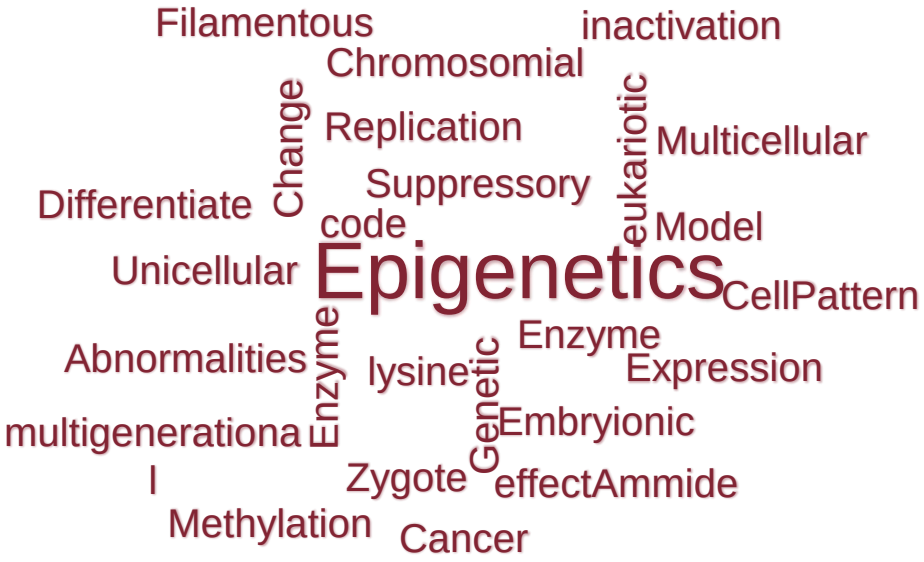
Rational Design of Potential Novel Pharmaceutical Compounds Through Computational Approaches and Machine Learning Techniques.

