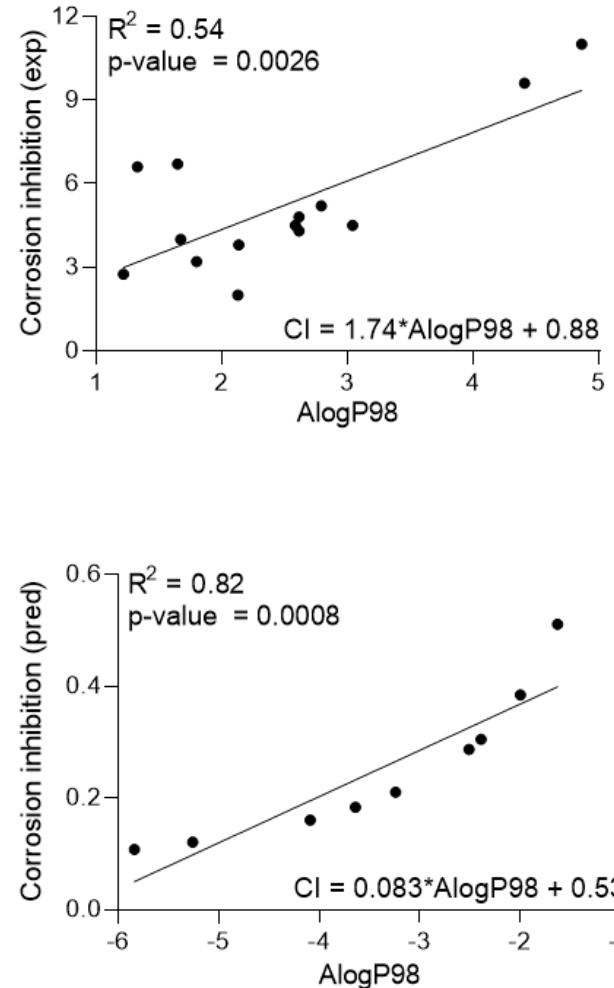
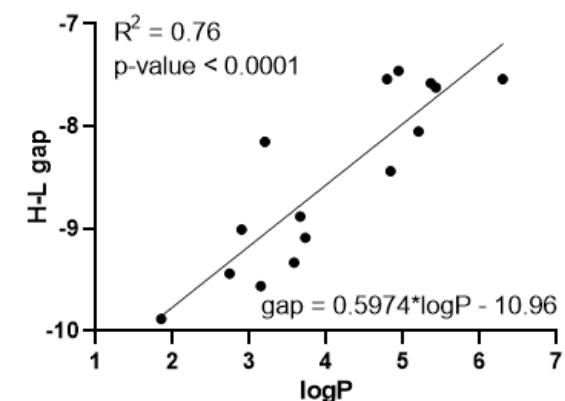
A. Schuffenhauer, P. Ertl, et al. *J. Chem. Inf. Model.*, 2007, 47 (1), 47-58

Molecular scaffold describes the core structure of a compound or series. Position (site) of chemical variation in a core structure carrying different substituents (functional groups, R-groups) in different analogs.

ITMO *more than a*
UNIVERSITY



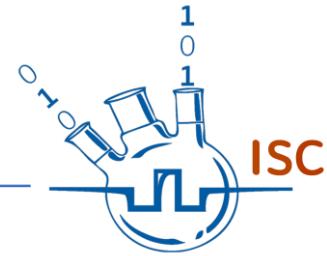
Cheminformatics and molecular descriptors

Structure	AlogP98	CI (pred)
1,3-linked β -D-galactopyranosyl	-5.84	0.107679
α -L-arabinofuranosyl	-3.24	0.210073
α -L-rhamnopyranosyl	-4.09	0.160254
β -D-glucopyranosyl	-5.26	0.120816
4-O-methyl- β -D-glucopyranosyl	-3.64	0.183263
galactose	-2.51	0.28659
arabinose	-2	0.384408
rhamnose	-1.63	0.510926
glucuronic acid	-2.39	0.304842

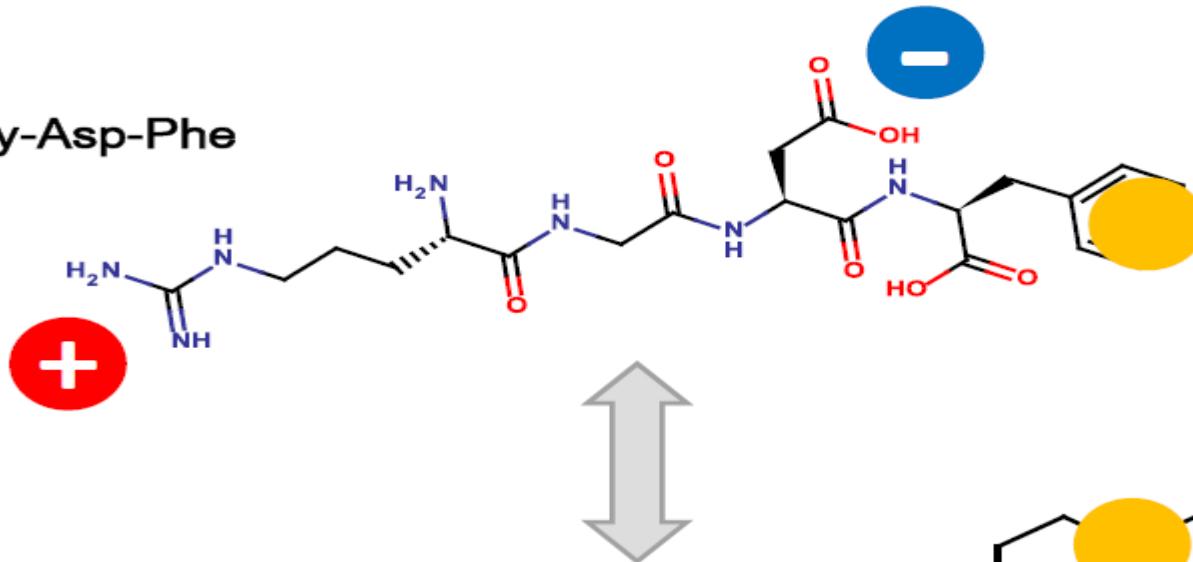


Similarity: pharmacophore

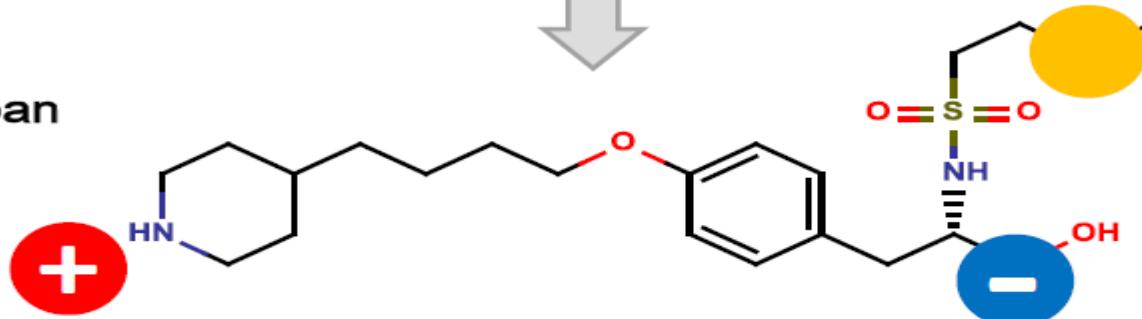


What is in common between these two molecules ?

Arg-Gly-Asp-Phe

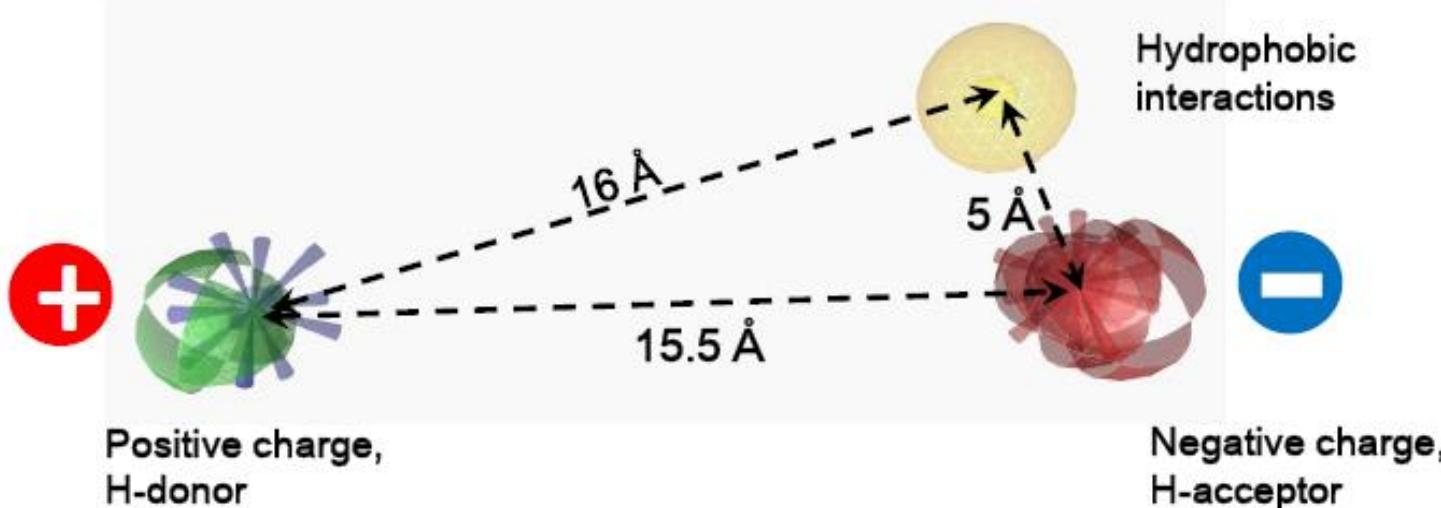
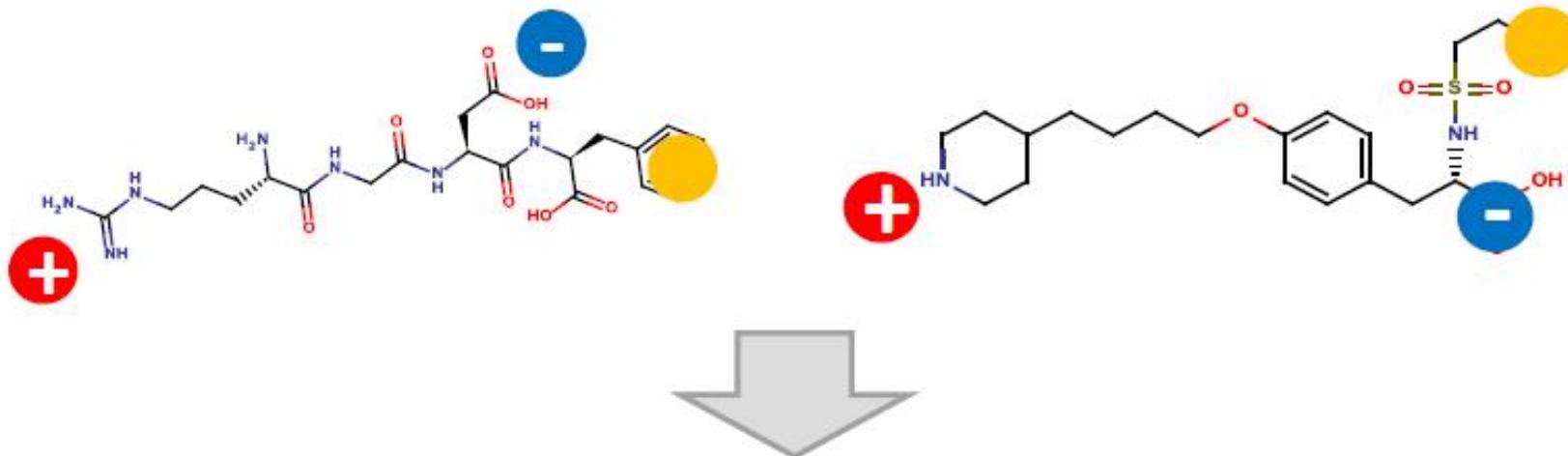
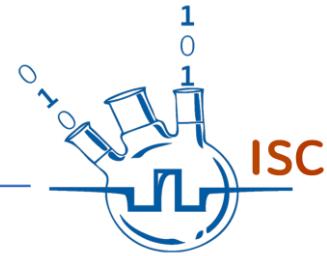


