

XXVIII Symposium on Bioinformatics and computer-aided drug discovery

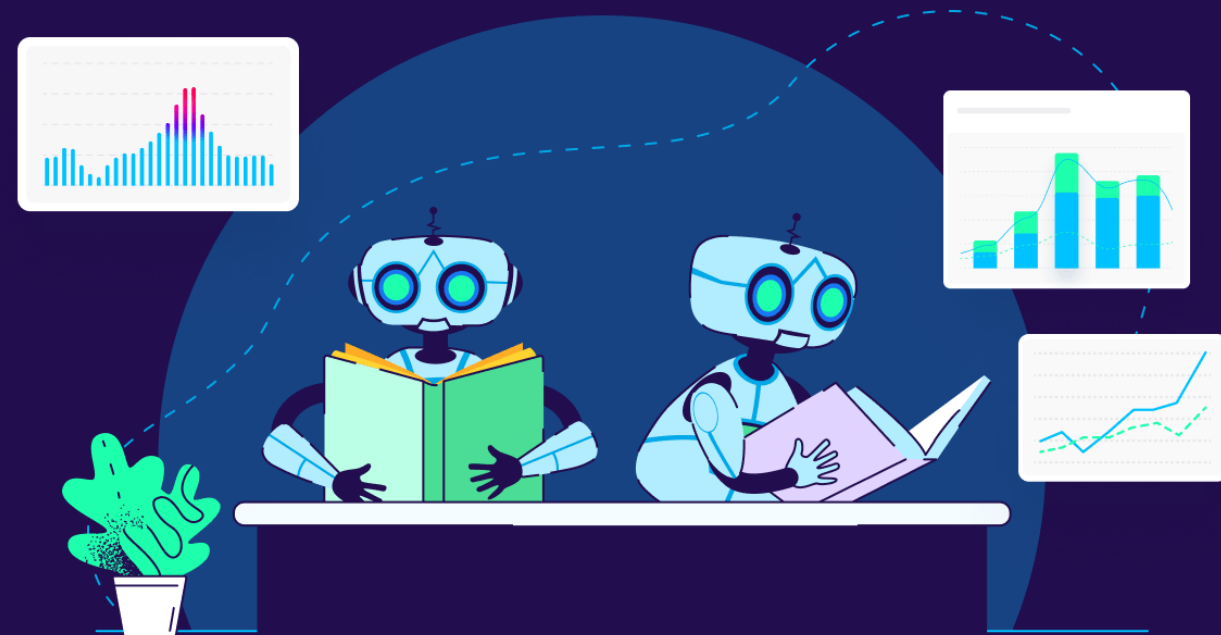
# EXPLORING THE SCORING FUNCTION SPACE FOR STRUCTURE-BASED DRUG DESIGN



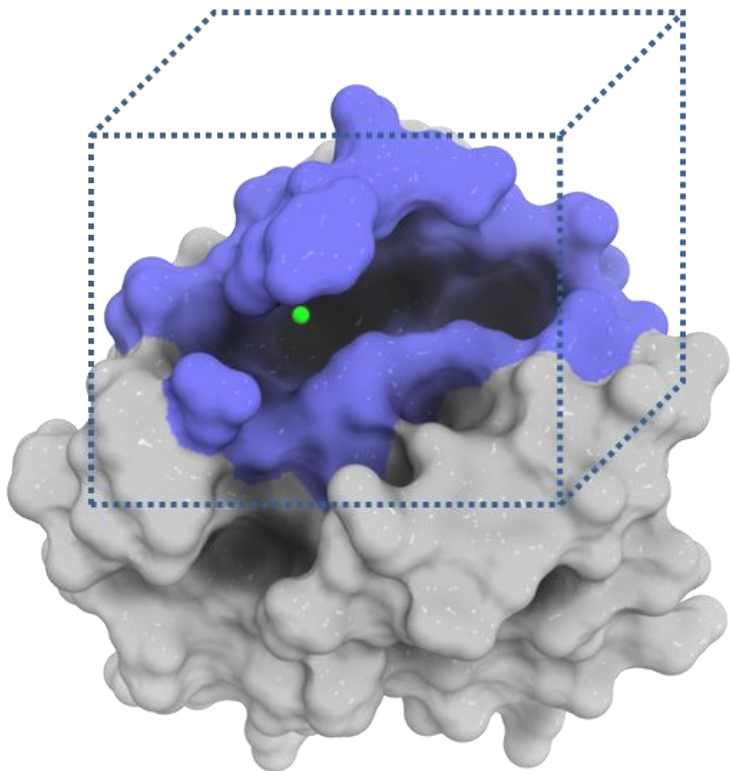
Gabriela Bitencourt Ferreira



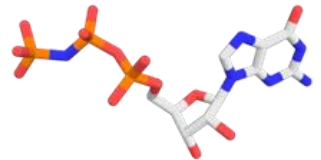
Pontifical Catholic University of Rio Grande do Sul, Brazil



