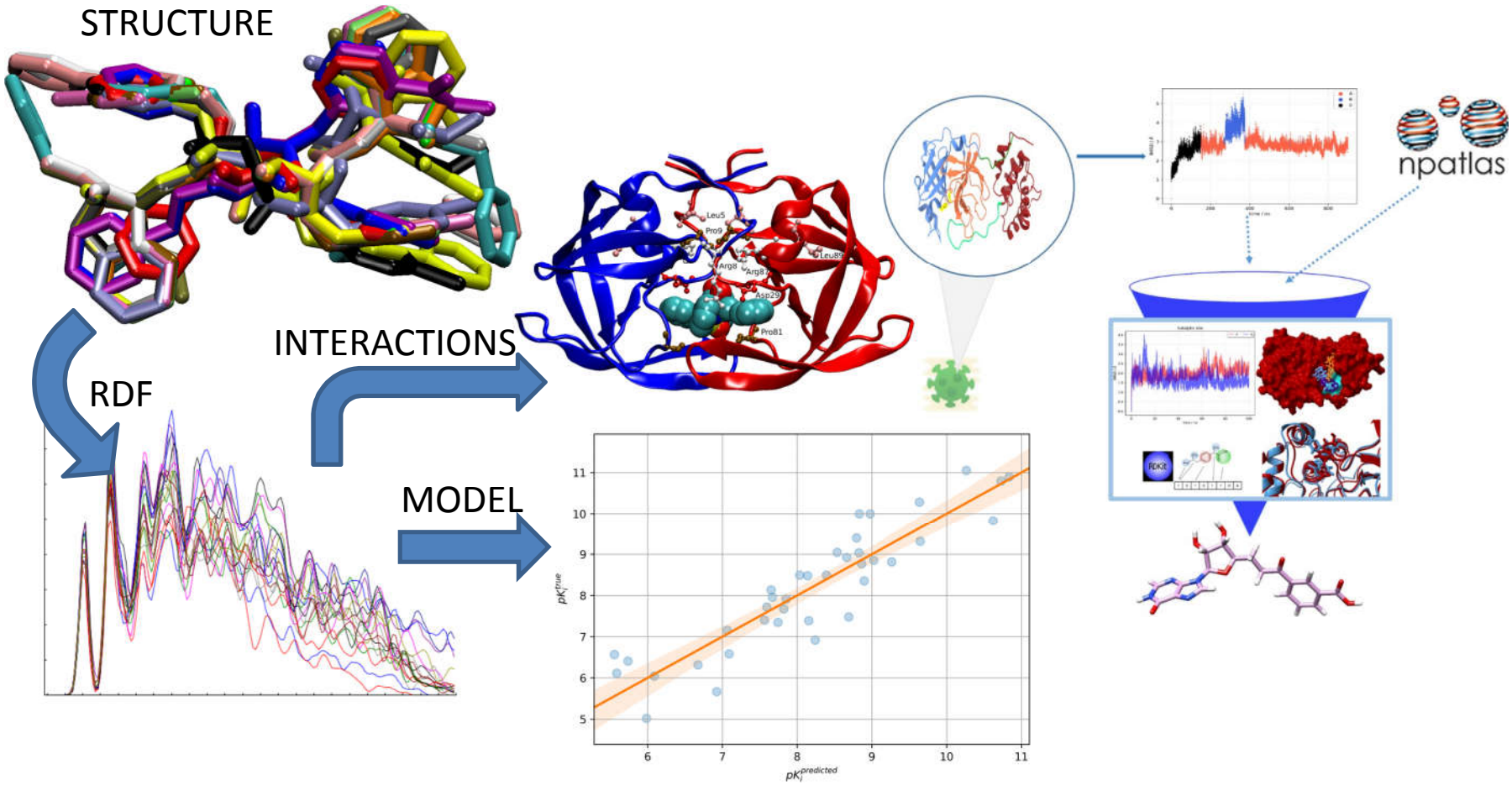
Computational chemistry artillery against viruses