Tirofiban



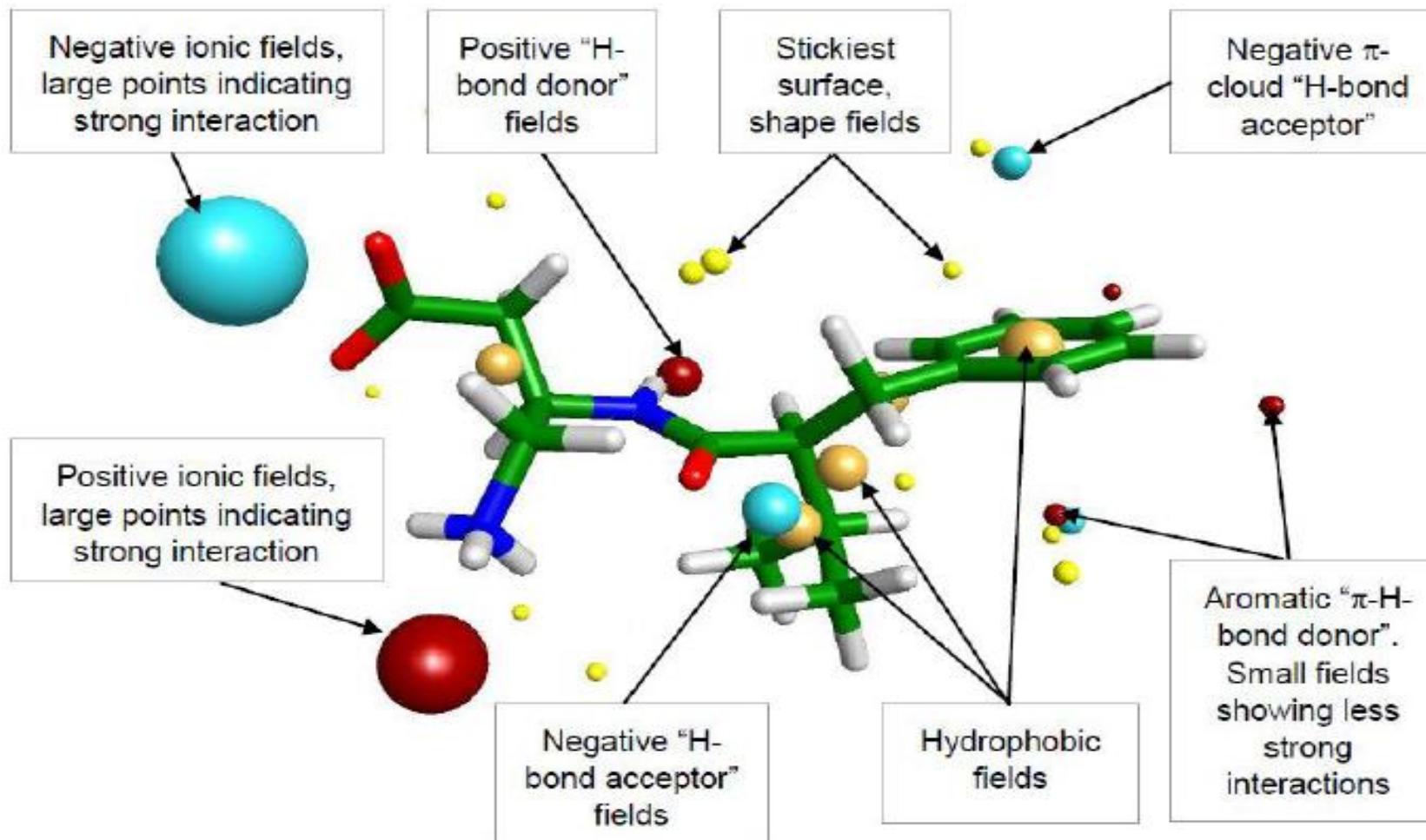


Similarity: pharmacophore





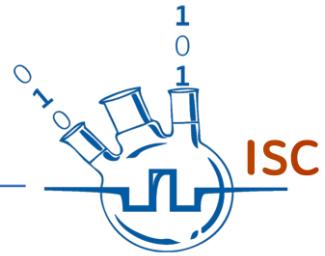
Molecular fields (CRESSET)



Interpretation of a field point pattern. The size of the point indicates the potential strength of the interaction

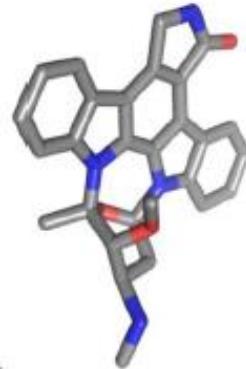


Protein-ligand molecular docking



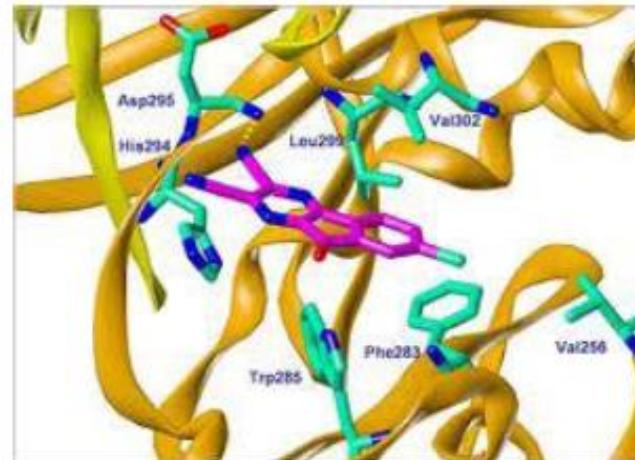
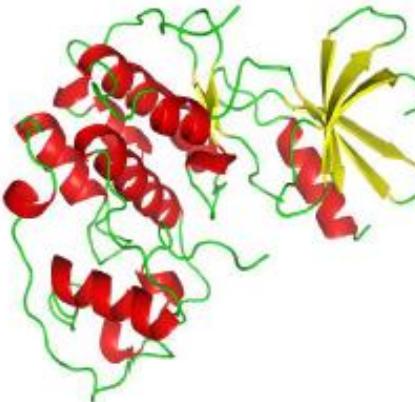
Hermann Emil Fischer

Key



+

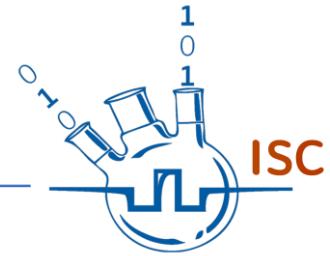
Lock



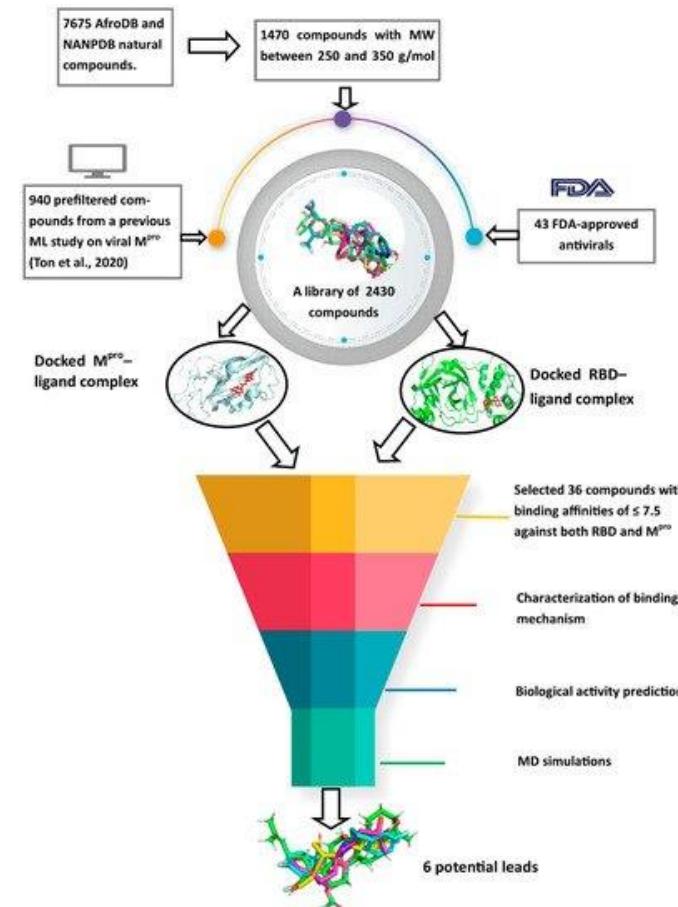
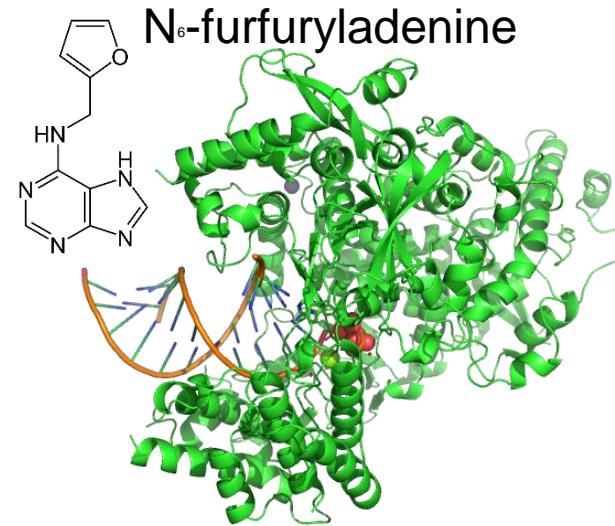
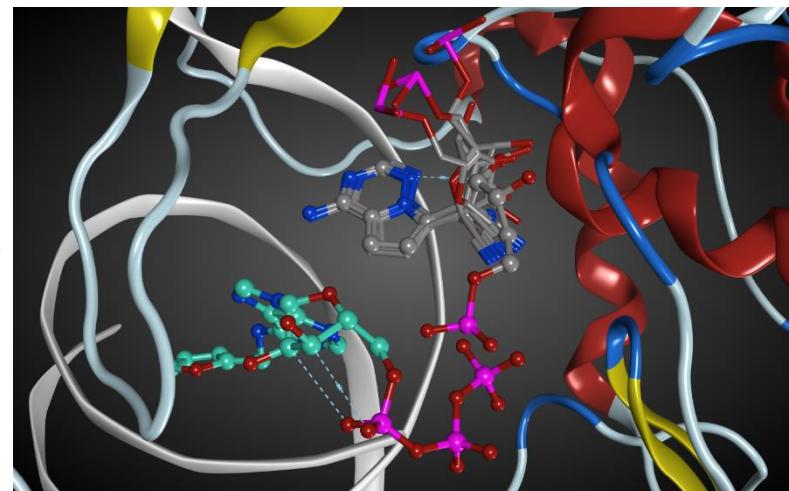
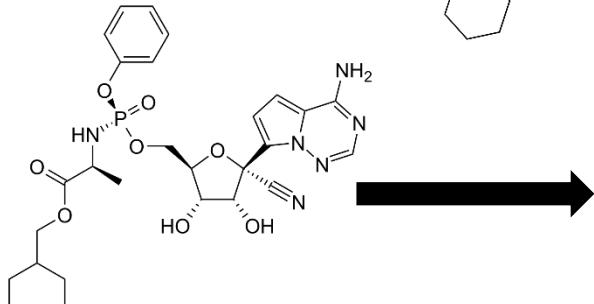
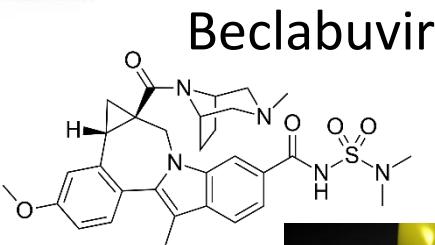
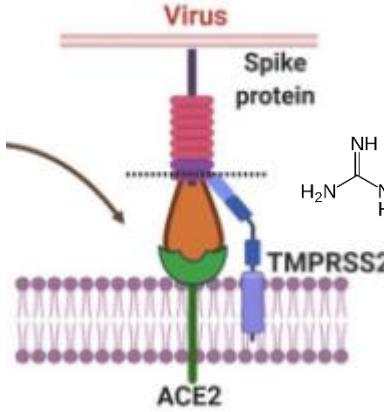
Ligand-Protein complex

ITMO *more than a*
UNIVERSITY

Cheminformatics and COVID-19 pandemics



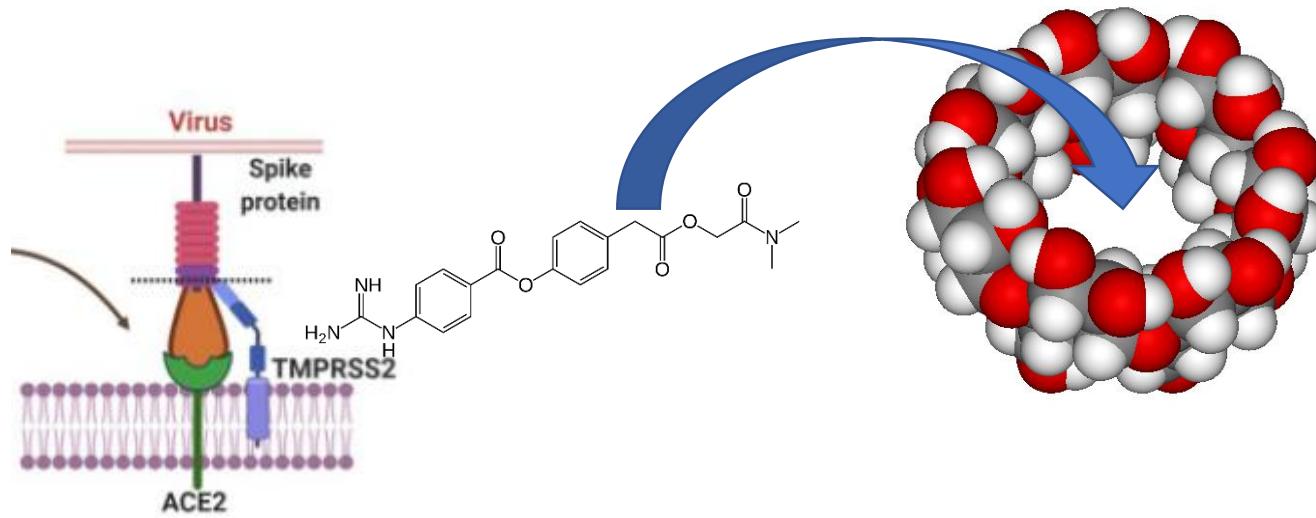
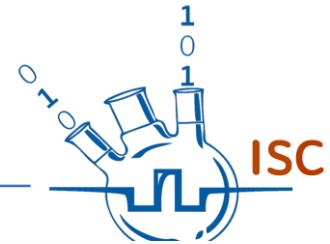
prototype drugs