# MOLECULAR DOCKING

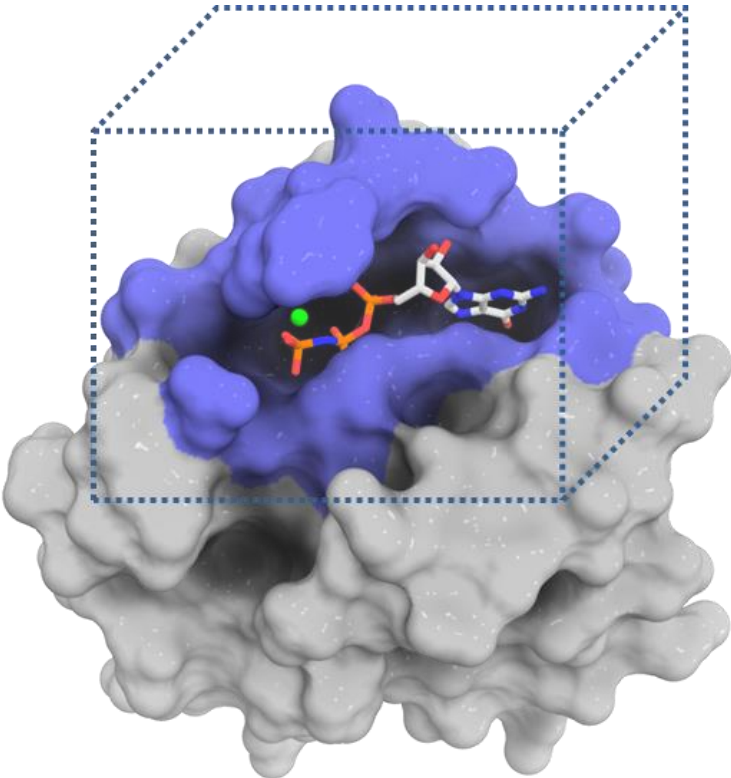


+

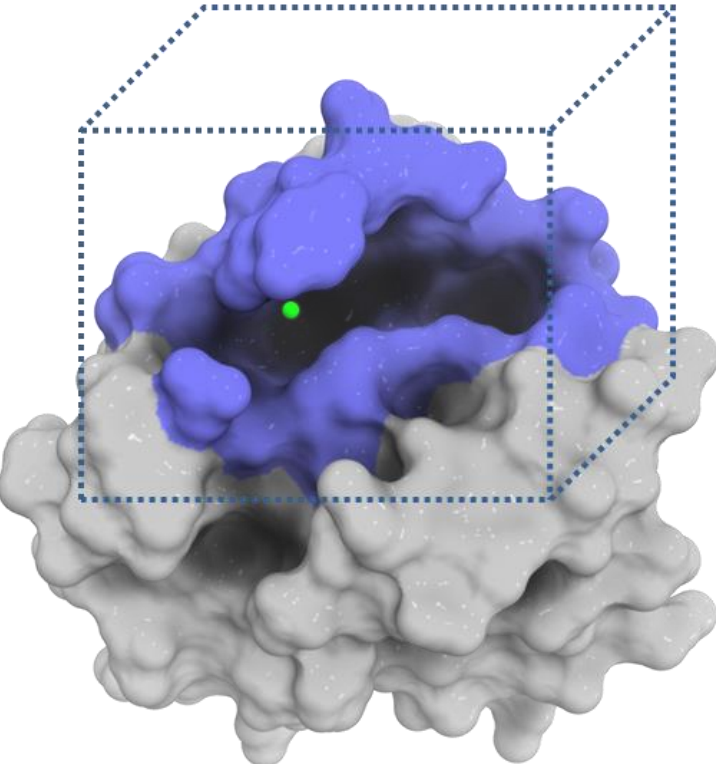


=

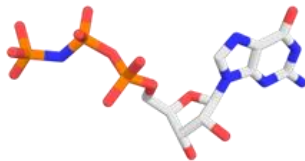
Scoring Function



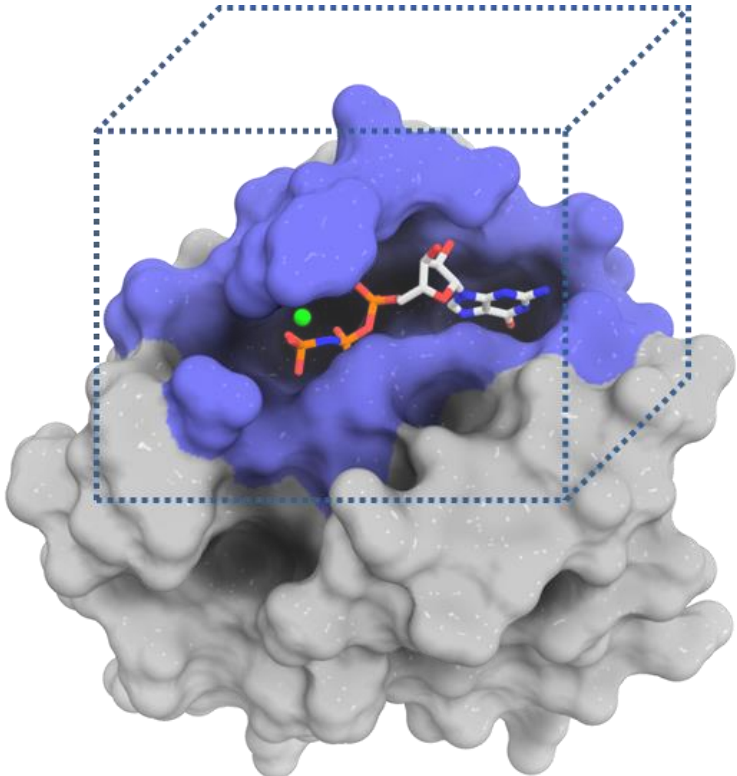
# MOLECULAR DOCKING



+



=



Classical Scoring Function

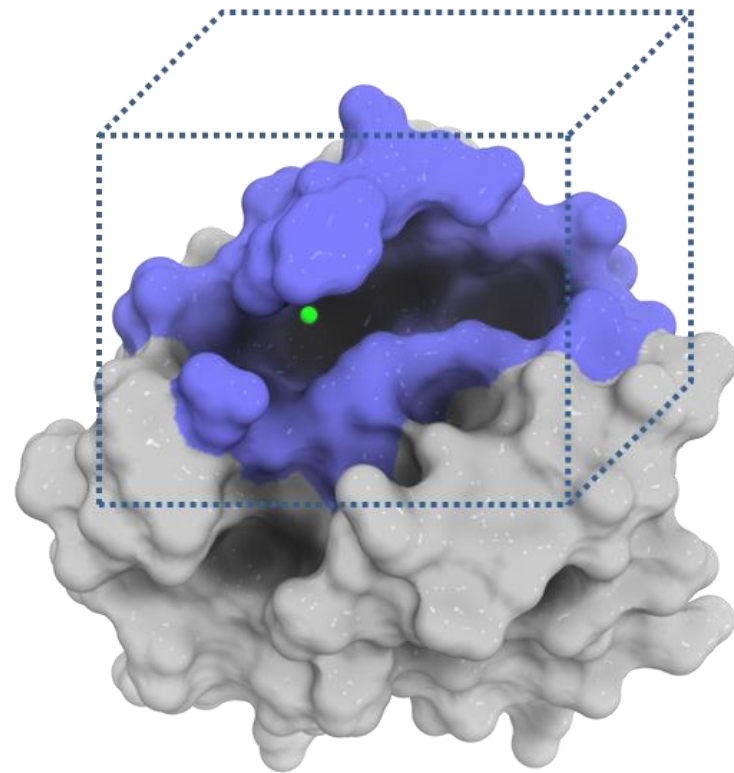
**Table 1.** Protein and Ligand Data Set Details

protein	target type	no. of ligands	no. of ligand classes	no. of cocrystals	max affinity (nM)	min affinity (nM)
Chk1	kinase	193	2	15	7	>10000
factorxa	serine protease	218	4	10	<1	5000
gyrase B	isomerase	138	3	7	4	>10000
HCV polymerase	polymerase	205	2	13	5.6	>10000
Met tRNA synthetase	synthetase	144	2	31	1	>10000
<i>E. coli</i> PDF	metalloprotease	199	3	2	1	>10000
<i>Strep</i> PDF	metalloprotease	186	3	4	<2	>10000
PPAR $\delta$	nuclear hormone receptor	206	5	54	0.3	>10000