Jurica Novak

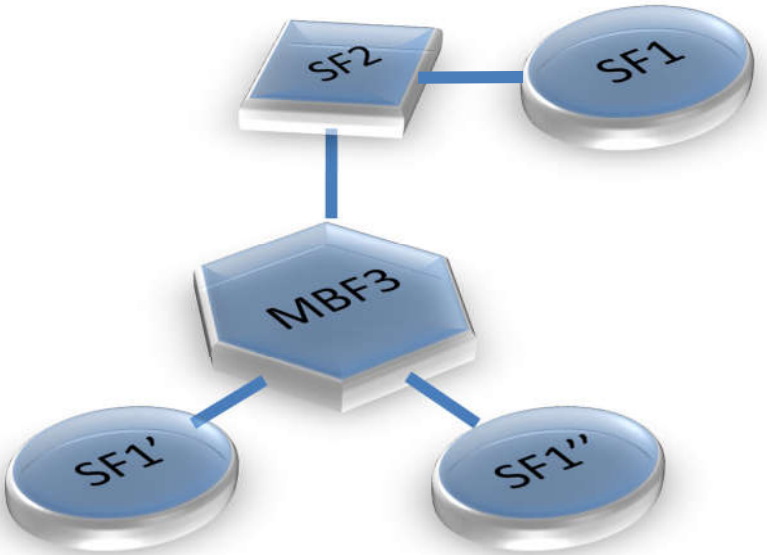
September 2022., Zagreb

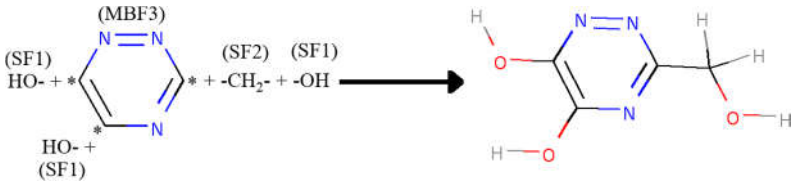
Introduction

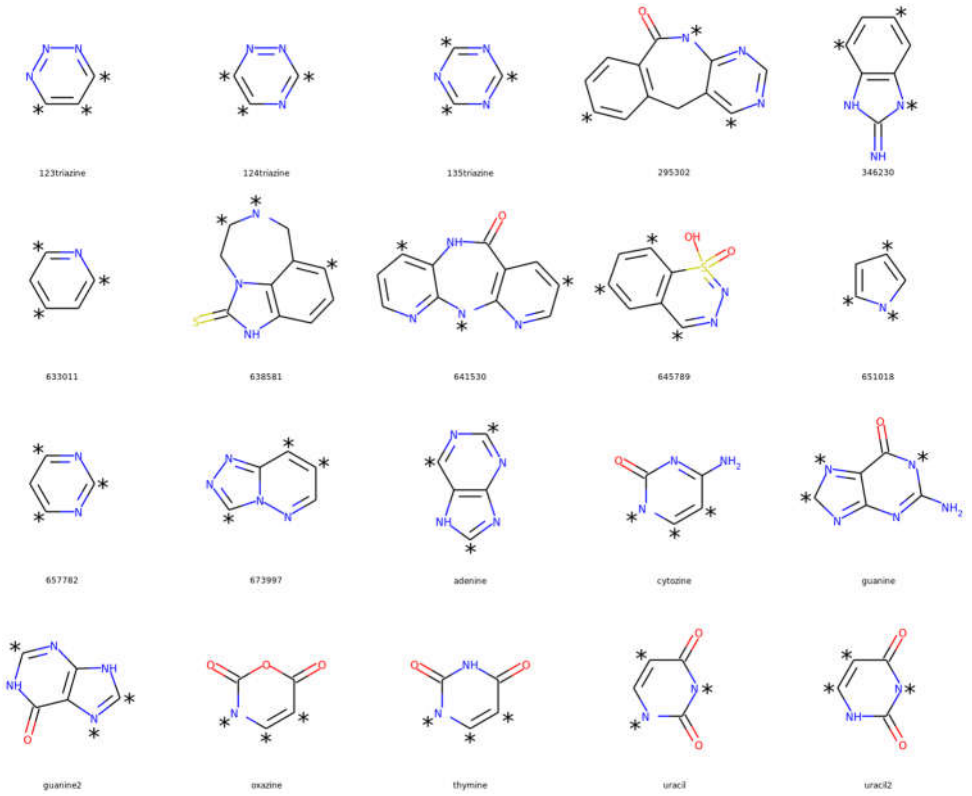


[1] JN *et al.* The Performance of Radial Distribution Function Based Descriptors in the Chemoinformatic Studies of HIV-1 Protease, FMC 12 (2020) 299
[2] JN *et al.* Novel radial distribution function approach in the study of point mutations: the HIV-1 protease case study, FMC 12 (2020) 1025
[3] JN *et al.* Can bacterial and fungal natural products stop SARS-CoV-2 virus? A virtual screening and molecular dynamics study of a natural product database, FMC 13 (2021) 363-378
[4] JN *et al.* Proposition of a new allosteric binding site for potential SARS-CoV-2 3CL protease inhibitors by utilizing molecular dynamics simulations and ensemble docking , JBSD

HIV







[5] JN *et al.* The Design of Compounds with Desirable Properties – the Anti-HIV Case Study (submitted to the Journal of Computational Chemistry)

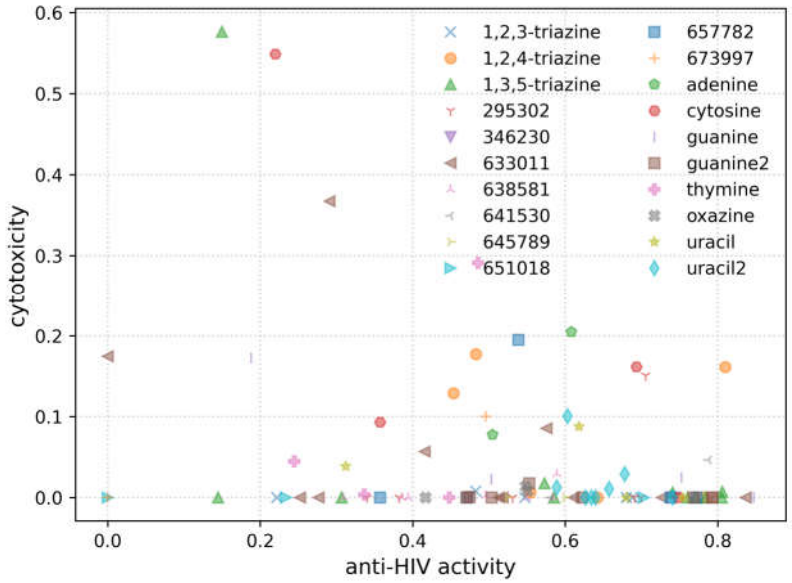
HIV

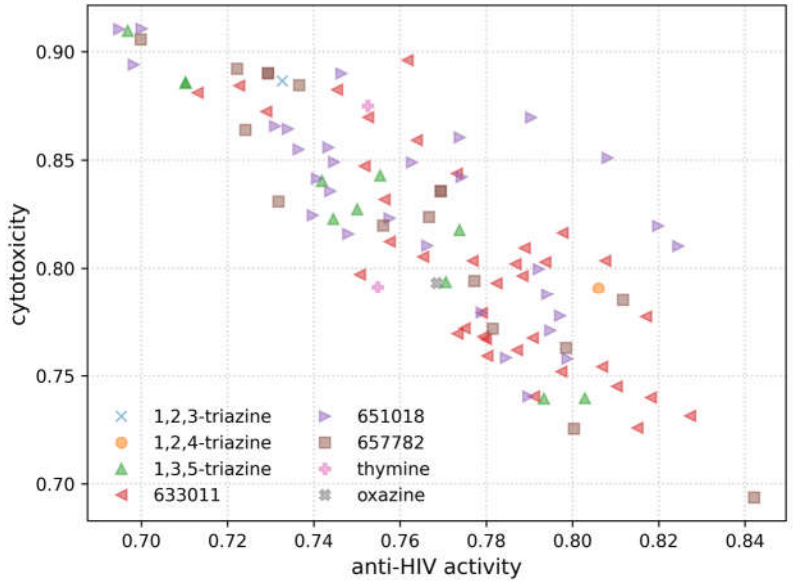
	MBF3	SF2	SF1	SF1’	SF1’’
Parent1	6	25	10	19	73
Parent2	2	39	46	24	6
Child1	2	25	10	19	73
Child2	6	39	46	24	6
Mutant1	2	25	10	19	79
Mutant2	6	39	46	60	6