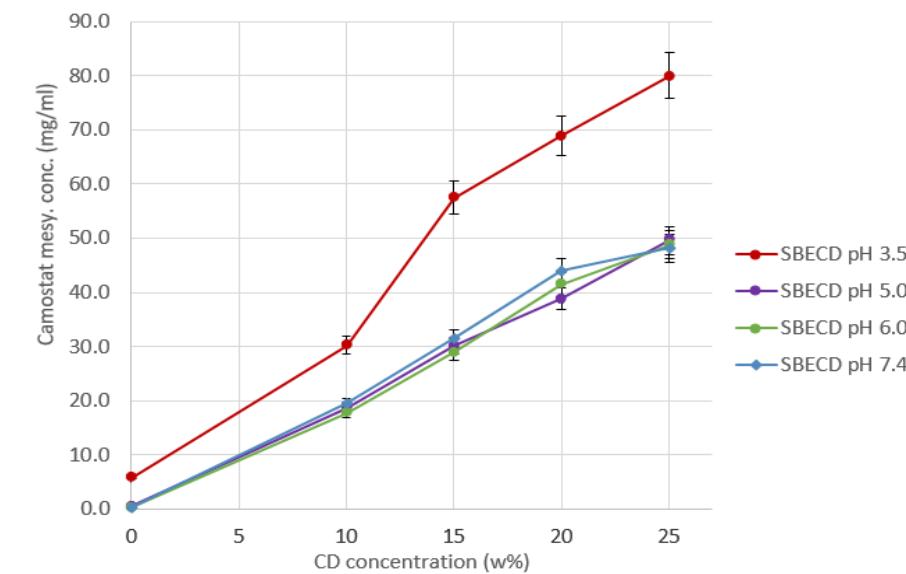
Shityakov et al., in progress

ITSMORE than a UNIVERSITY

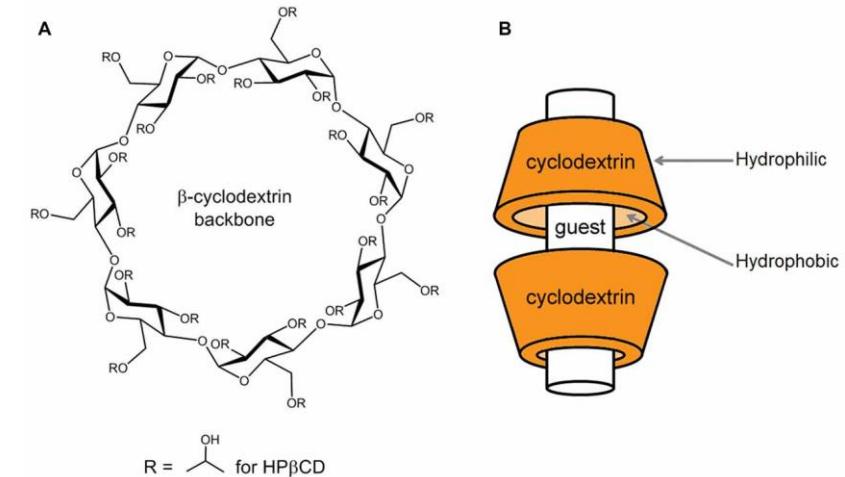
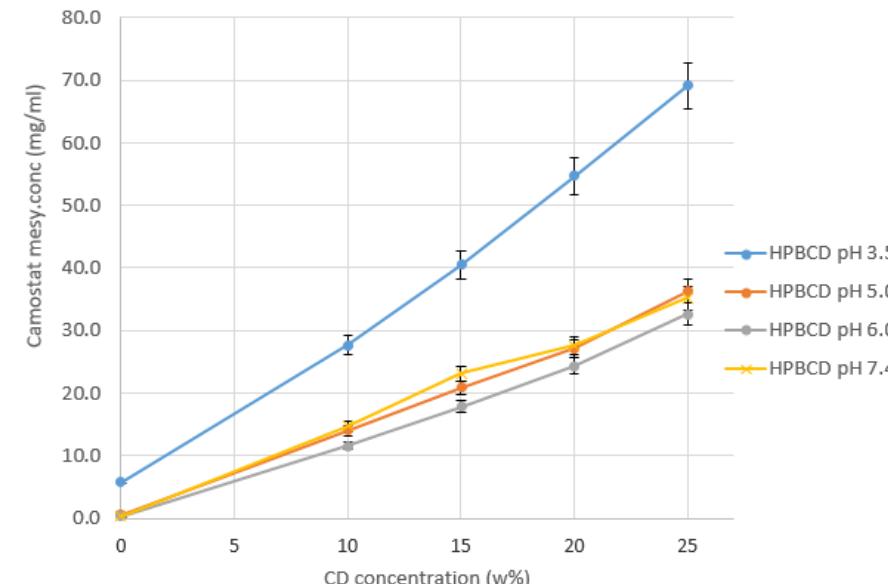
Cheminformatics and COVID-19 pandemics



Prototype drug solubility isotherms - SBECD

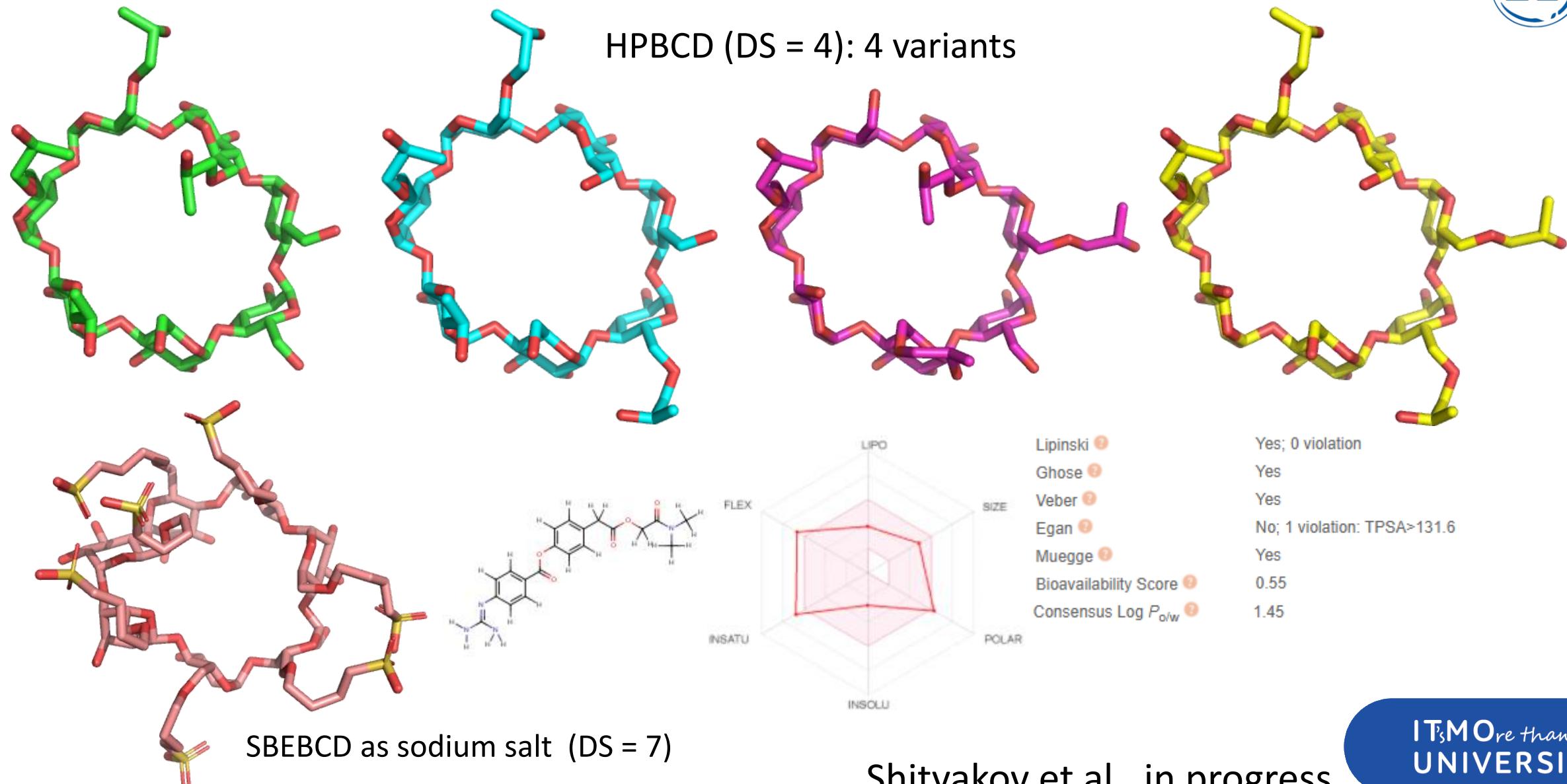
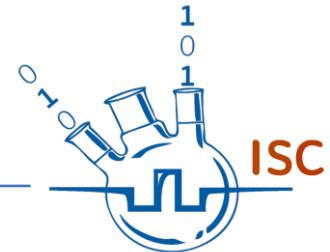


Prototype drug solubility isotherms - HPBCD



Shityakov et al., in progress

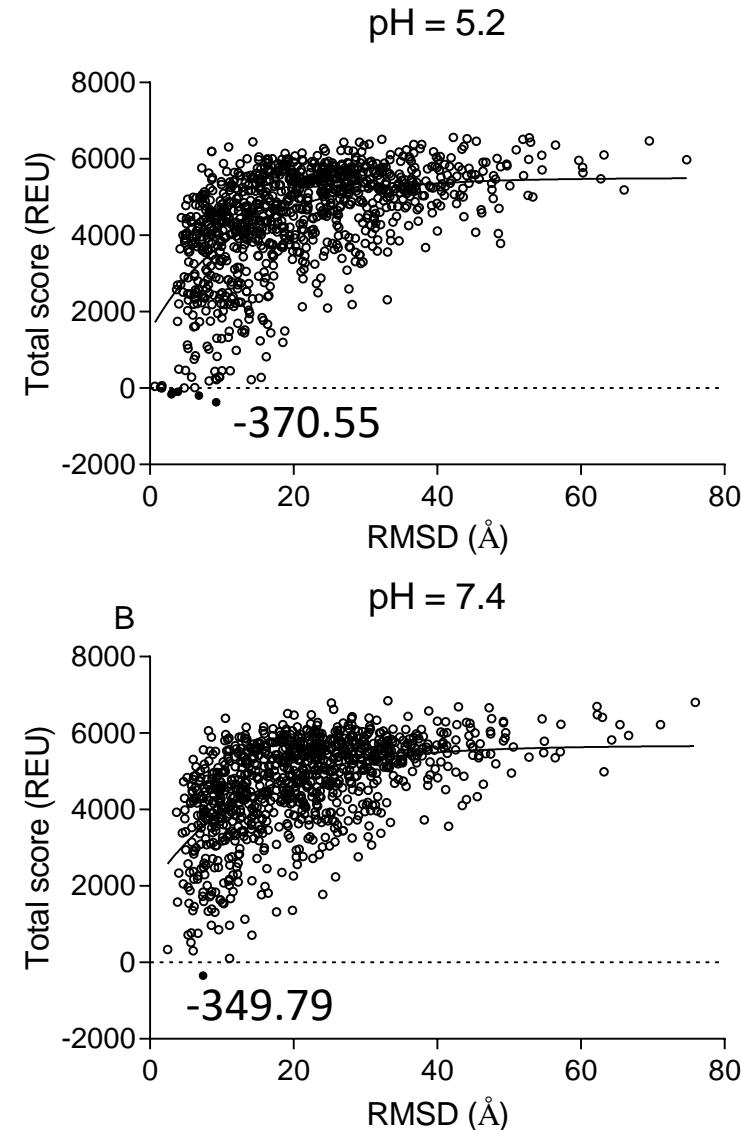
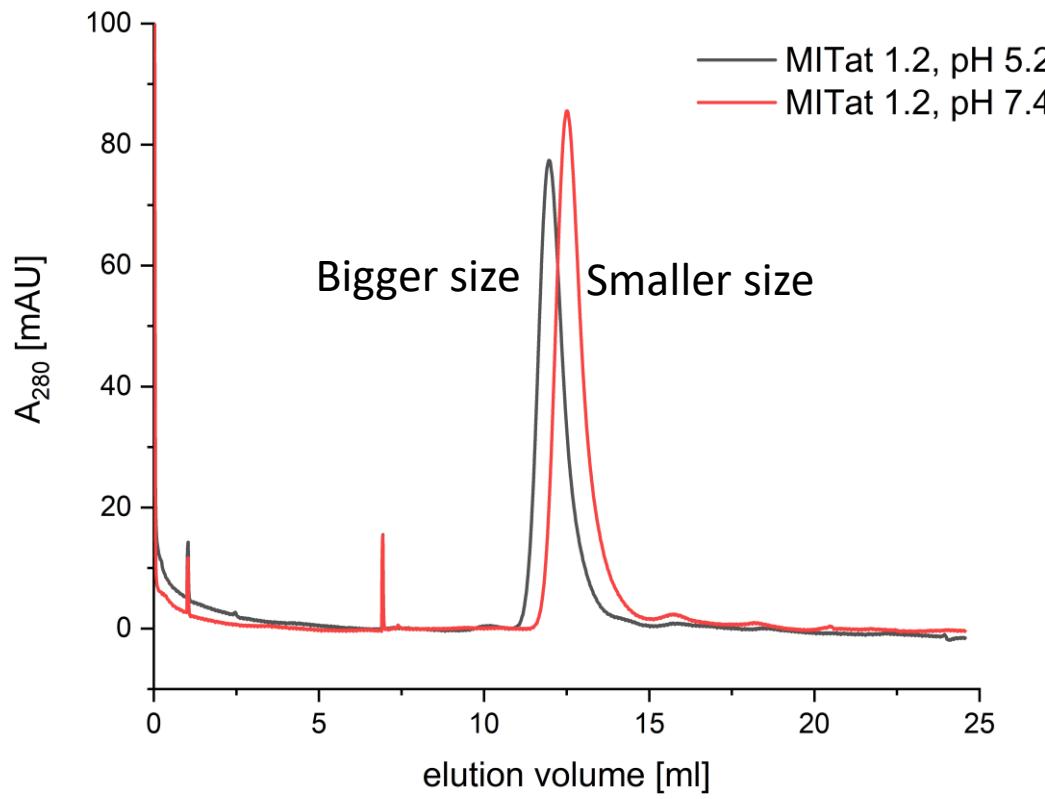
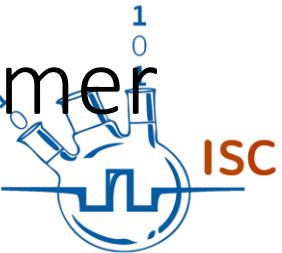
Cheminformatics and COVID-19 pandemics