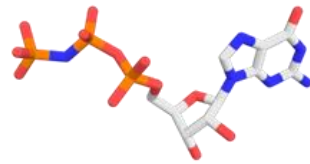
**Table 7.** Best Correlation Coefficient  $r$  between the  $-\log$  Affinity (pAffinity) and Docking Score for All Programs across All Targets

program	Chk1	FXa	gyrase B	HCVP	MRS	<i>E. coli</i> PDF	<i>Strep</i> PDF	PPAR $\delta$
Dock4	-0.33	-0.31	-0.39	0.00	-0.13	-0.38	-0.34	0.07
DockIt	-0.49	-0.19	-0.37	0.04	-0.28	-0.13	-0.30	-0.34
FlexX	-0.57	-0.31	-0.39	-0.12	-0.01	-0.42	-0.25	-0.36
Flo+	-0.44	-0.38	-0.36	-0.09	0.05	-0.27	-0.39	-0.42
Fred	-0.14	0.01	-0.13	-0.07	0.13	0.07	-0.24	0.06
Glide	-0.47	-0.08	-0.21	-0.04	0.08	-0.13	-0.12	-0.35
Gold	-0.42	-0.05	-0.14	-0.09	0.04	-0.12	-0.11	-0.43
LigandFit	-0.45	-0.13	-0.39	-0.06	-0.15	-0.21	-0.49	-0.10
MOEDock	-0.29	0.00	0.07	-0.01	-0.13	0.08	0.20	0.17
MVP	-0.26	0.10	-0.33	-0.01	-0.18	-0.17	-0.16	-0.18