[5] JN *et al.* The Design of Compounds with Desirable Properties – the Anti-HIV Case Study (submitted to the Journal of Computational Chemistry)

HIV

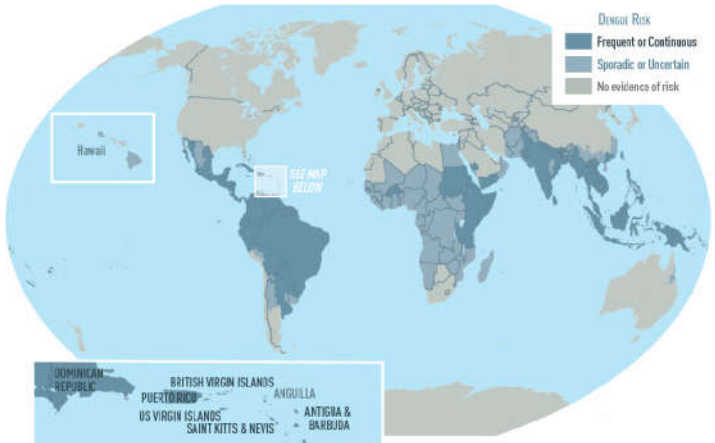
$$g = \left(\prod_{i=1}^{n=18} p_{anti-HIV,i} \prod_{j=1}^{m=12} p_{tox,j} \right)^{\frac{1}{n+m}}$$



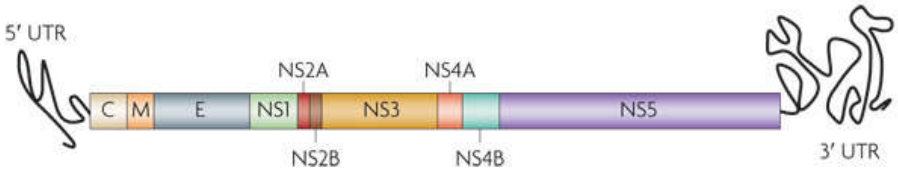


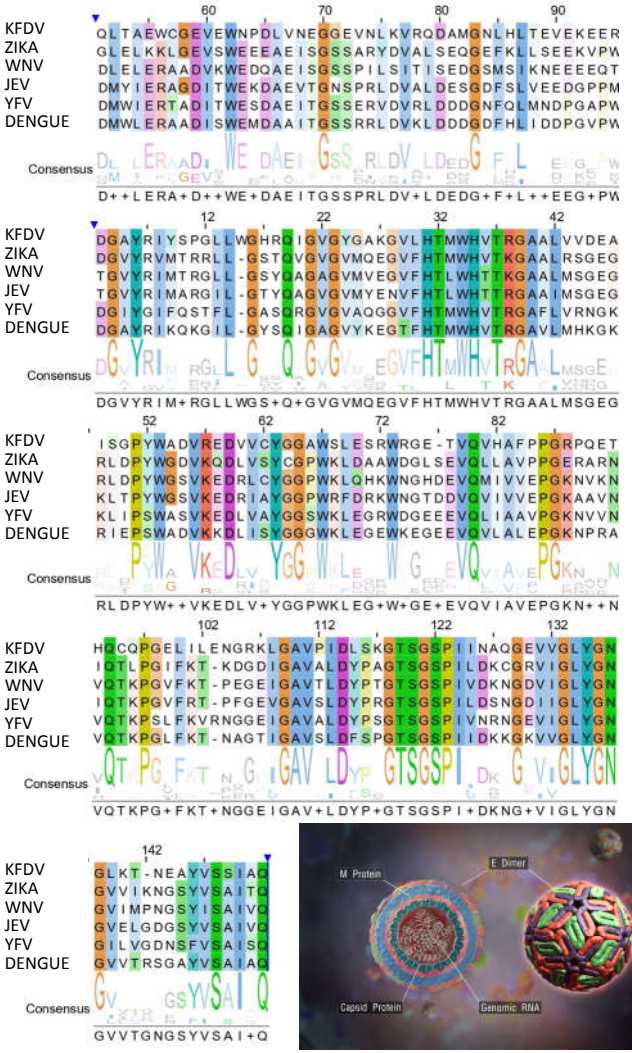
[5] JN *et al.* The Design of Compounds with Desirable Properties – the Anti-HIV Case Study (submitted to the Journal of Computational Chemistry)

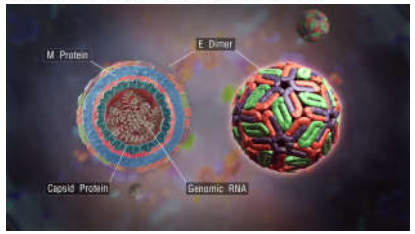
Flavivirus


Dengue Risk
Frequent or Continuous
Sporadic or Uncertain
No evidence of risk
Hawaii
SEE MAP BELOW
DOMINICAN REPUBLIC
PUERTO RICO
BRITISH VIRGIN ISLANDS
ANGUILLA
US VIRGIN ISLANDS
SAINT KITTS & NEVIS
ANTIGUA & BARBUDA
GUADELOUPE
DOMINICA
MARTINIQUE
SAINT LUCIA
ARUBA
CURAÇAO
SAINT VINCENT & THE GRENADINES
GRENADA
BARBADOS
TRINIDA & TOBAGO