Shityakov et al., in progress

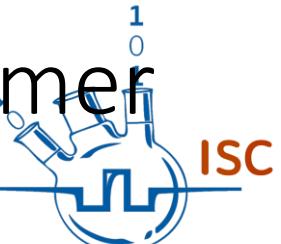
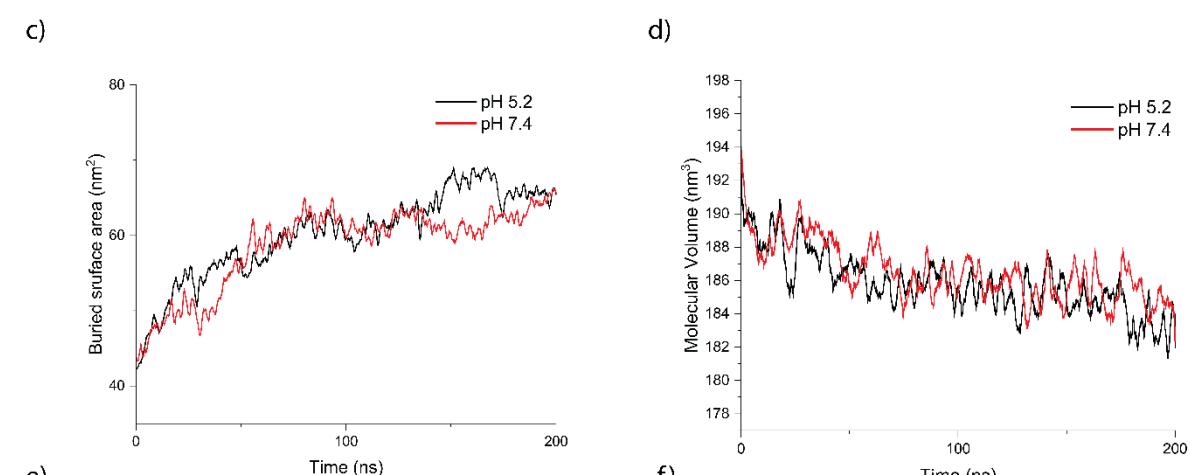
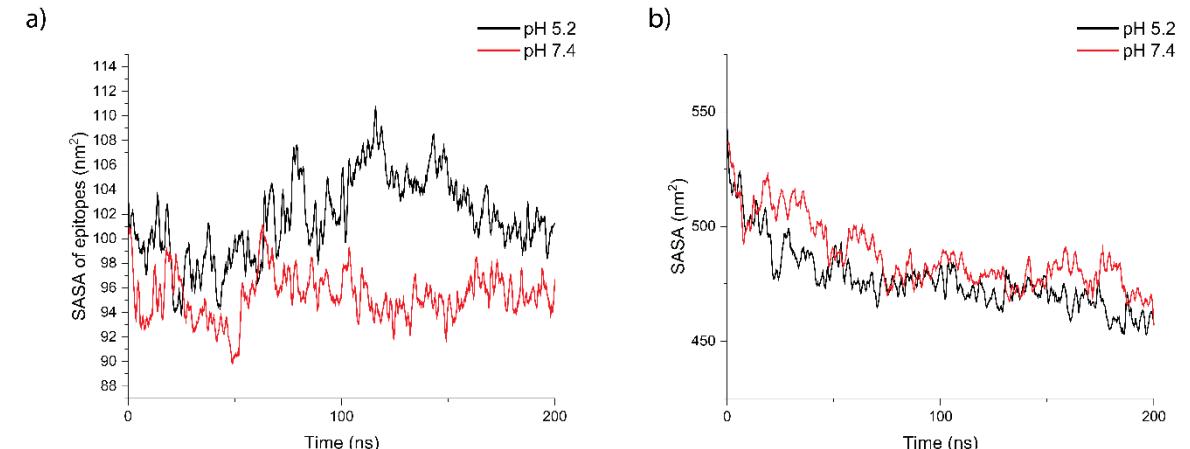
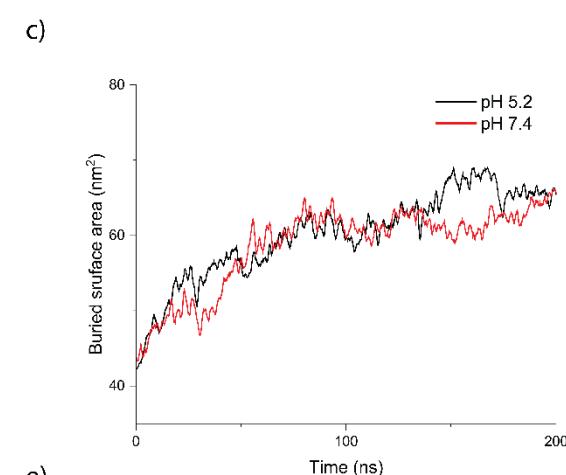
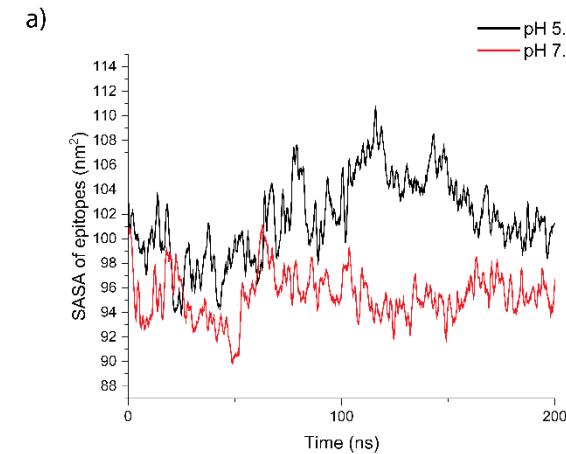
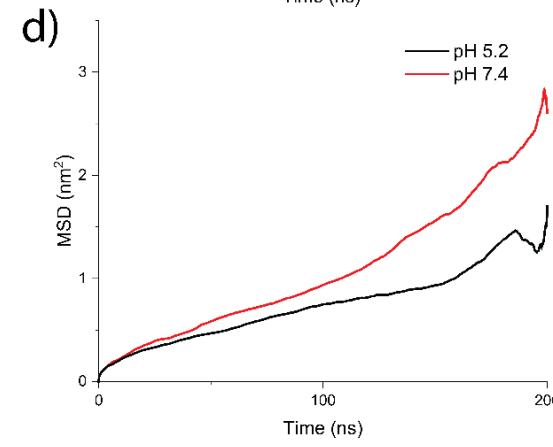
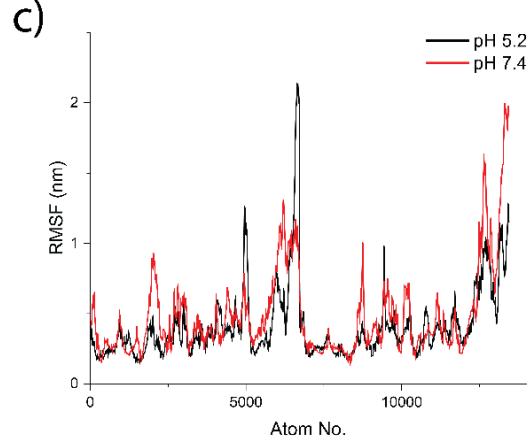
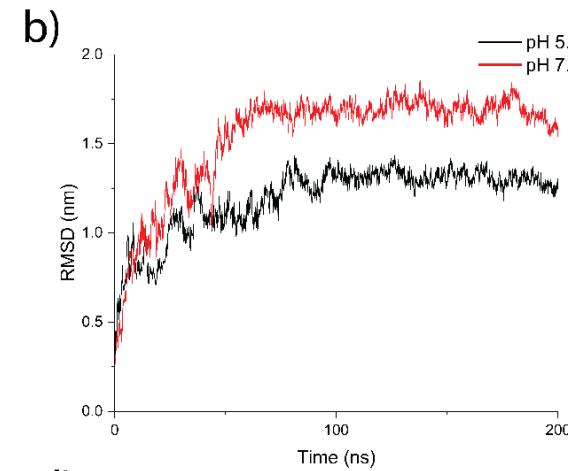
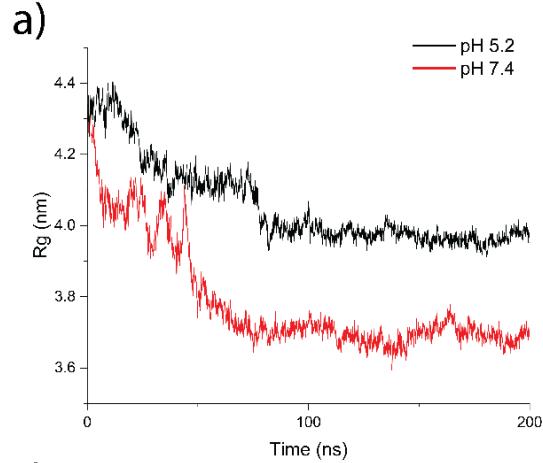
IT₃MOre than a
UNIVERSITY

pH-dependent conformational changes of VSG dimer



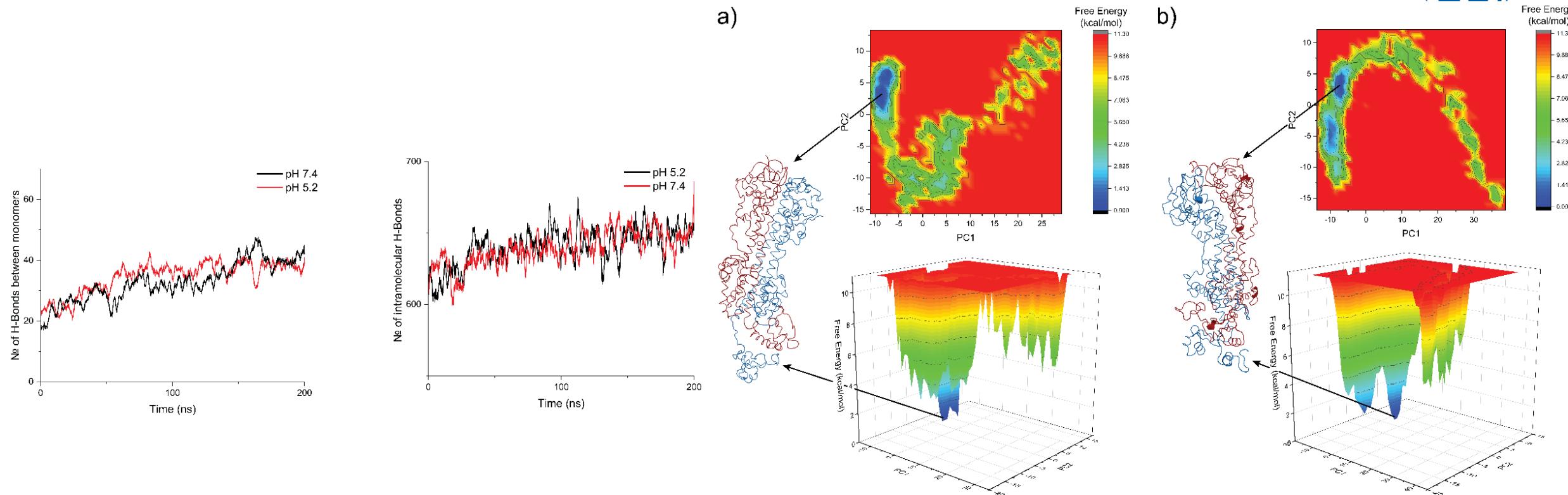
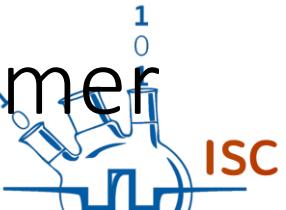


pH-dependent conformational changes of VSG dimer

IT₃MOre than a
UNIVERSITY

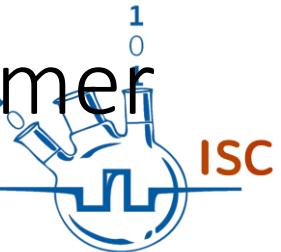
pH-dependent conformational changes of VSG dimer



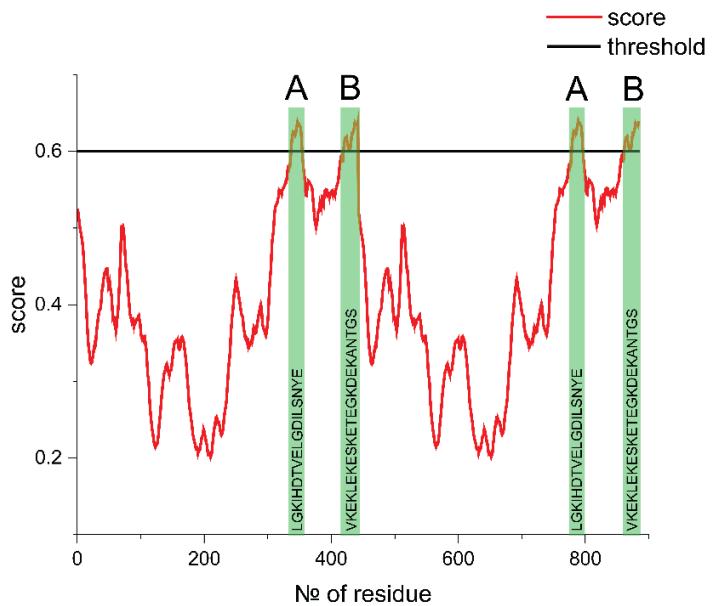
pH	SASA of epitopes, nm ²		BSA, nm ²	Molecular volume, nm ³	Number of H-bonds between monomers	Number of intramolecular H-bonds	Rg, nm
	SASA, nm ²	BSA, nm ²					
5.2	469.0±8.1	104.0±2.6	64.0±3.0	185.0±2.2	37.0±4.5	646.0±10.0	3.97±0.02
7.4	478.0±7.8	95.0±1.3	62.0±1.9	186.0±2.1	38.0±2.2	647.0±8.5	3.69±0.03