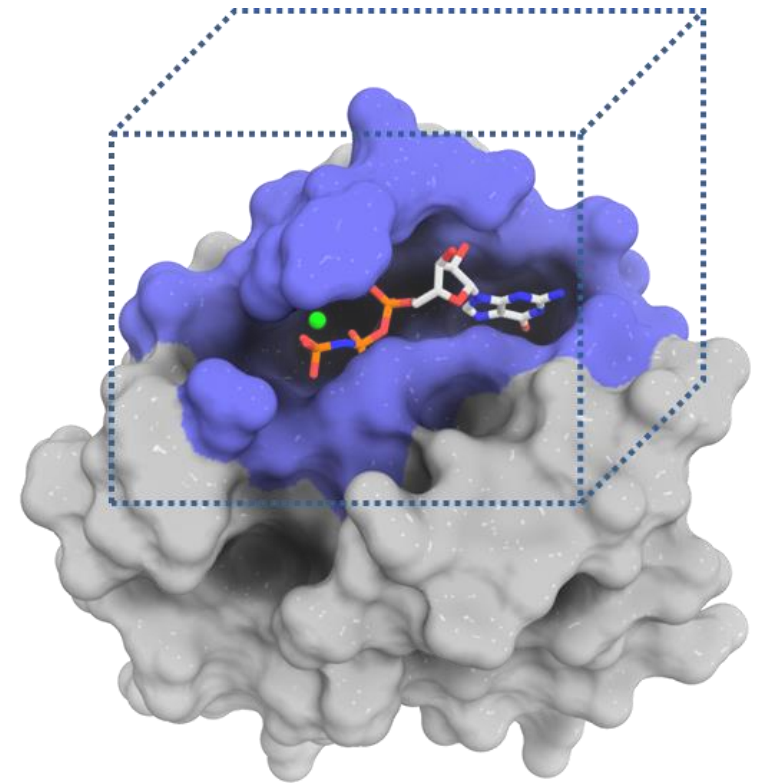
# MOLECULAR DOCKING



+



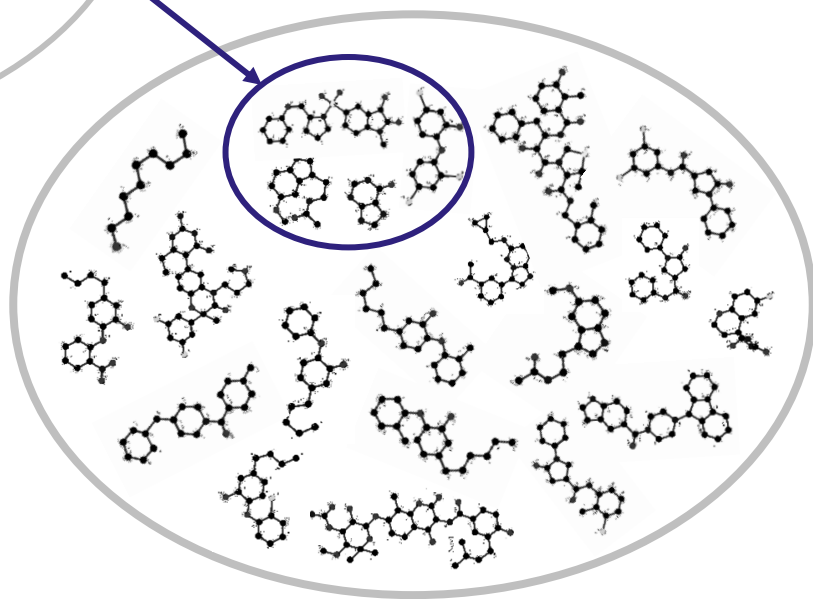
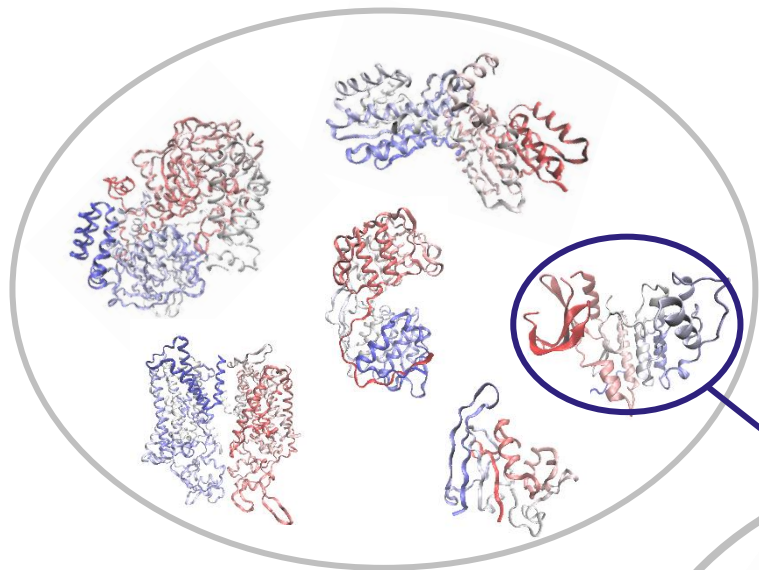
=



Targeted Scoring Function

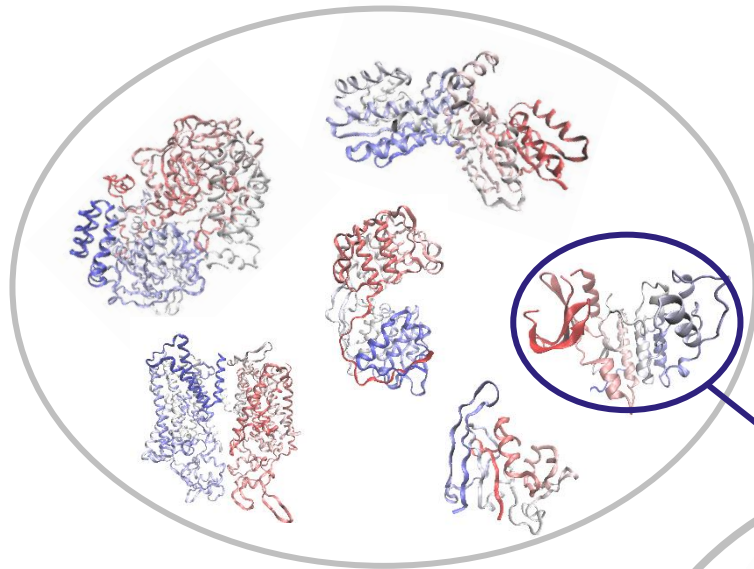
We suggest the use of targeted scoring functions, specific to the protein we are studying