CDC

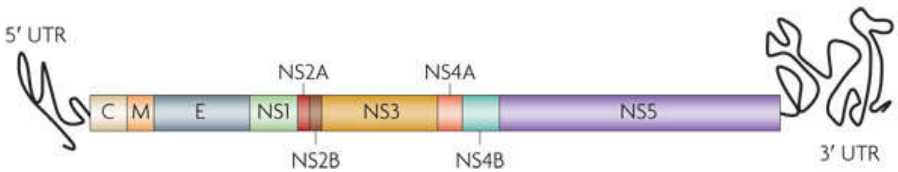

5′ UTR
C
M
E
NS1
NS2A
NS2B
NS3
NS4A
NS4B
NS5
3′ UTR

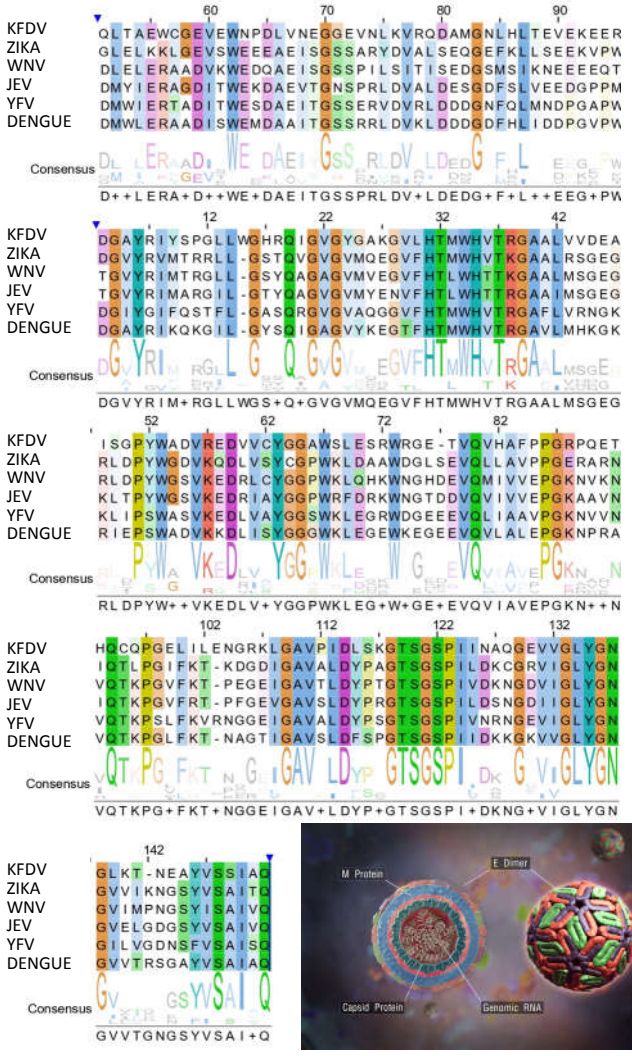

KFDV
ZIKA
WNV
JEV
YFV
DENGUE
Consensus
D++LERA+D++WE+DAEITGSSPRLDV+LDEDG+F+L++EEG+PW
DGVYRIM+RGLLWGS+Q+GVGVMQEGVFHTMWHVTRGAALMSGEG
RLDPYW++VKEDLV+YGGPWKLEG+W+GE+EVQVIAVEPGKN++N
GVVTGNGSYVSAI+Q

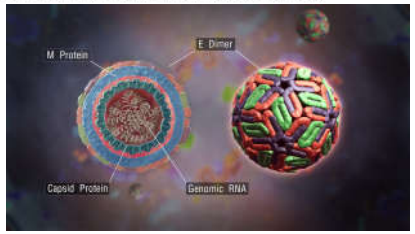

M Protein
E Dimer
Capsid Protein
Genomic RNA

Flavivirus

	KFDV NS3	ZIKA NS3	WNV NS3	JEV NS3	YFV NS3	DENV NS3
KFDV NS3	100	48.98	46.26	44.22	50.68	53.06
ZIKA NS3	48.98	100	69.33	62.67	56.67	57.33
WNV NS3	46.26	69.33	100	76	57.33	56.67
JEV NS3	44.22	62.67	76	100	54.67	52
YFV NS3	50.68	56.67	57.33	54.67	100	60.93
DENV NS3	53.06	57.33	56.67	52	60.93	100