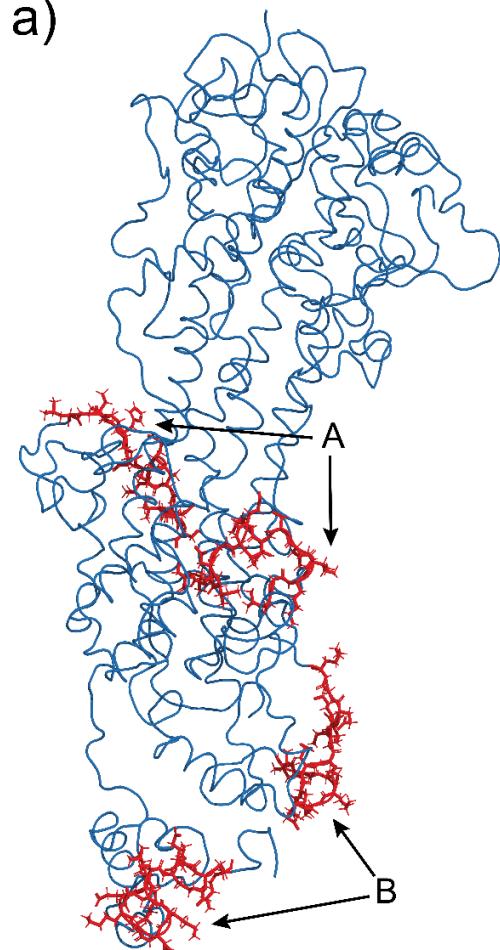
pH-dependent conformational changes of VSG dimer



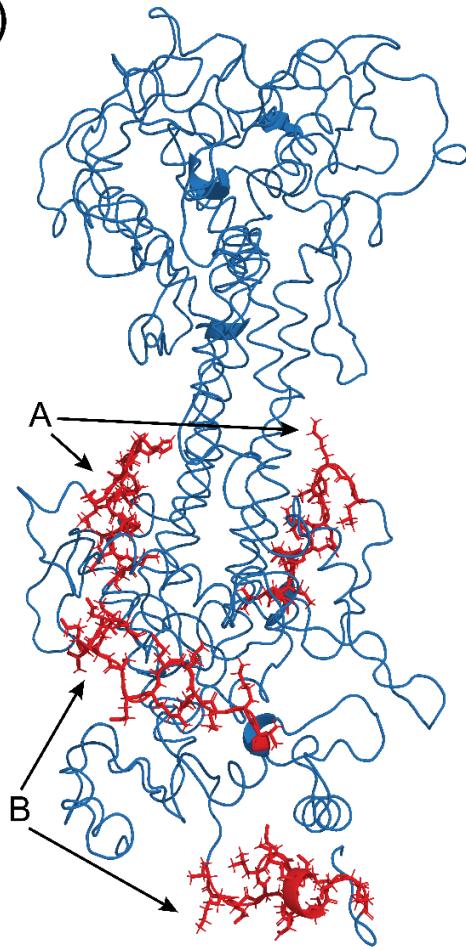
c)



a)



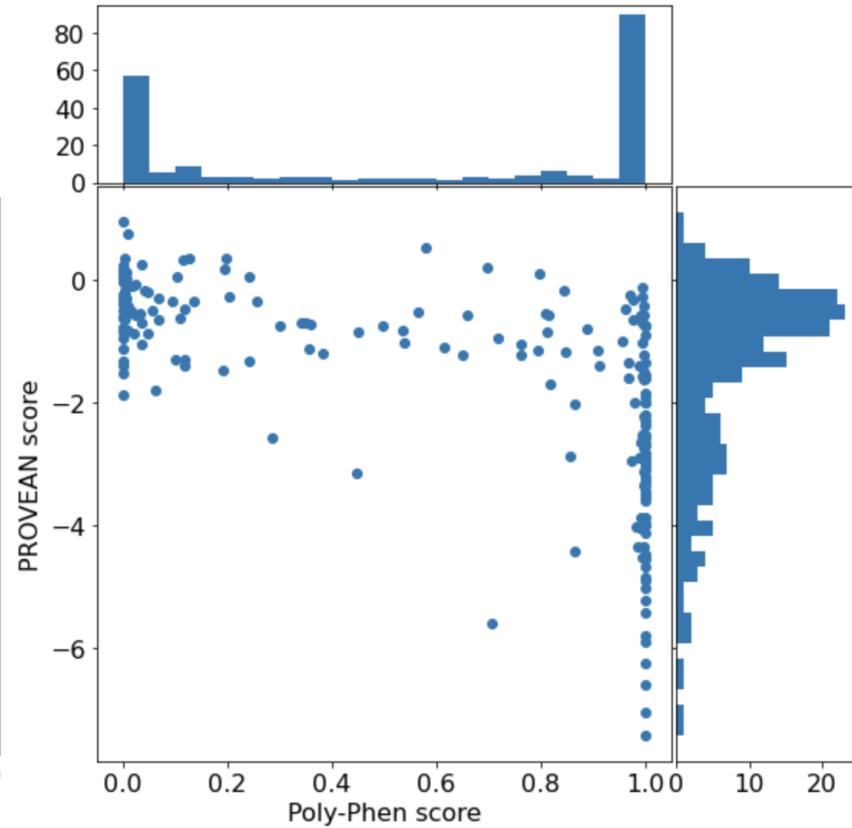
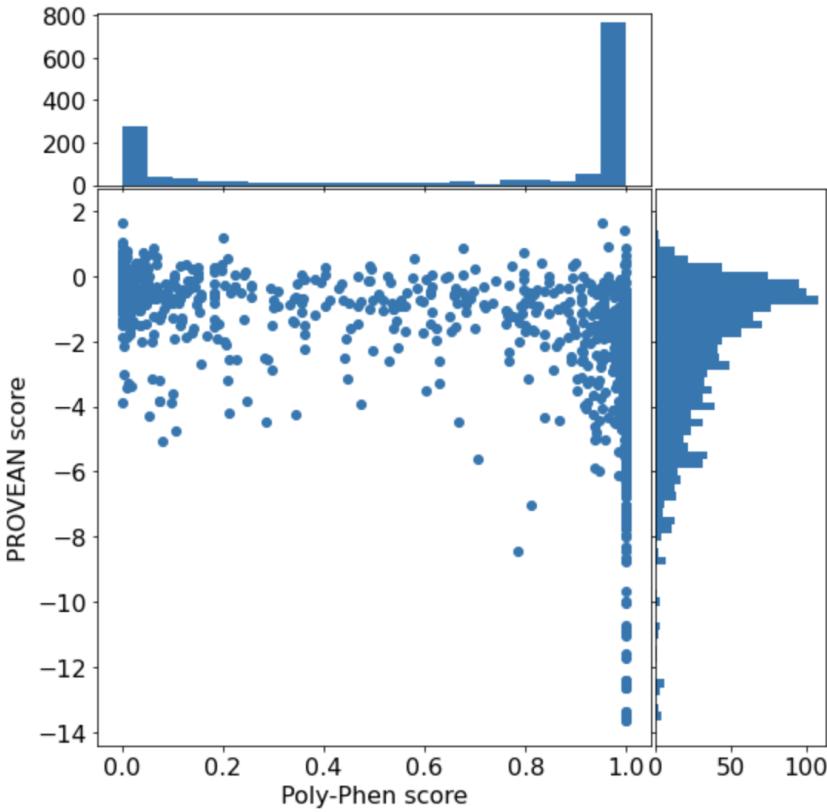
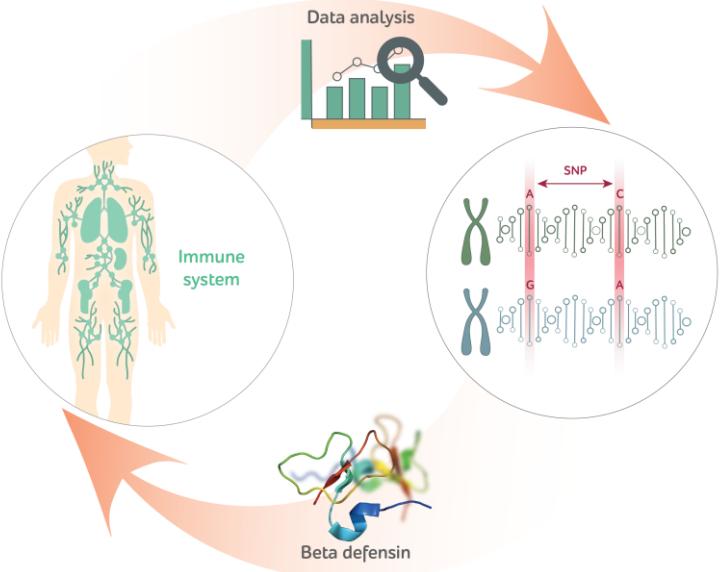
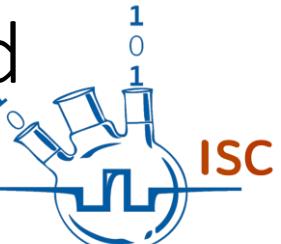
b)



pH	SASA, nm ²	SASA of epitopes, nm ²	BSA, nm ²	Molecular volume, nm ³	Number of H- bonds between monomers	Number of intramolecular H-bonds	Rg, nm	dG, kcal/mo
5.2	451	102	66	182	42	653	3.95	-353.37
7.4	459	96	65	183	39	686	3.71	-324.07

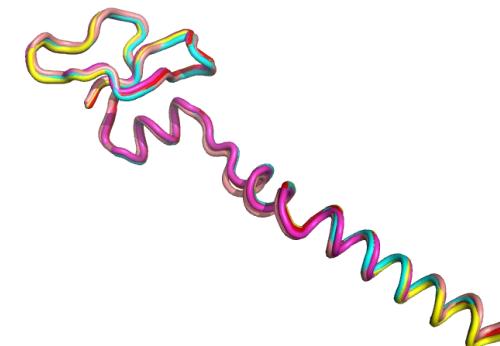
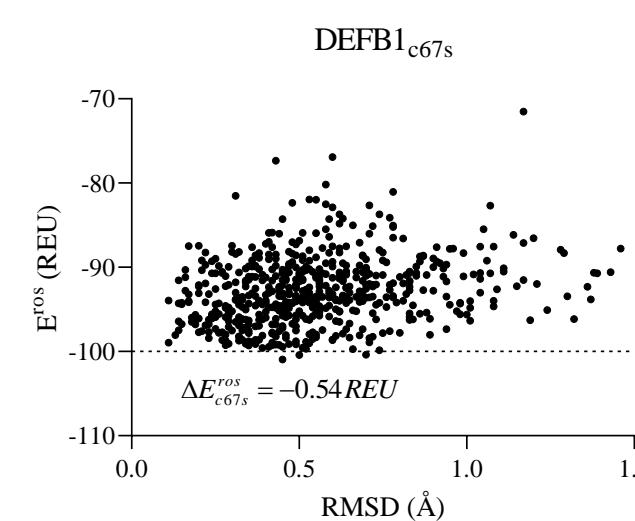
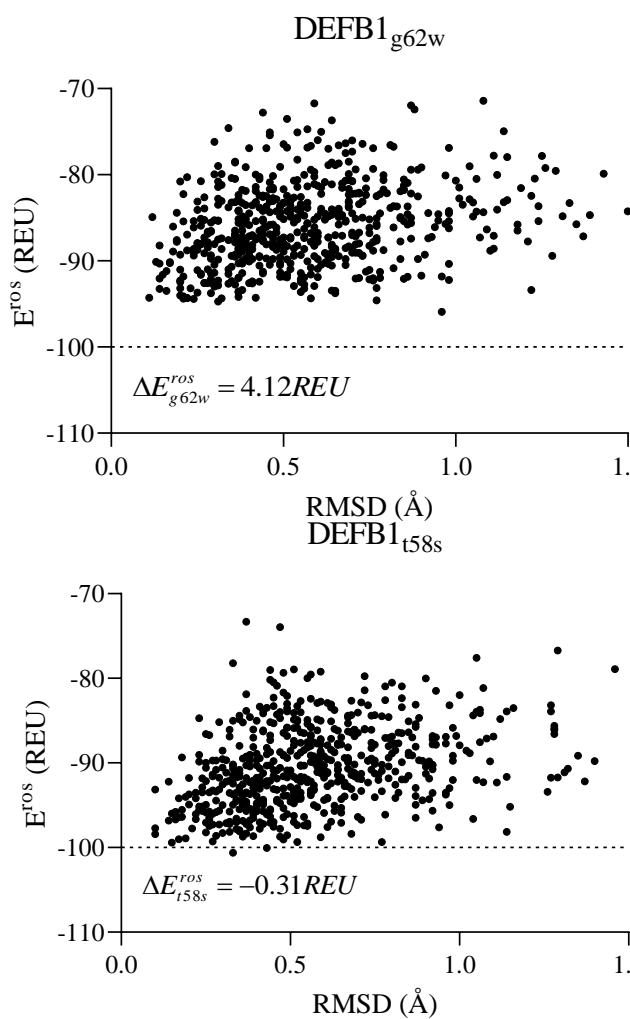
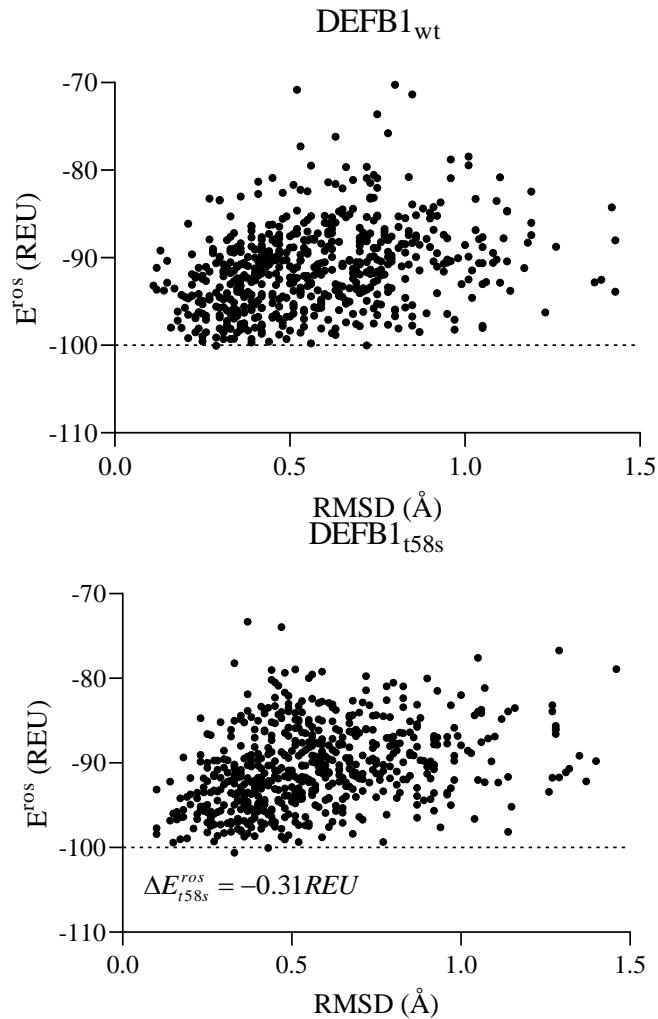