Protein Space



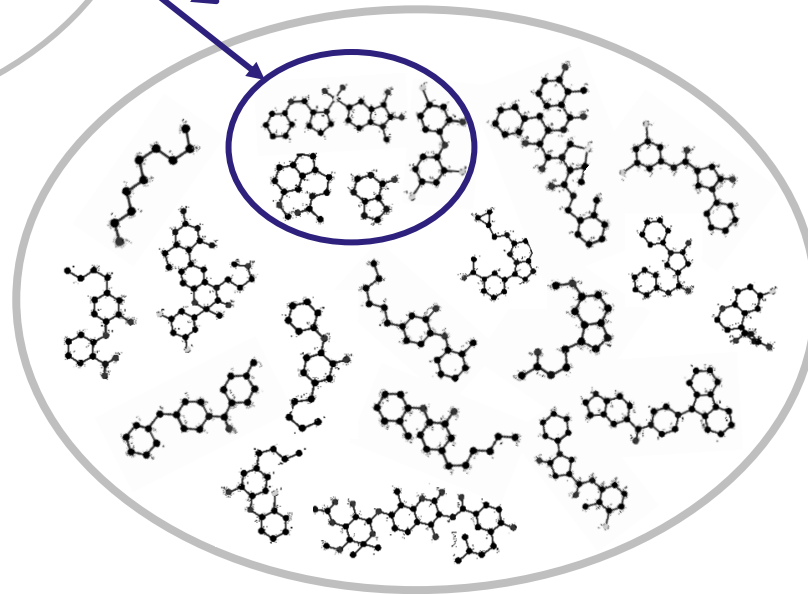
Chemical Space

## Protein Space

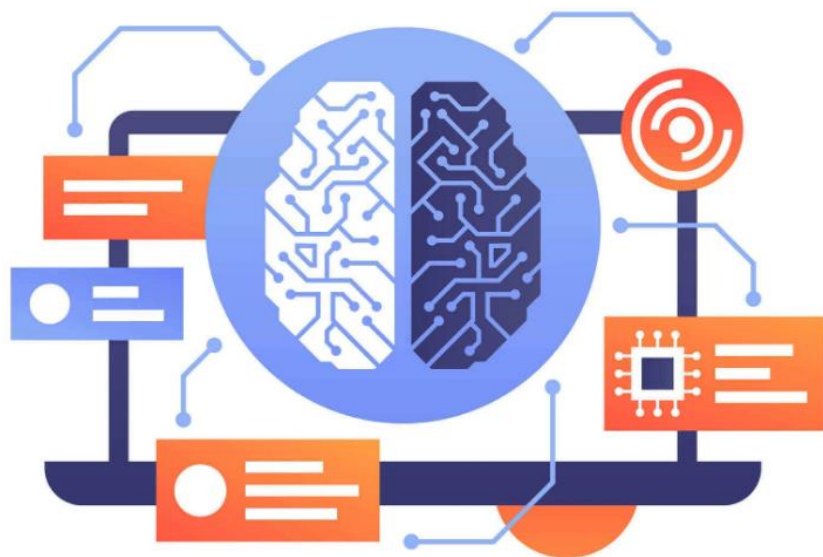


## Scoring Function Space

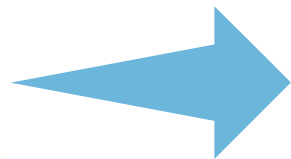
$$\log(IC_{50}) = \sum_{i=0}^N \omega_i x_i + \sum_{j=0}^N \alpha_j x_j^i$$
$$\Delta G = \sum_{i=0}^N \omega_i x_i \quad f = \sum_{i=1}^N \alpha_i x_i - x_j^{-3} + \sum_{j=1}^M x$$
$$\Delta S = \alpha_j - x_i \sum_{i=1}^N x_i y_j \quad f = \alpha_j \beta_i + x$$
$$\log(K_I) = \sum_{i=0}^N \omega_i x_i + \sum_{j=1}^M \sum_{i=1}^N \lambda$$



## Chemical Space



Machine Learning methods



## Scoring Function Space

$$\log(IC_{50}) = \sum_{i=0}^N \omega_i x_i + \sum_{j=0}^N \alpha_j x_j^i$$

$$A = \pi r^2$$

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left( a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right) \frac{[E][I]}{[EI]}$$

$$\log(EC_{50}) = \sqrt{x_i} + \sum_j^i x$$

$$\Delta G = \sum_{i=0}^N \omega_i x_i$$

$$a^2 + b^2 = c^2$$

$$\Delta S = \alpha_j - x_i \sum_{i=1}^N x_i y_j$$

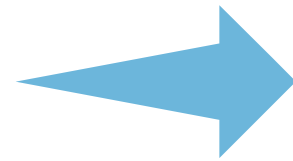
$$f = \alpha_j \beta_i + x$$

$$\log(K_I) = \sum_{i=0}^N \omega_i x_i + \sum_{j=1}^M \sum_{i=1}^N \lambda$$





Sixty-four different regression methods



## Scoring Function Space

$$\log(IC_{50}) = \sum_{i=0}^N \omega_i x_i + \sum_{j=0}^N \alpha_j x_j^i$$

$$A = \pi r^2$$

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left( a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right) \frac{[E][I]}{[EI]}$$