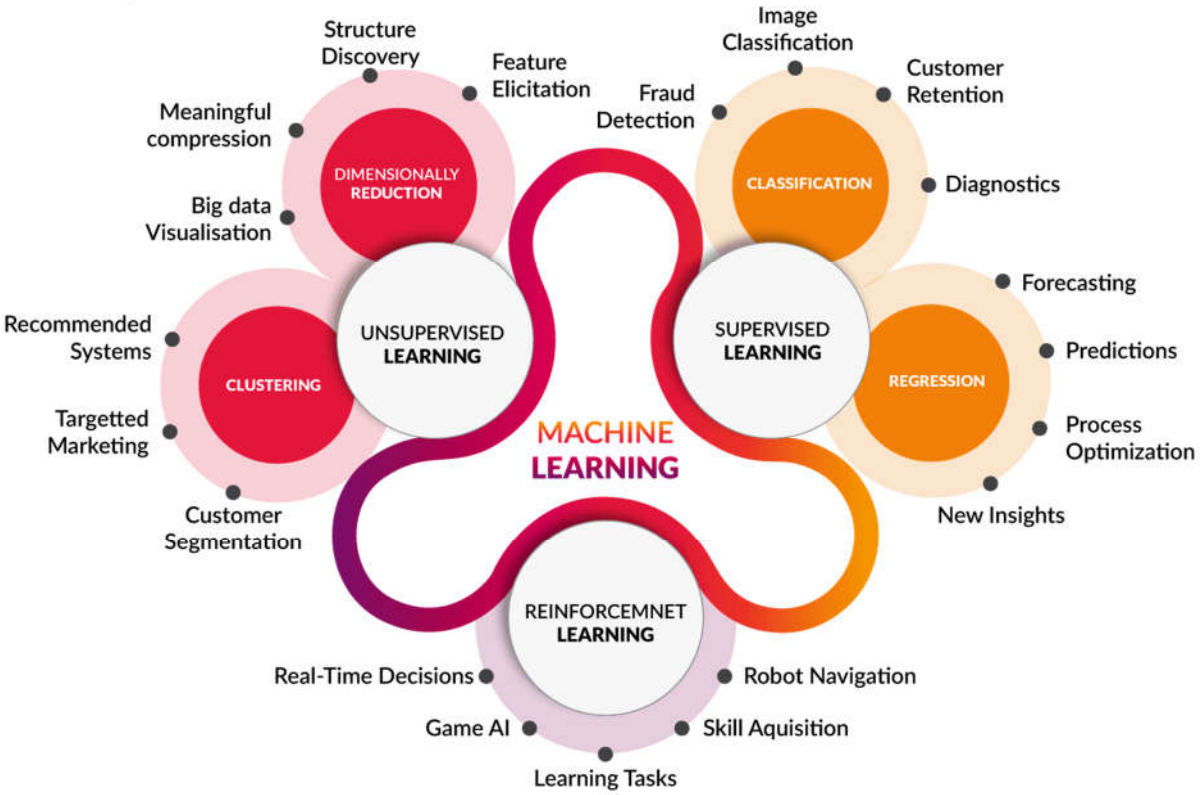




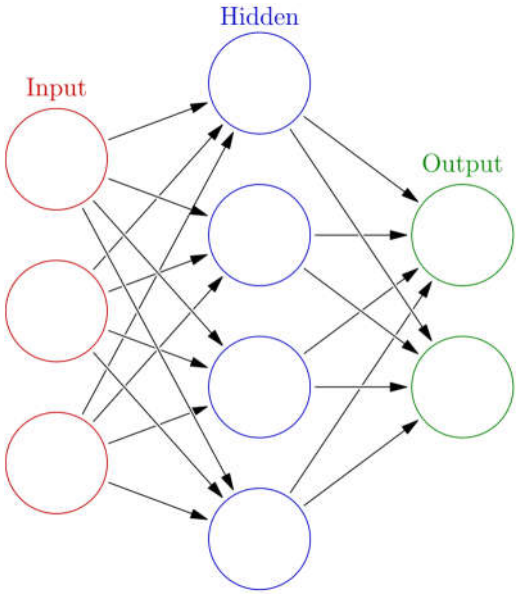


Machine learning

- a technique which uses computers to discover patterns or information about data



https://augnitive.com/introduction-to-machine-learning/



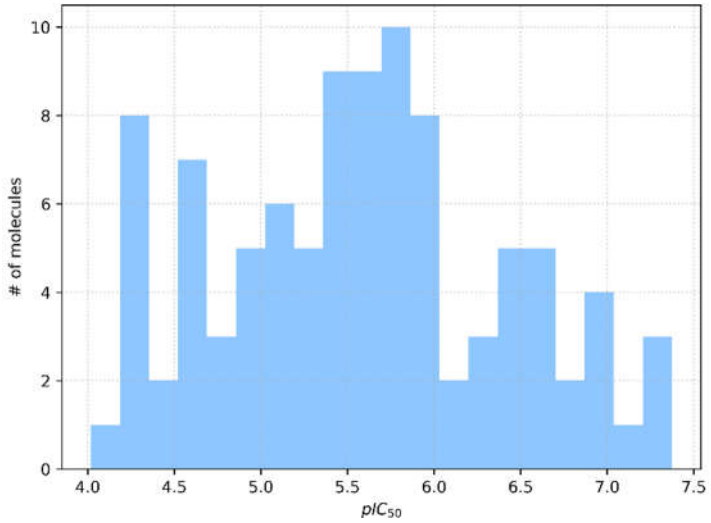
ML scheme

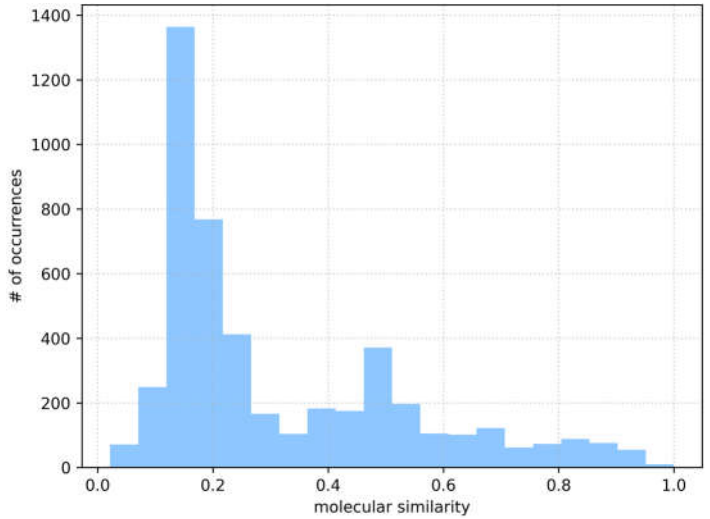
ChEMBL	descriptors	preparation	CV grid search	Sequential Feature Elimination	Train model evaluation
Experimental bioactivities SMILES Removal of mixtures Removal of duplicates	Mordred, RDKit 2D/3D SMILES to 3D Geometry optimization	Removal of highly correlated descriptors Standardization RReliefF	DNN RF SVM XGB		