nsSNPs in DEFB1 gene reveal impact on protein-ligand binding sites



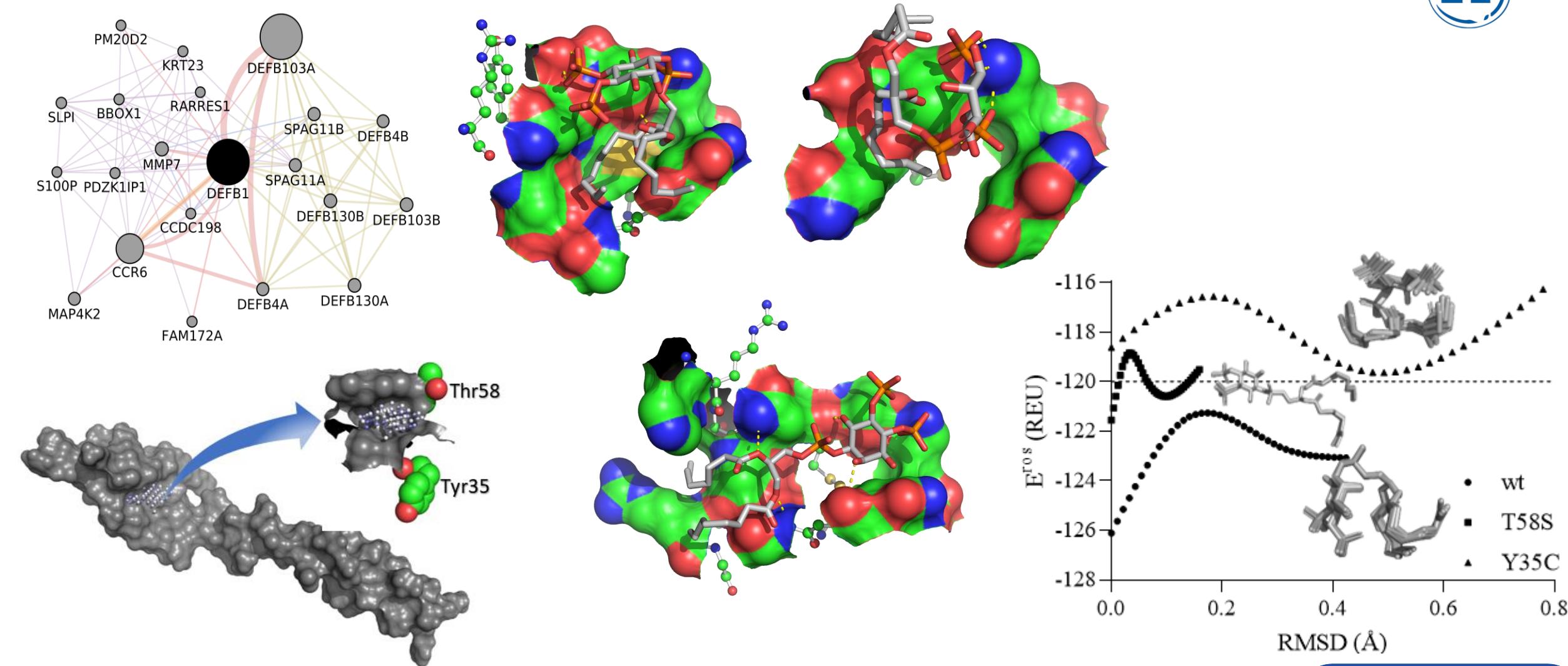
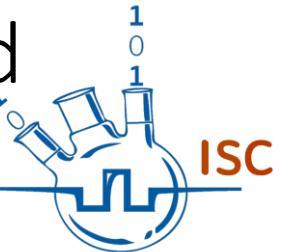


nsSNPs in DEFB1 gene reveal impact on protein-ligand binding sites



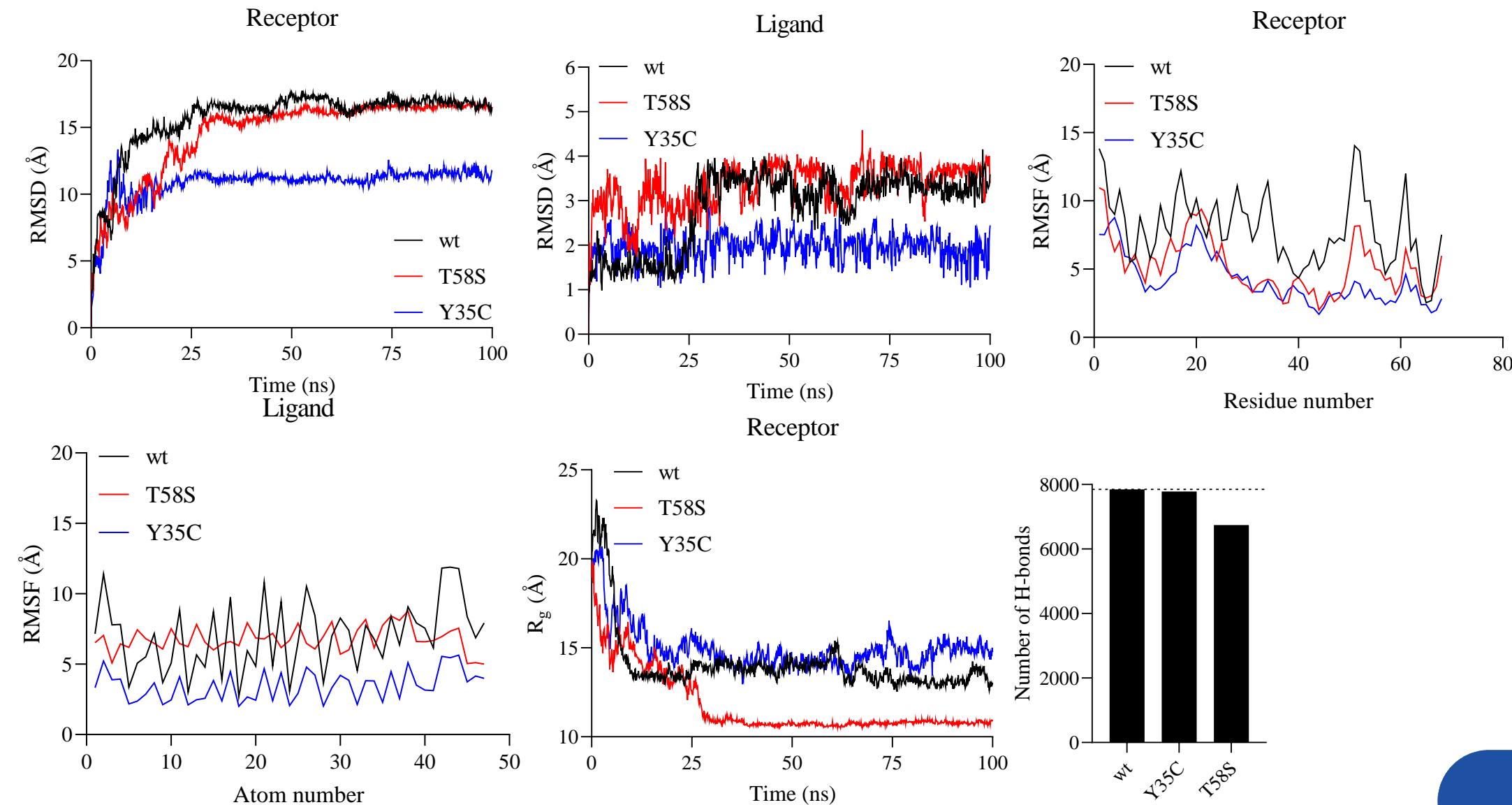


nsSNPs in DEFB1 gene reveal impact on protein-ligand binding sites



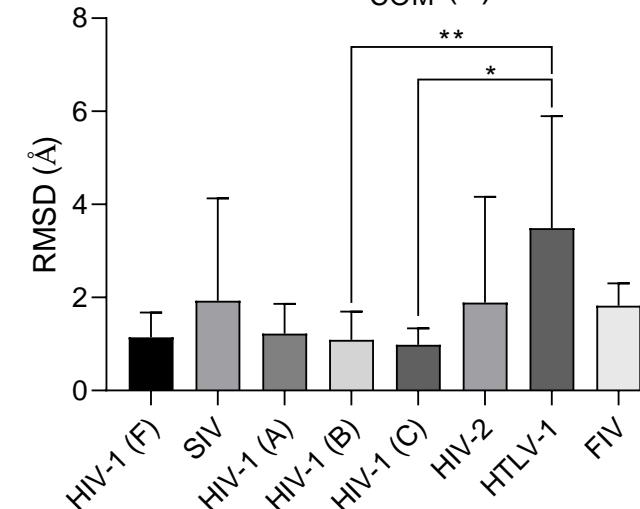
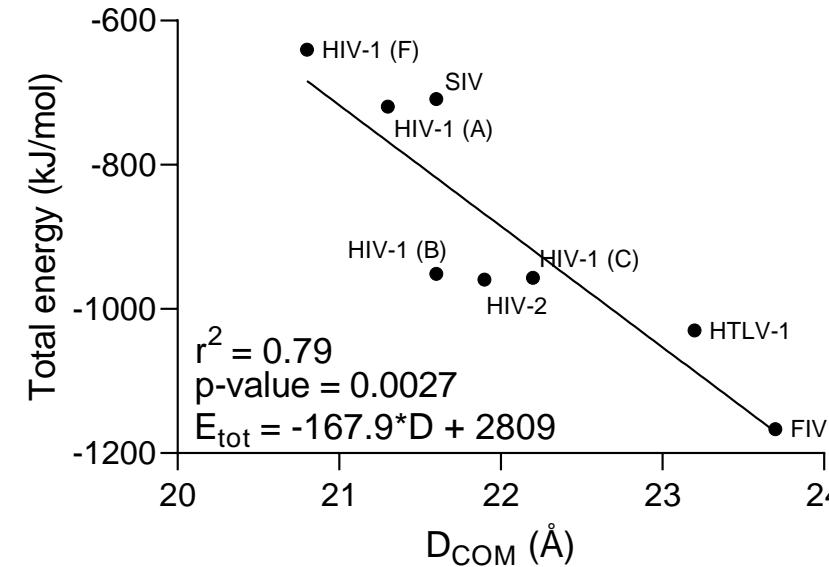
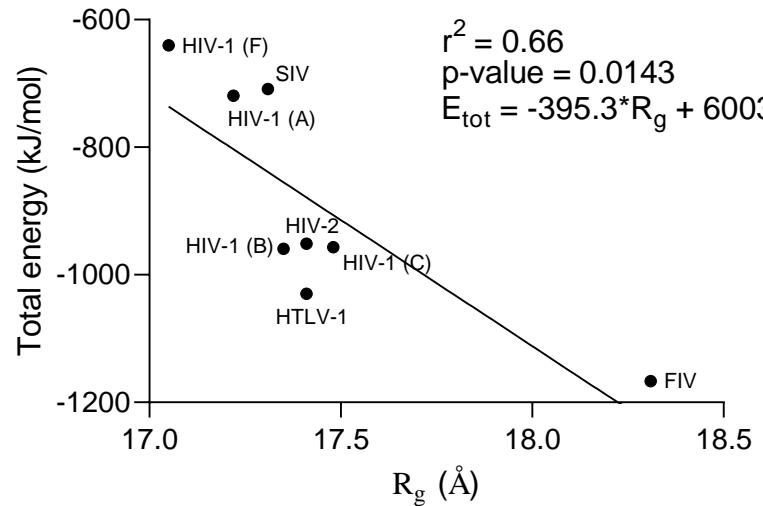
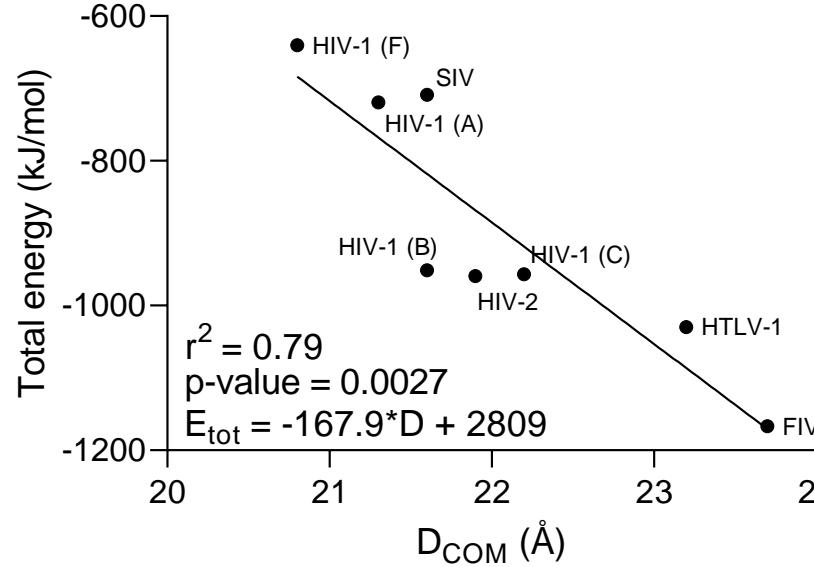
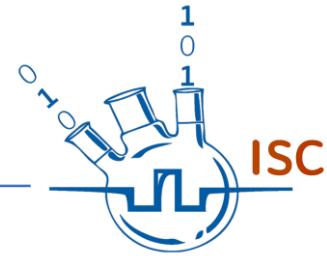