$$\log(EC_{50}) = \sqrt{x_i} + \sum_j^i x$$

$$\Delta G = \sum_{i=0}^N \omega_i x_i$$

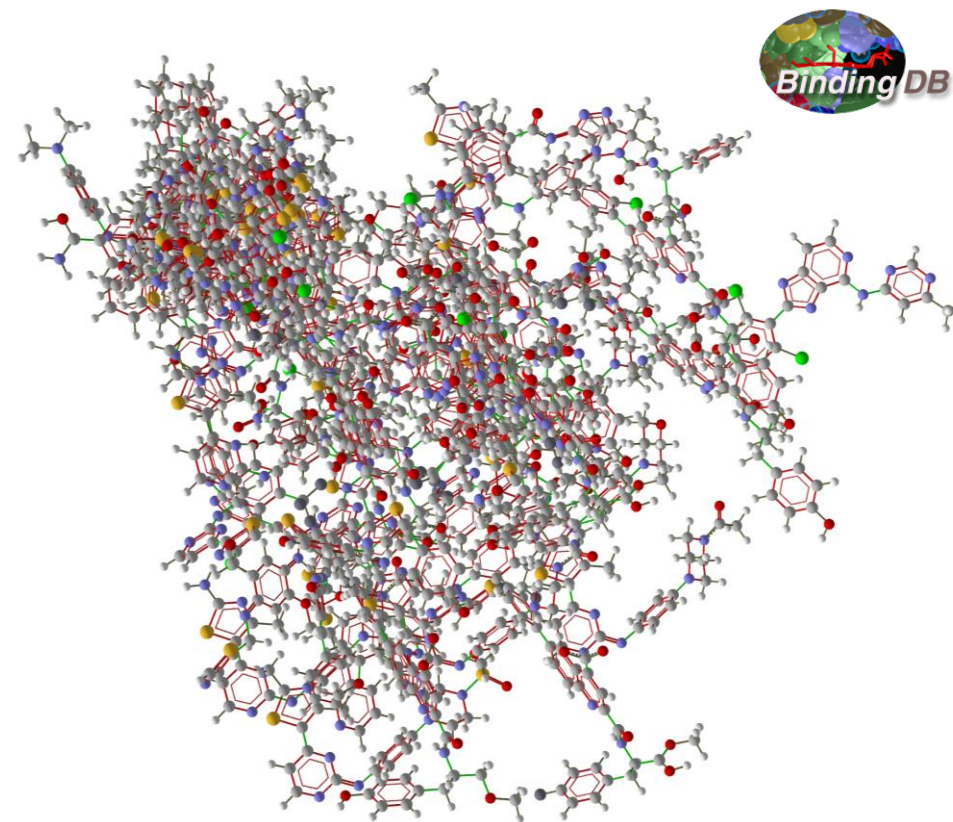
$$a^2 + b^2 = c^2$$

$$\Delta S = \alpha_j - x_i \sum_{i=1}^N x_i y_j$$

$$f = \alpha_j \beta_i + x$$

$$\log(K_I) = \sum_{i=0}^N \omega_i x_i + \sum_{j=1}^M \sum_{i=1}^N \lambda$$

# DATA



77 ligands



Sixty-four different regression methods

SAnDReS 2.0

Setup Edit Dataset Docking Hub Scoring Functions Virtual Screening Machine Learning Box Statistical Analysis Help