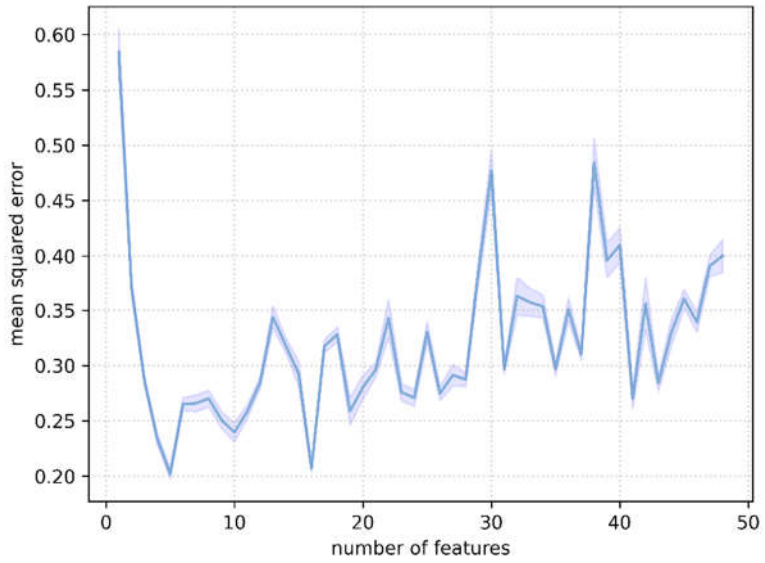
isabella
CLUSTER

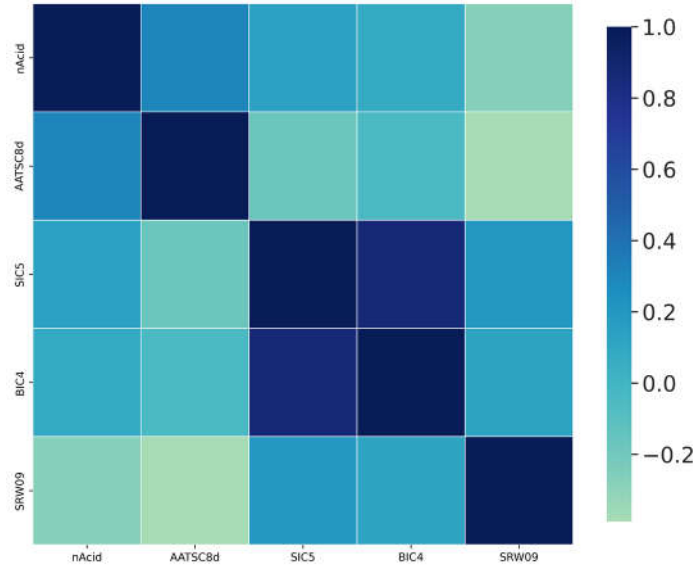
cnrm
Center for Advanced Computing and Modelling