nsSNPs in DEFB1 gene reveal impact on protein-ligand binding sites



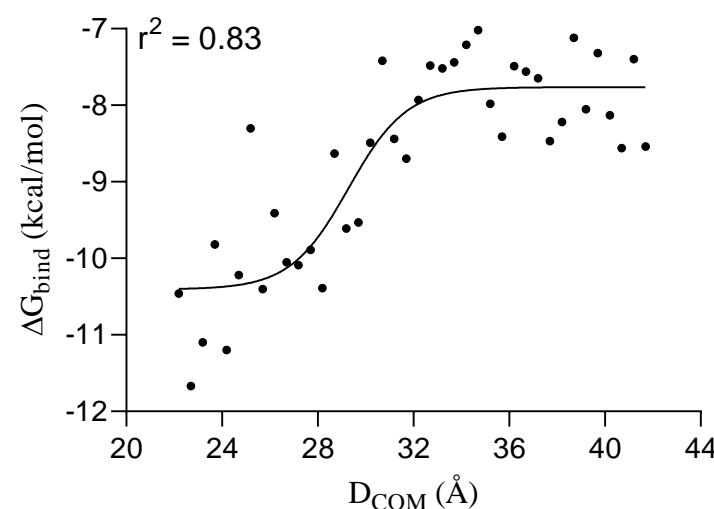
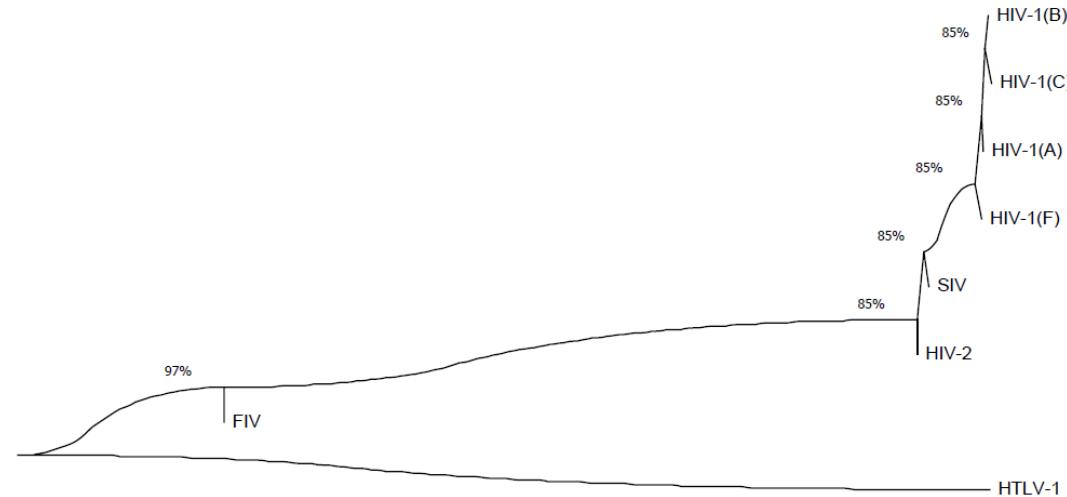
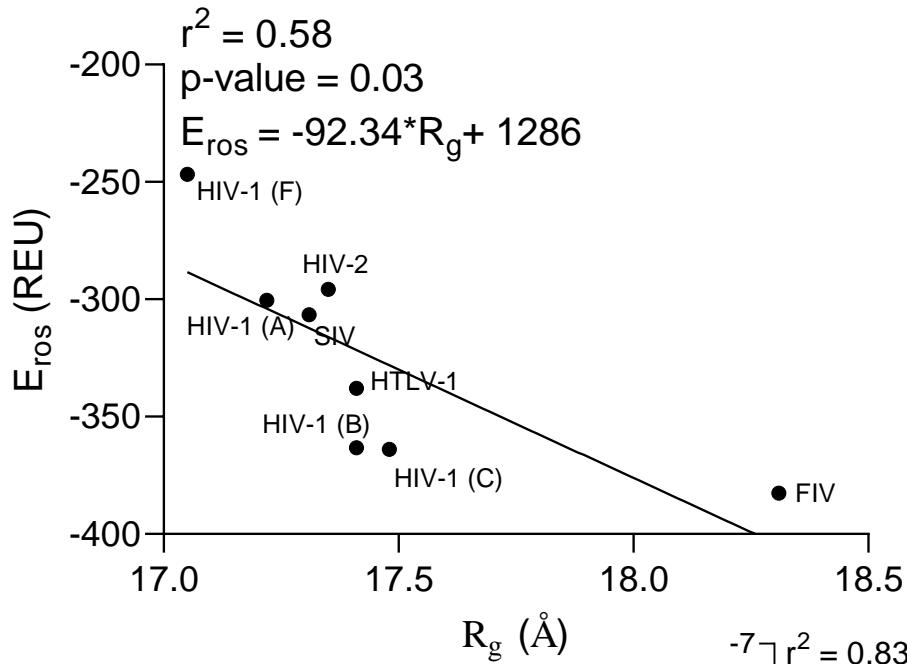
Dimerization and inhibition of retroviral proteases



Kruskal-Wallis test (p-value = 0.0086)
*:p-value = 0.04
**:p-value = 0.03

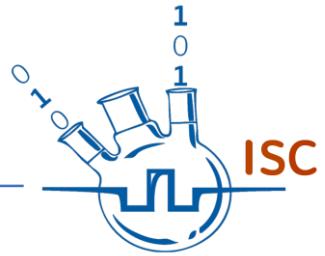


Dimerization and inhibition of retroviral proteases

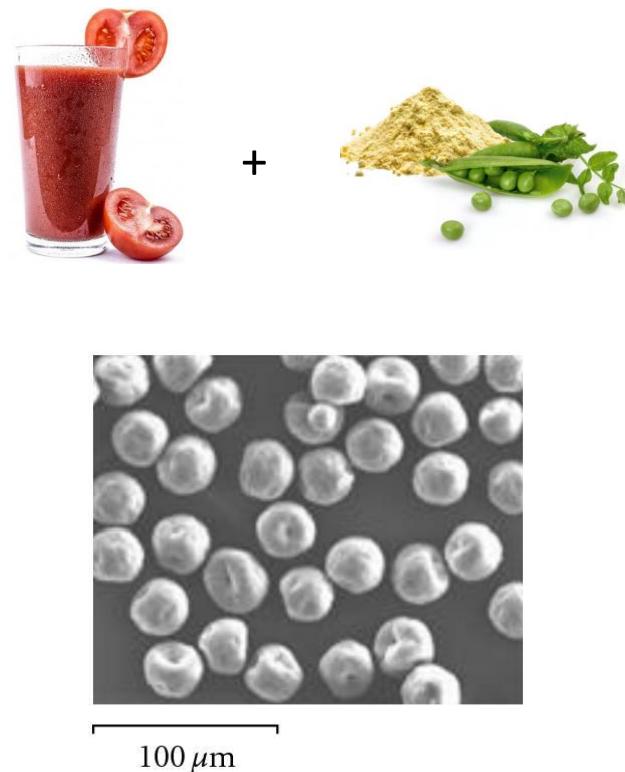
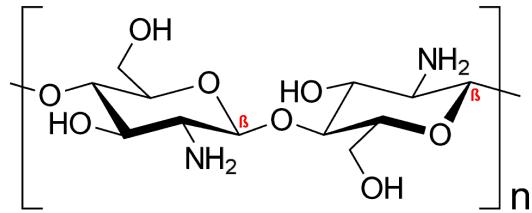
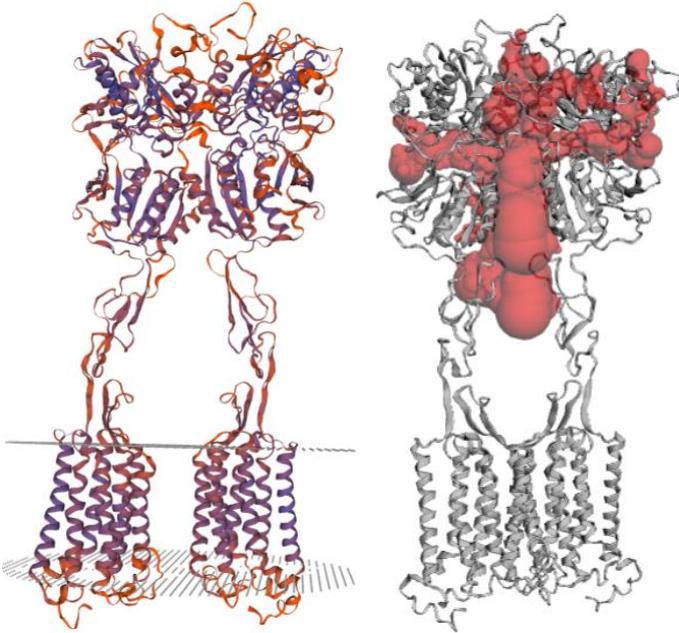