S A n D R e S

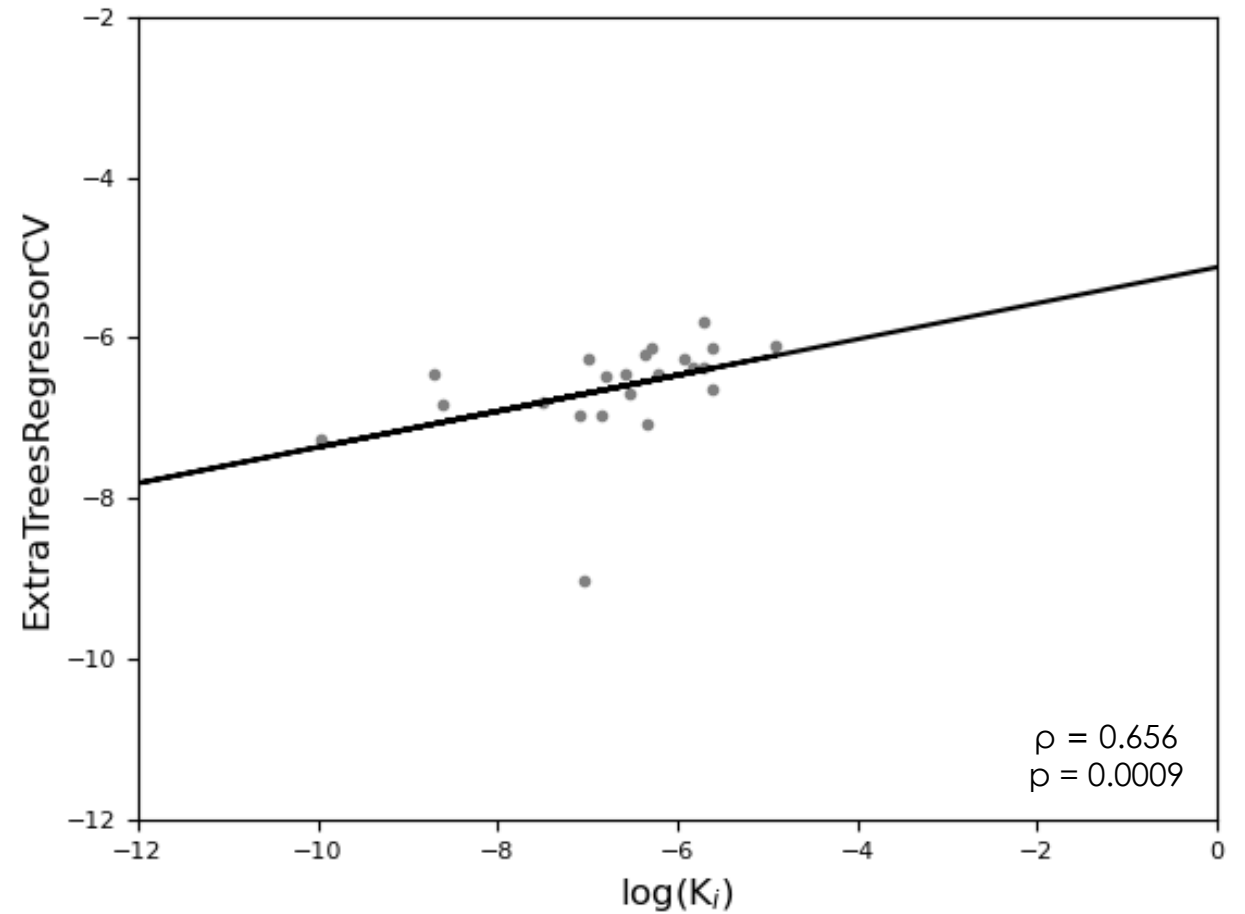
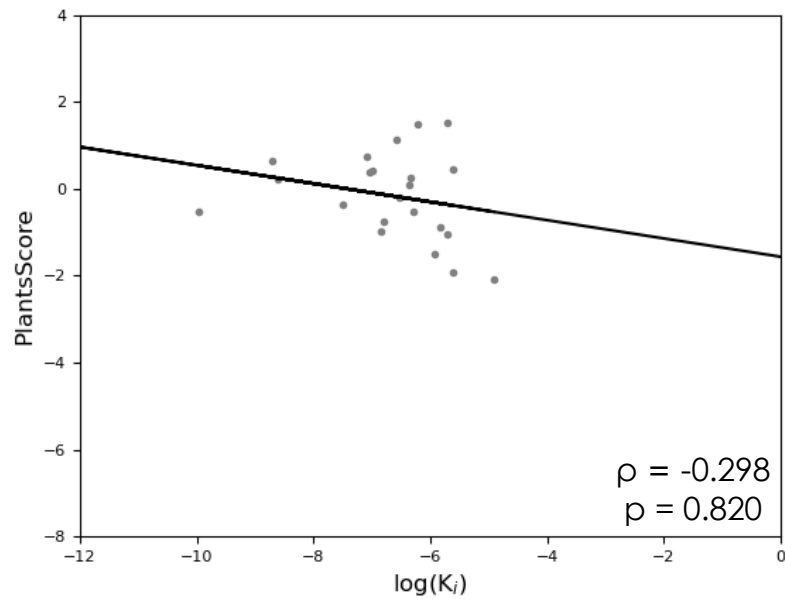
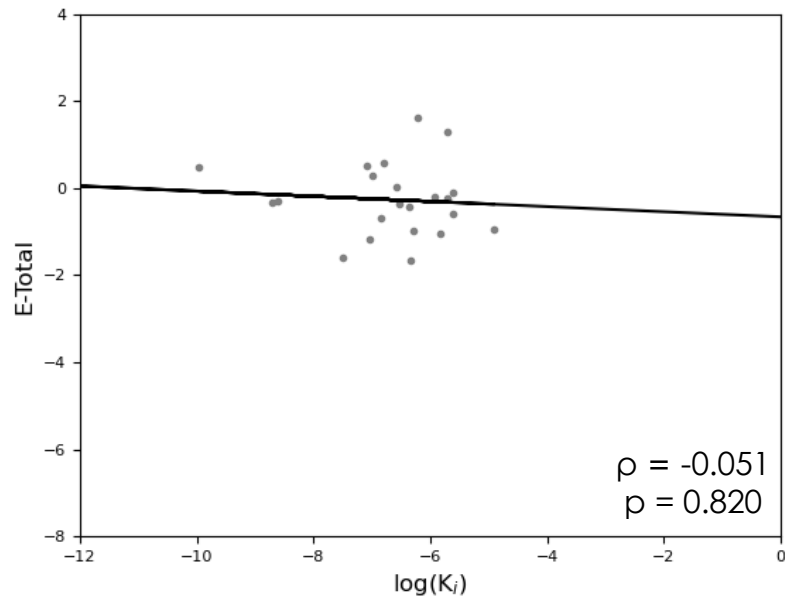
SAnDReS opened misc/data/ml\_par.csv file Site RCSB