DNN

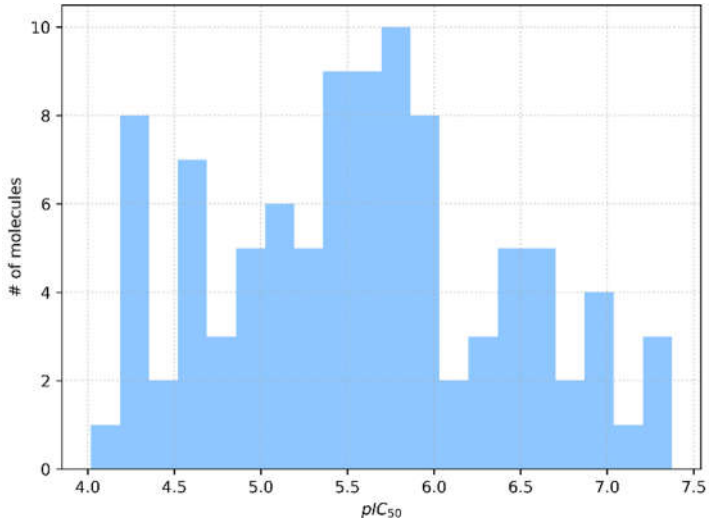

# of molecules
pIC_{50}

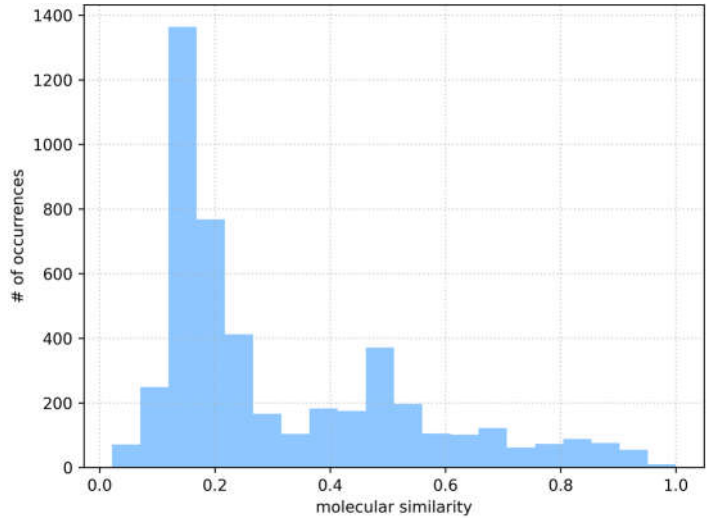

# of occurrences
molecular similarity

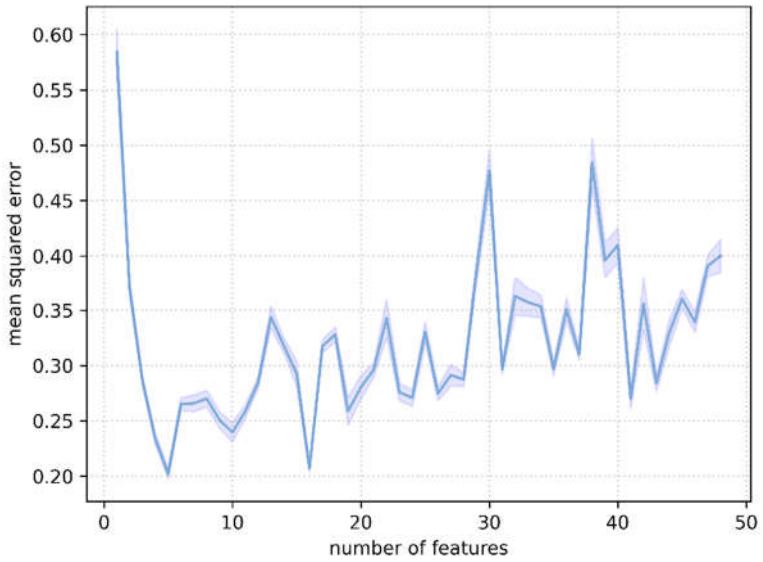

mean squared error
number of features


nAcid
AATSC8d
SIC5
BIC4
SRW09

DNN



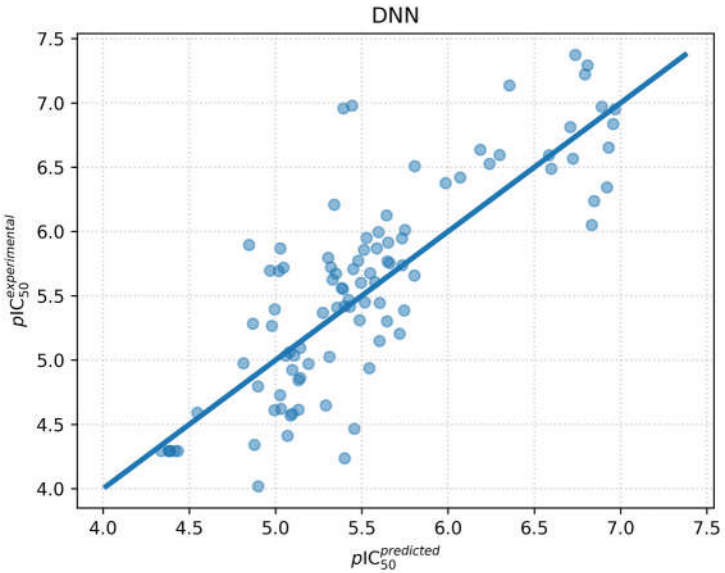


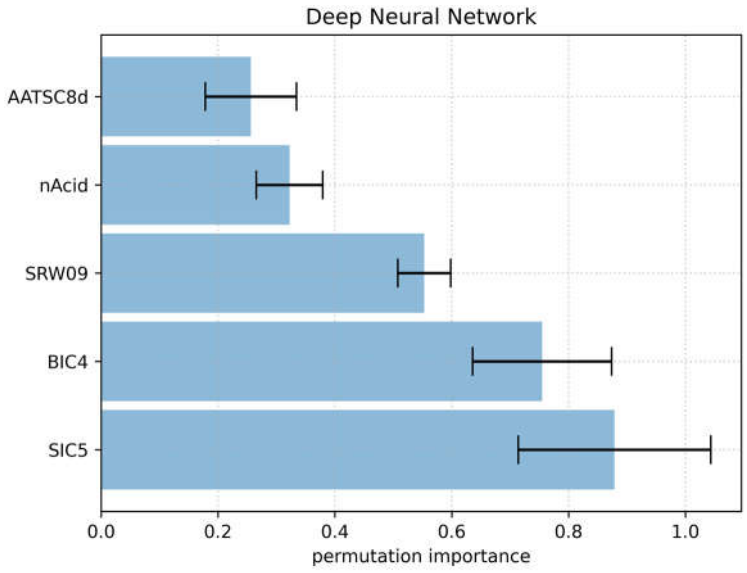


dropout_rate: [**0.0**, 0.1, 0.2, 0.3]
n_layers: [8, 16, 32, **64**, 128]
n_neurons: [32, 64, 128, **256**]
L2 weight_decay: [0.0, 0.1, 0.001, **1e-05**]
Early stopping
Trainable params: 4080897

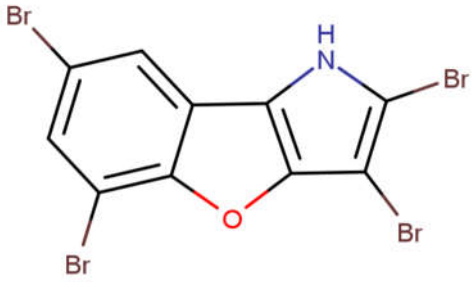
WNV NS2B/NS3

model	N	train			test		
		MAE	MSE	RMSE	MAE	MSE	RMSE
DNN	5	0.250	0.116	0.341	0.253	0.112	0.335
RF	5	0.159	0.049	0.221	0.413	0.285	0.534
SVM	12	0.188	0.082	0.286	0.404	0.300	0.548
XGB	8	0.134	0.032	0.184	0.352	0.232	0.482