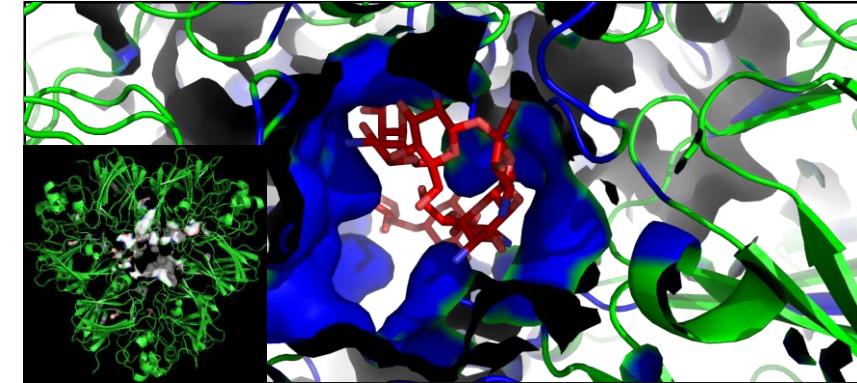
Chemoinformatics and foodinformatics



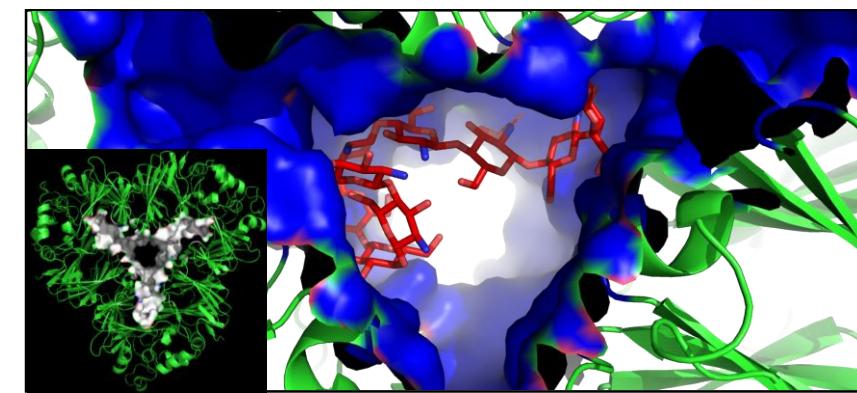
Artificial tongue biosensor



legumin-chitosan complex
 $\Delta G = -7.9 \text{ kcal/mol}$

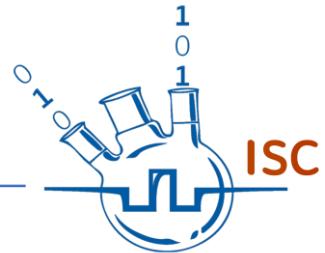


vicilin-chitosan complex
 $\Delta G = -7.5 \text{ kcal/mol}$





Chemoinformatics and triboinformatics

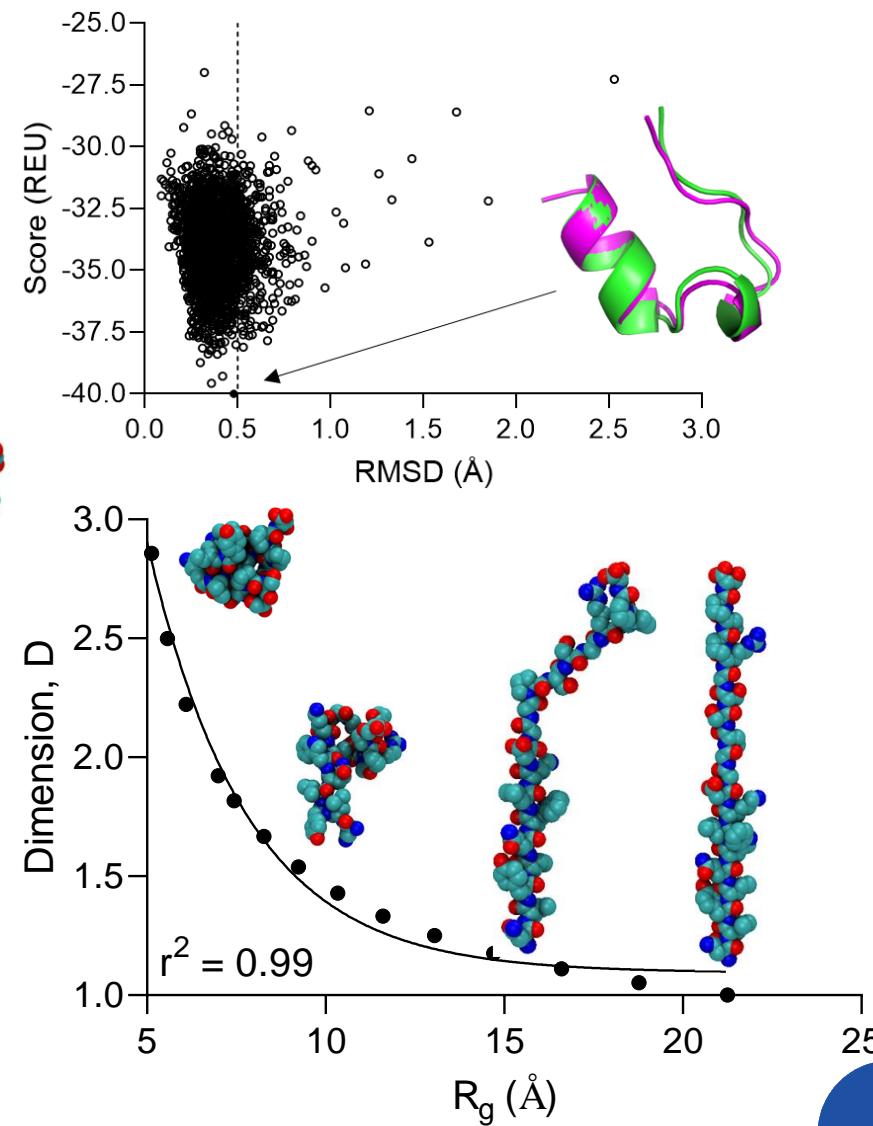
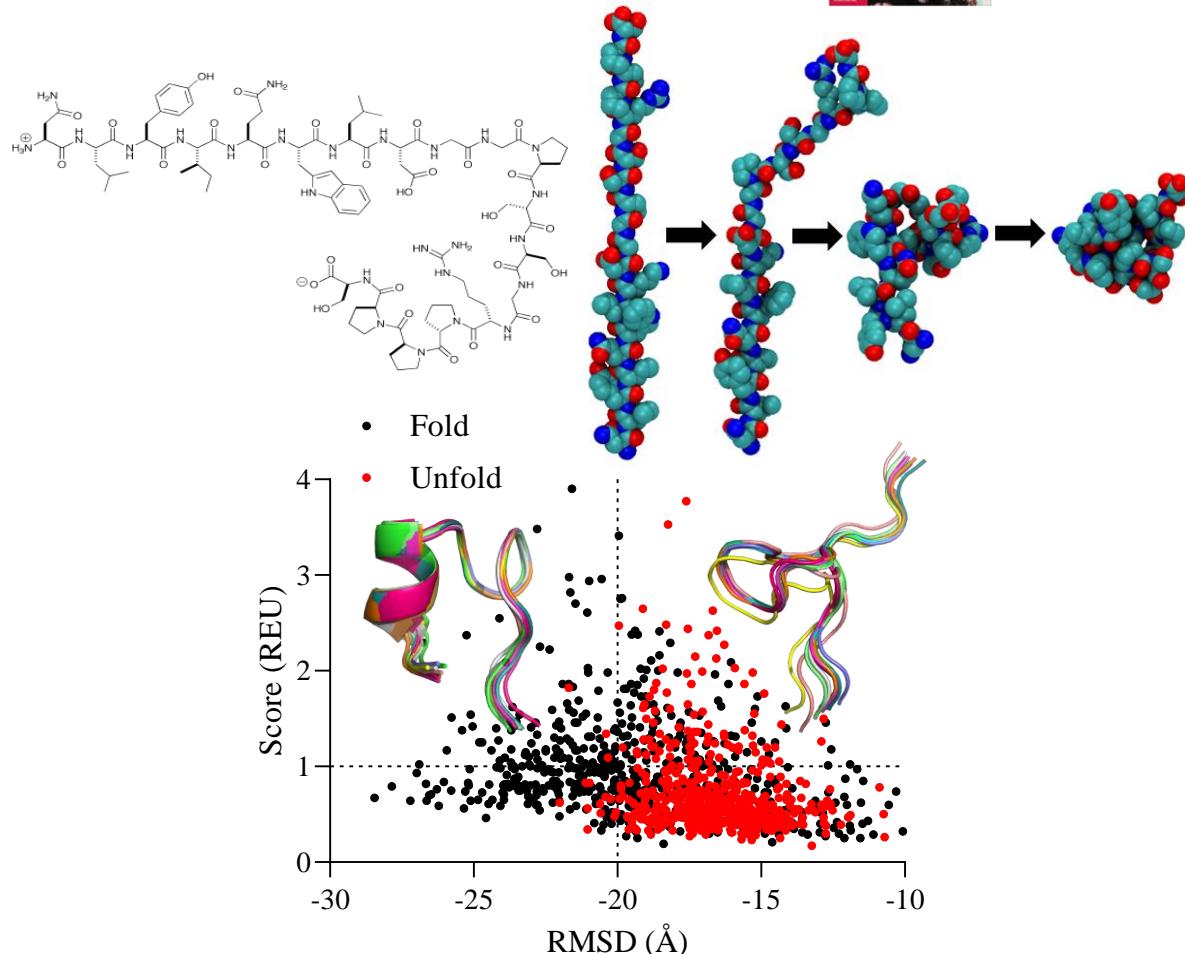


Research articles

Topological bio-scaling analysis as a universal measure of protein folding

Sergey Shityakov, Ekaterina V. Skorb and Michael Nosonovsky✉

Published: 13 July 2022 | <https://doi.org/10.1098/rsos.220160>



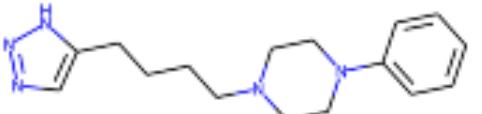
IT₃MOre than a
UNIVERSITY

Chemoinformatics and drug design

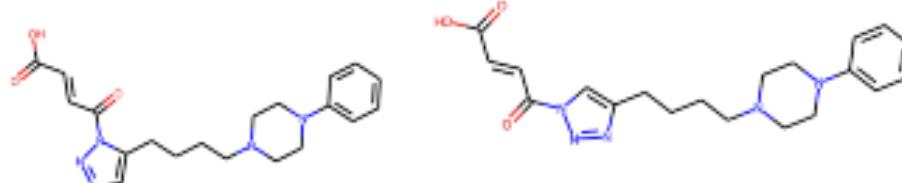


BBB project Dataset construction

The scaffold of interest:



The library of organic azides is used to generate a dataset of click-compounds



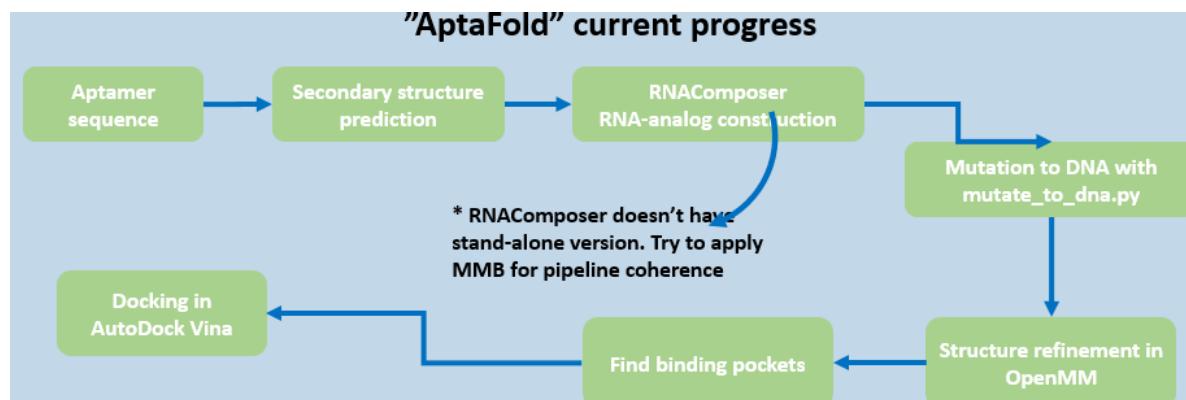
In total 18 844 molecules generated including both 1,4 and 1,5-isomers (9 422 radicals introduced)

Using RDkit following descriptors were calculated:

2D descriptors (13 descriptors)

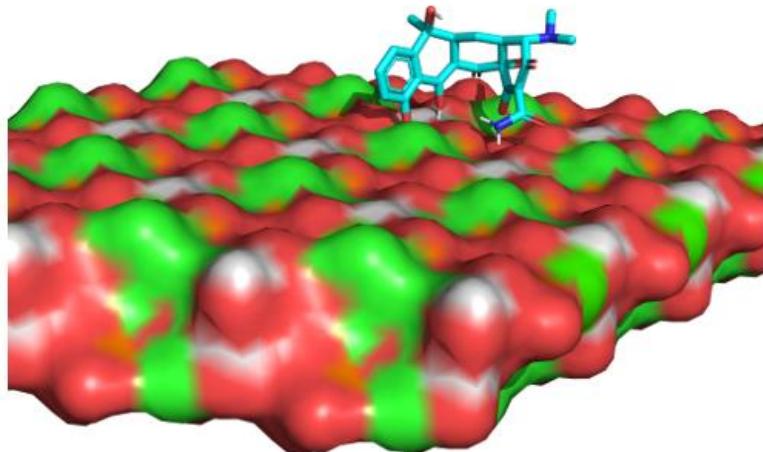
3D descriptors (5 descriptors)

- SASA
- V_m
- Eccentricity

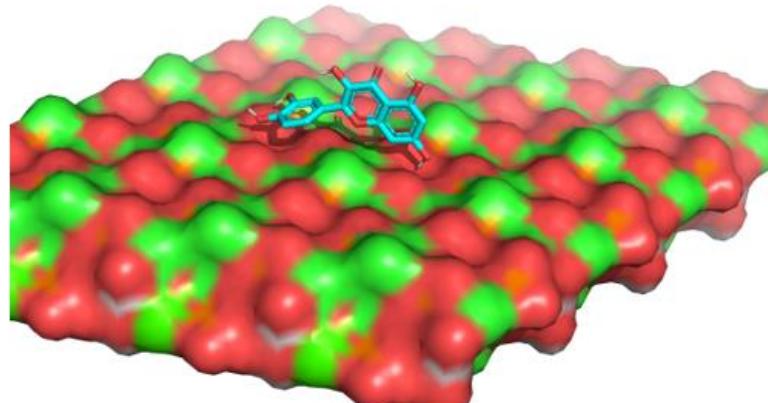


ITSMORE than a
UNIVERSITY

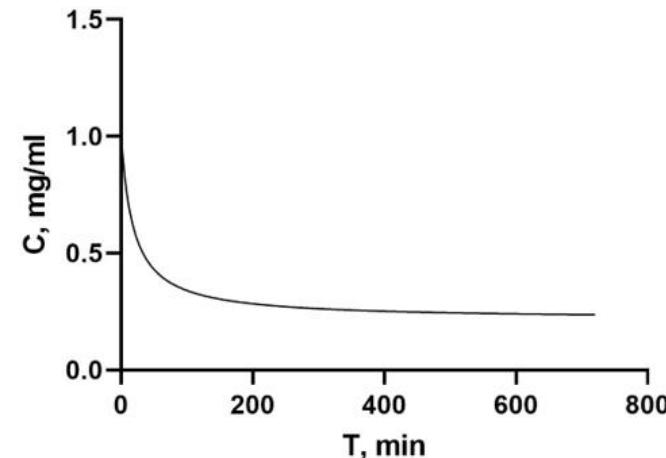
HA-Tetracycline



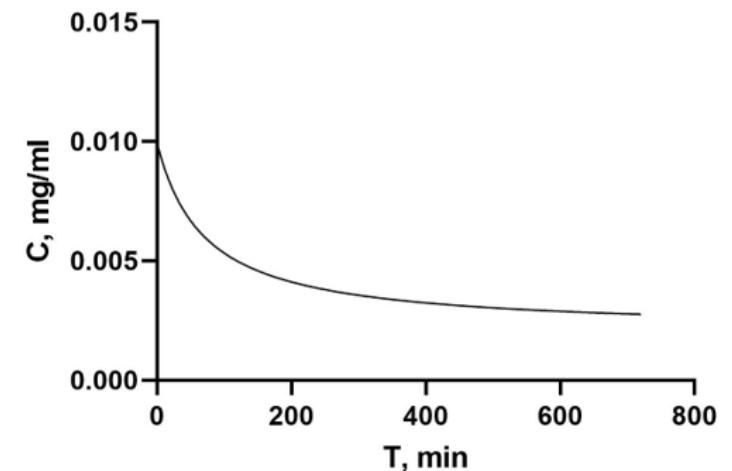
HA-Quercetin



Tetracycline Hydrochloride 1 mg/ml

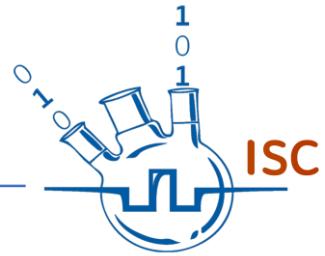


Tetracycline Hydrochloride 0.01 mg/ml

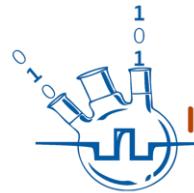




Chemoinformatics: areas of application



- Rational drug design and discovery (pharmacokinetics and pharmacodynamics)
- Protein-ligand interactions
- Host-guest interactions
- Prediction of physico-chemical properties
- Materials design
- Synthesis design



Thank you for your attention