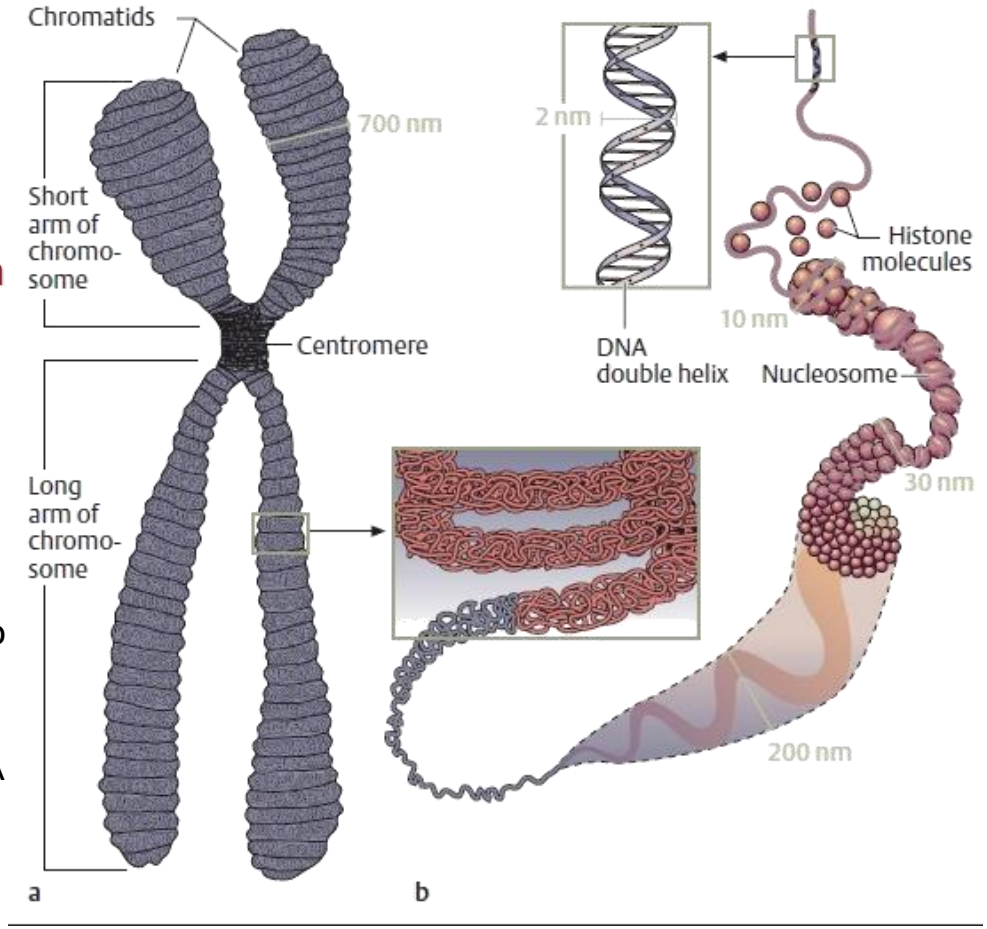
SAPIENZA
UNIVERSITÀ DI ROMA

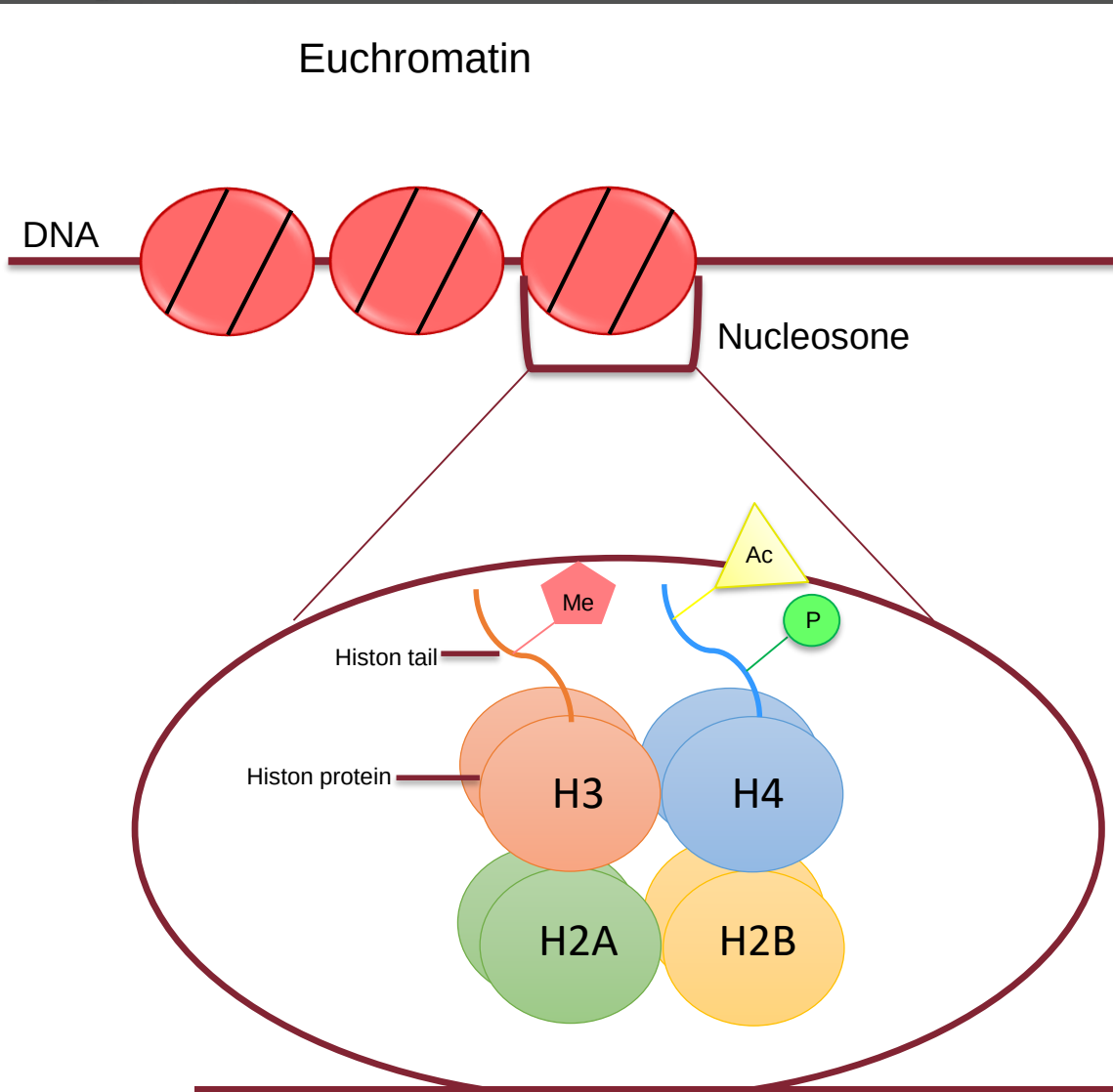
PhD School of Pharmaceutical Sciences XXX Cycle
Department of Chemistry and Drug Technologies
Faculty of Pharmacy and Medicine
"Sapienza" University of Rome