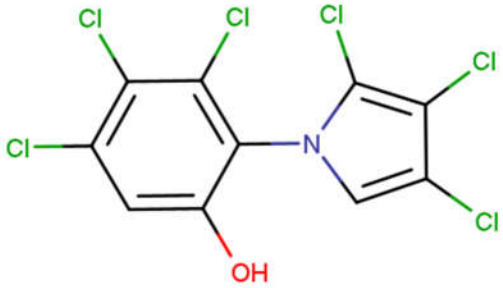




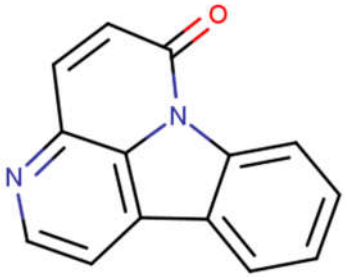
WNV NS2B/NS3



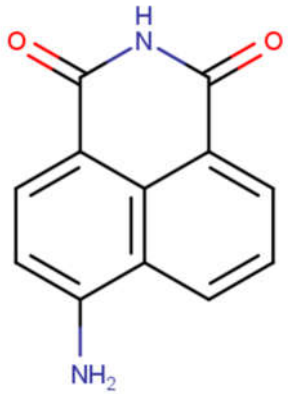
pIC_{50} = 12.5



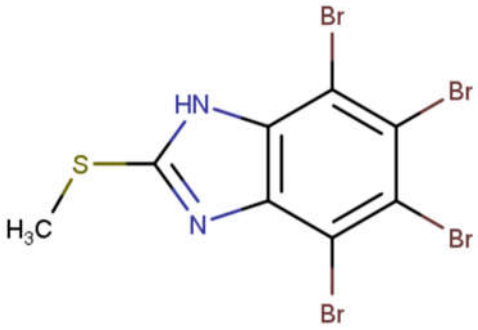
pIC_{50} = 11.3



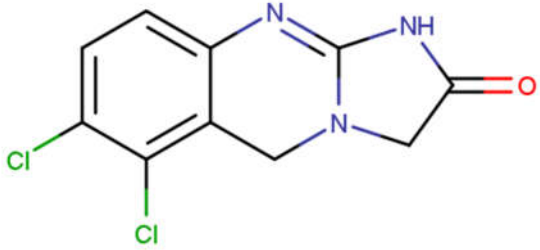
pIC_{50} = 10.7



pIC_{50} = 10.0

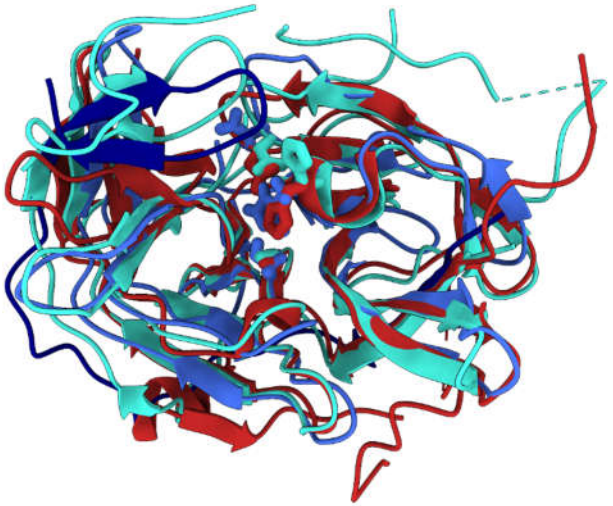


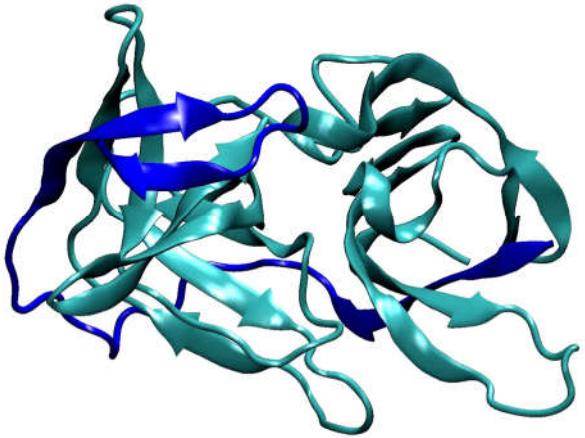
pIC_{50} = 10.0

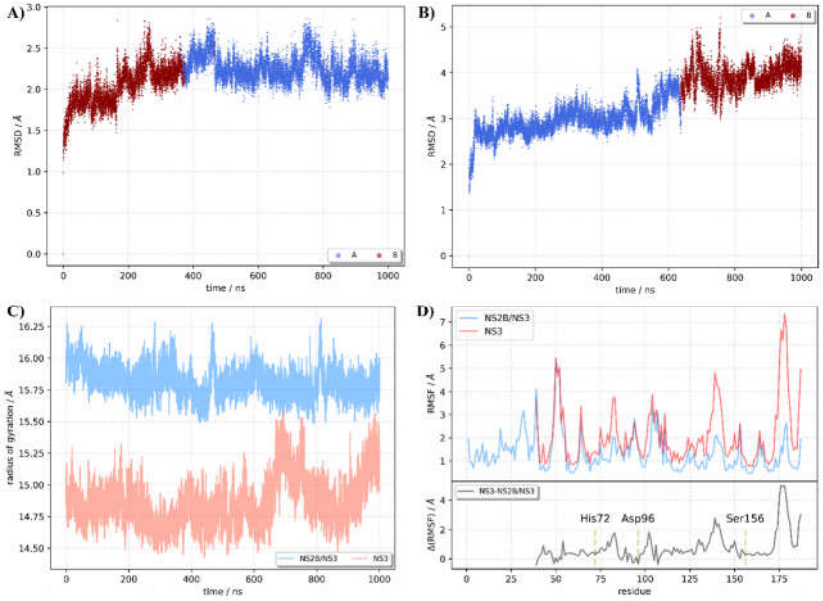


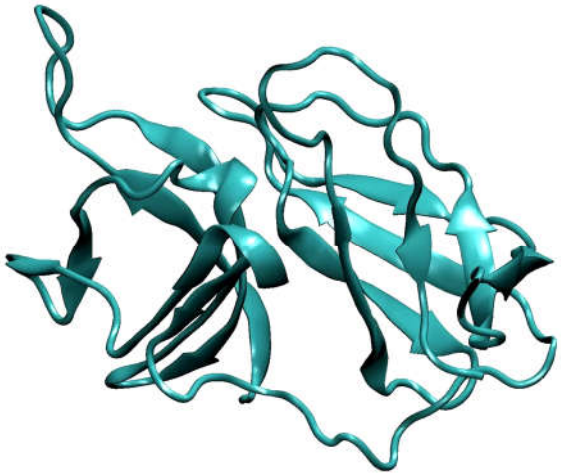
pIC_{50} = 9.9

KFDV NS2B/NS3

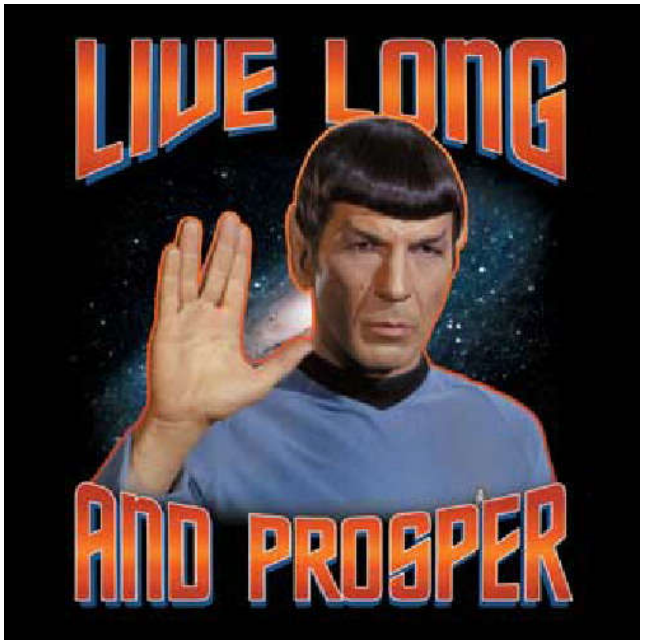








[6 s
si
[7 e
K,

Thanks

- Shivananda Kandagalla
- Hrvoje Rimac
- Vladimir Potemkin
- Vladislav Naumovich
- Prateek Pathak
- Maria Grishina
- Gordan Janeš
- Ivan Mitrović