Fast Editor

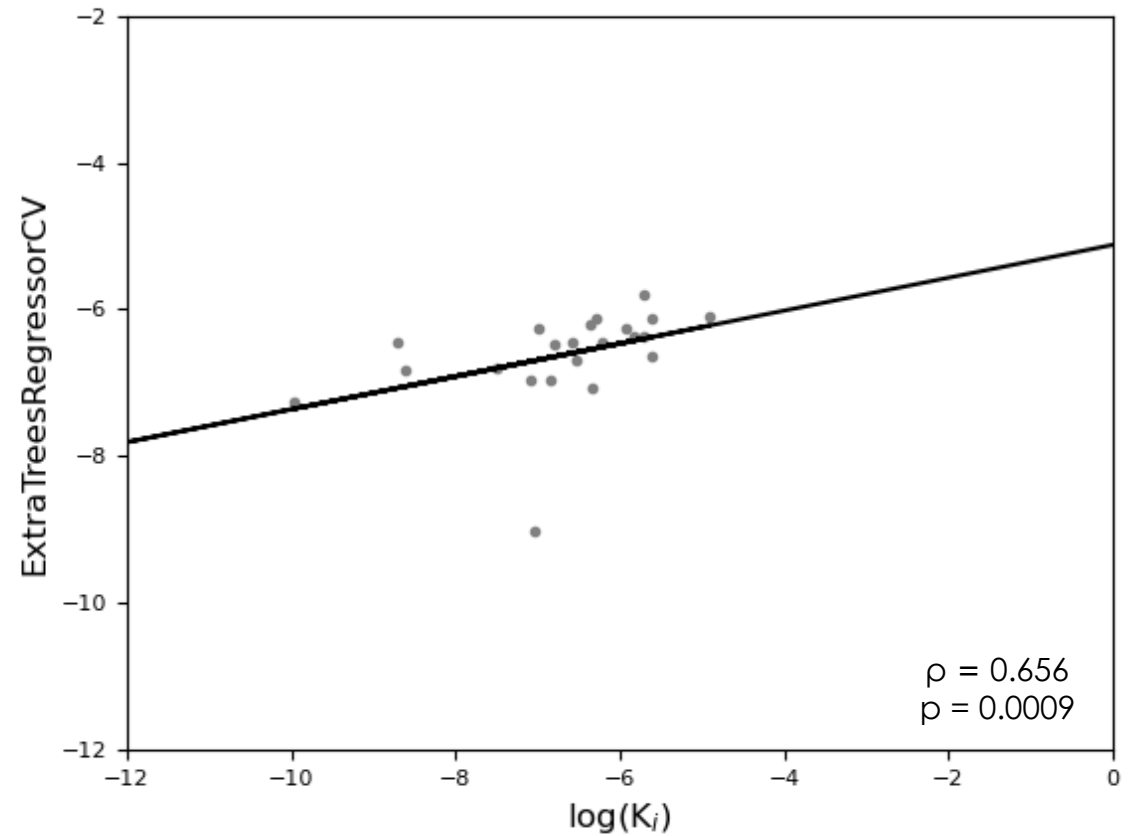
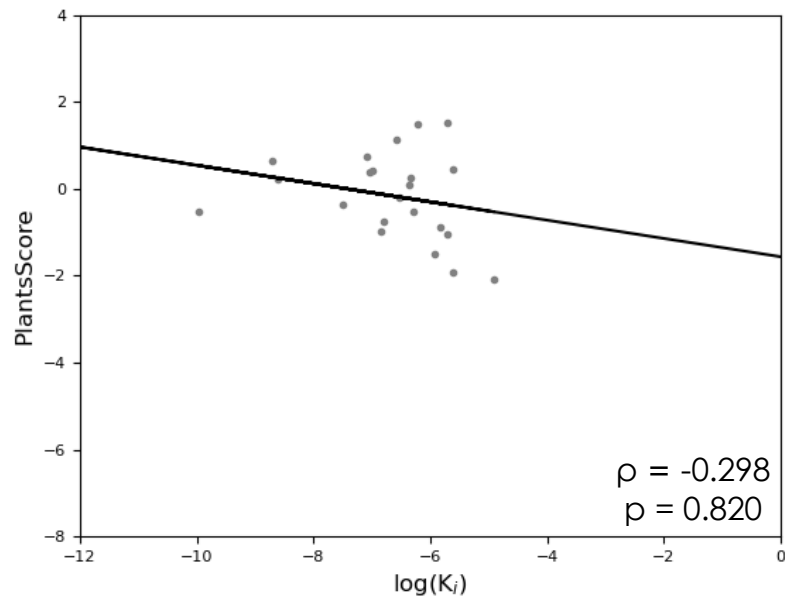
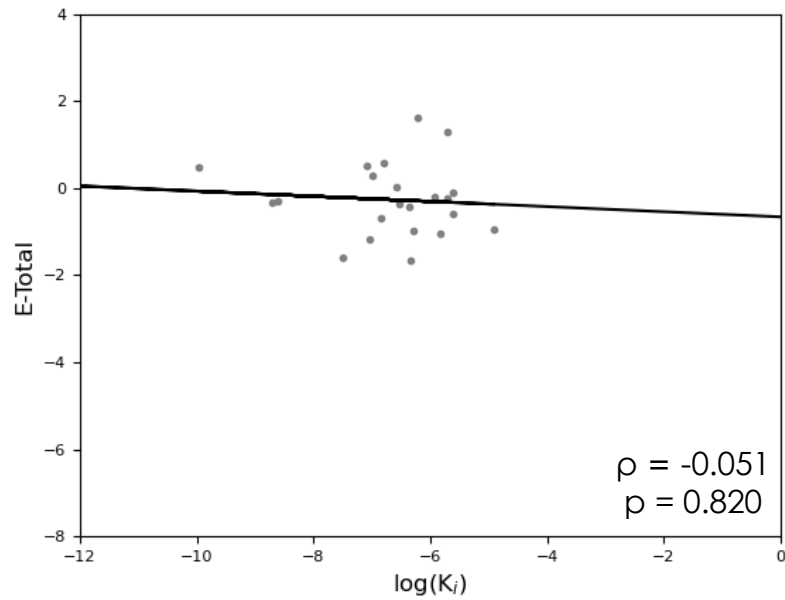
File: misc/data/ml\_par.csv Read Clear Save Close

```
on methods)
#
# Set up input parameters
n_features, 6
features_in, Csp3, ElectroLong, LE1, E-Intra (steric), E-Intra (tors-ligandatoms), E-Intra (vdw)
test_size_in, 0.3
seed_in, 271828
"
```

# MODEL TEST SET



# MODEL TEST SET



We concluded that the use of targeted scoring functions can be a new approach to predict the binding affinity

# Thank you!



[g.bitencourt@edu.pucrs.br](mailto:g.bitencourt@edu.pucrs.br)  
[gabriela@azevedolab.net](mailto:gabriela@azevedolab.net)



<https://azevedolab.net/>