Manuela Sabatino



The word “epigenetic” literally means “in addition to changes in genetic sequence.”
The term has evolved to include any process that alters gene transcription without changing the DNA sequence, and leads to modifications that can be transmitted to daughter cells





In cells the Chromatin switches between two states:
Euchromatin → relaxed state, the gene trascription is activated;
Heterochromatine → Condensed state, the gene trascription is largely repressed;
This is possibile through a marker, for example a istonic modication, that are recognized by other proteic complex.

Epigenetics

Acetyltransferases
- HAT1
- CBP/P300
- ScSAS

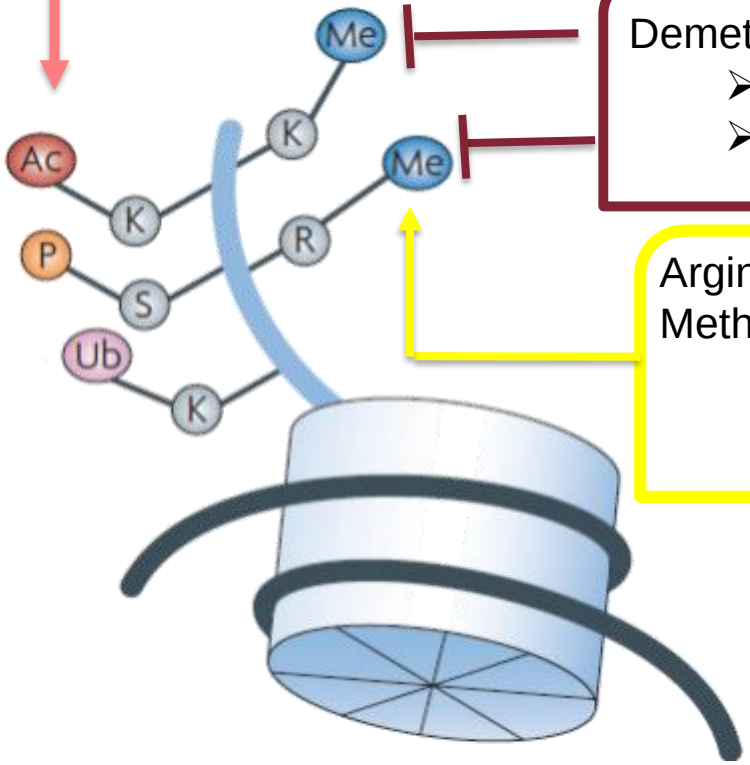
Deacetylases
- SirT2

Kinases
- MSK1
- MSK2
- CKII

Lysine-Methyltransferases
- SUV39
- MLL(1-5)
- EZH
- DOT1L

Demethylases
- LSD1
- JHDM

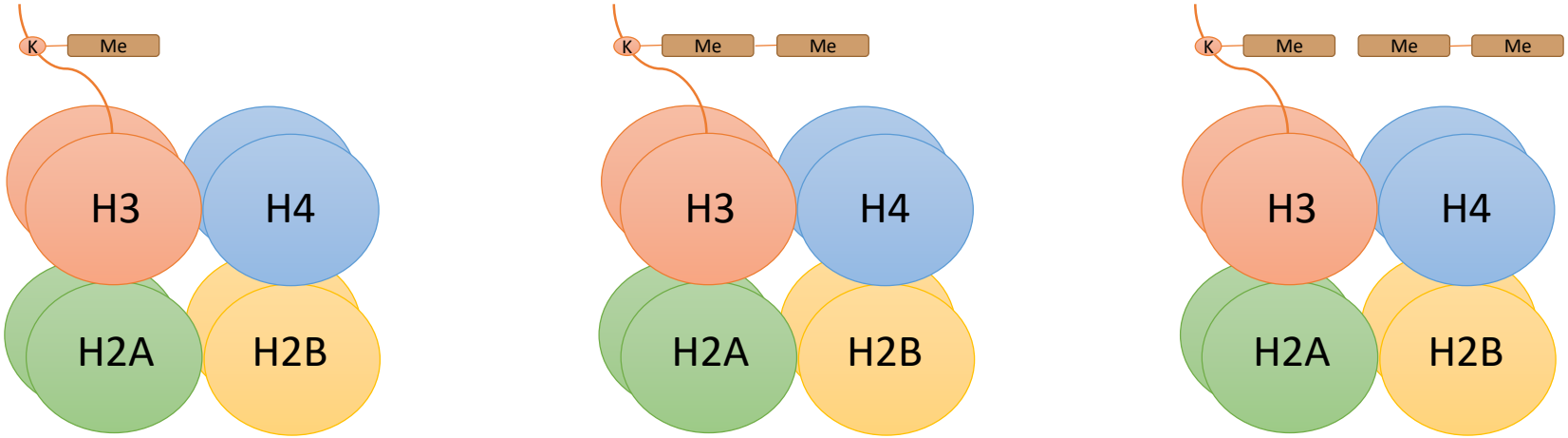
Arginine-Methyltrasferases
- CARM1
- PRMT4
- PRMT5



Dot1l catalitic mechanism

by www.RCMD.it

Disruptor Of Telomeric Silencing 1-Like Methyltransferase H3K79 selective

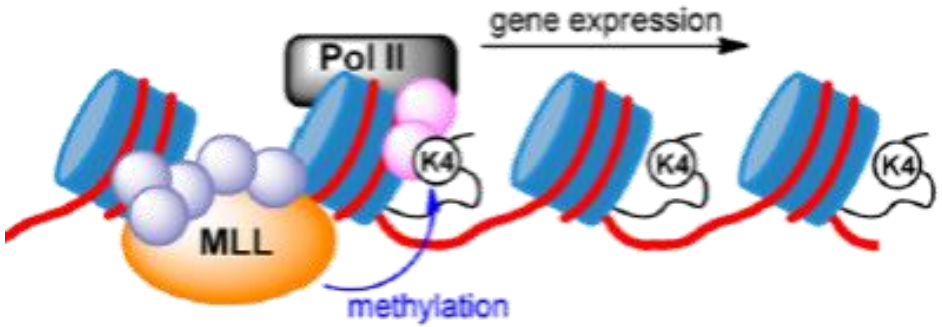


TRANSCRIPTIONAL ACTIVATIONS

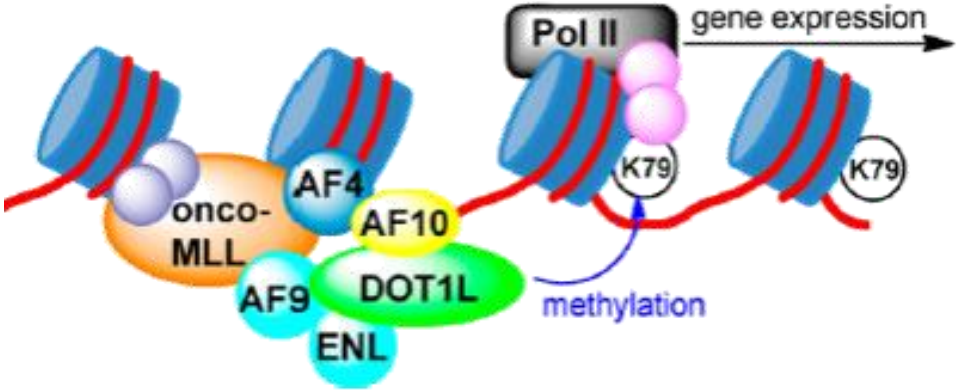
Dot1l and MLL

MLL is a methyltrasferase and modifies the lysine 4 of the histone 3

Wild-type



Myeloid Lynphoid Leukemia



Objective Of The Study

by www.RCMD.it

DOT1L Inhibitors Study

↓

Quantitative Predictive Models

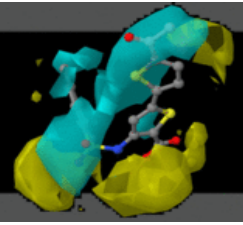
↓

Design of New Compounds

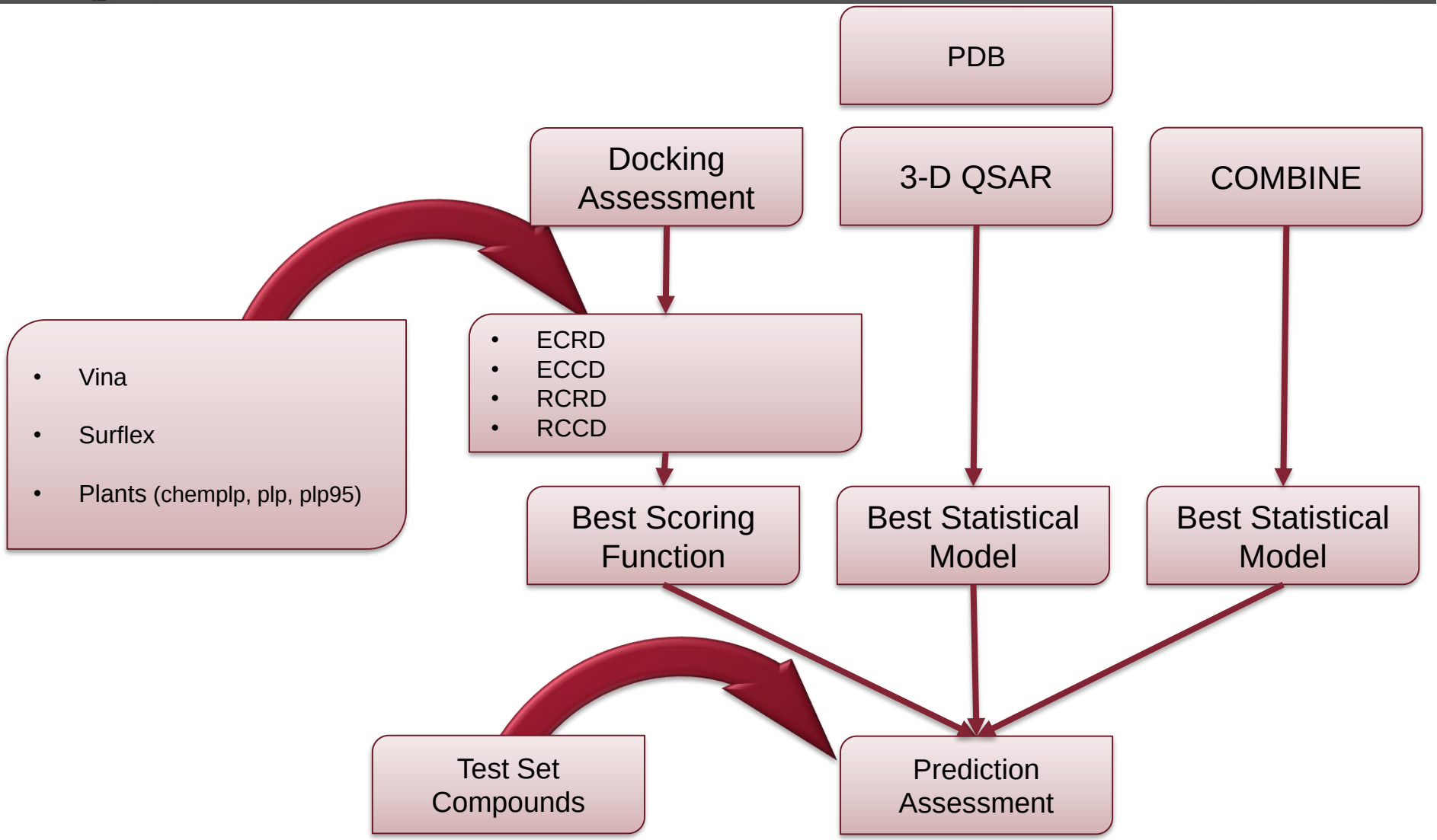
↓

Prioritize Synthesis

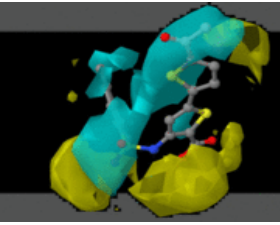
Workflow

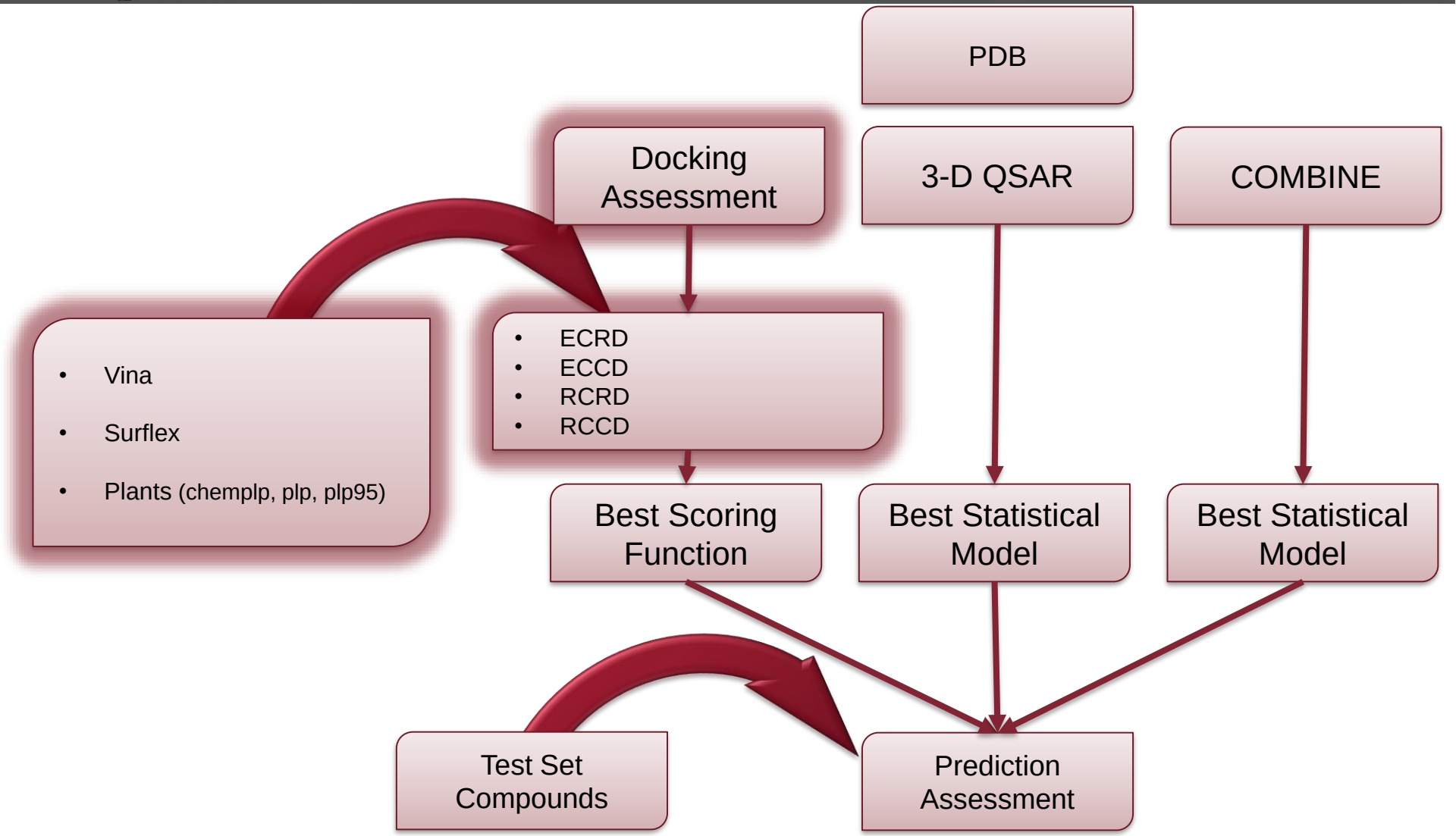


by www.RCMD.it



Workflow

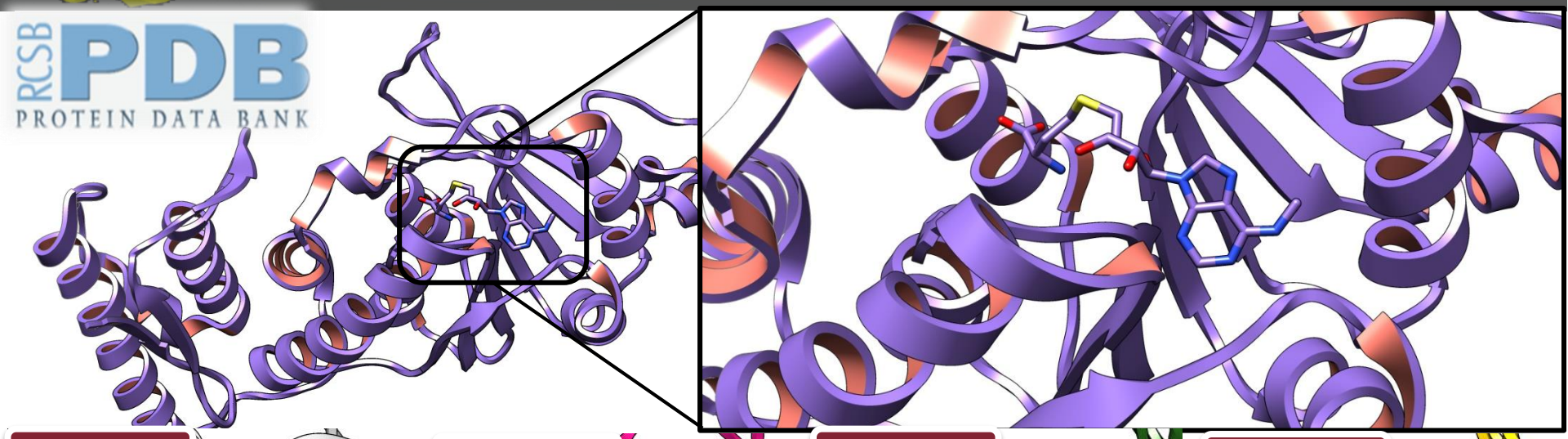



PDB
Docking Assessment
3-D QSAR
COMBINE
Vina
Surflex
Plants (chemplp, plp, plp95)
ECRD
ECCD
RCRD
RCCD
Best Scoring Function
Best Statistical Model
Best Statistical Model
Test Set Compounds
Prediction Assessment

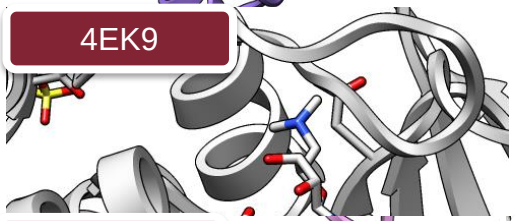
Docking Assessment

by www.RCMD.it

RCSB PDB PROTEIN DATA BANK

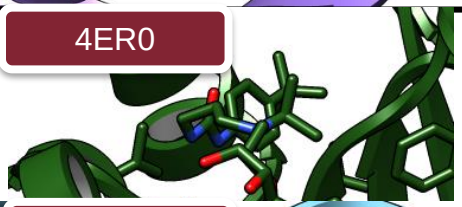


4EK9

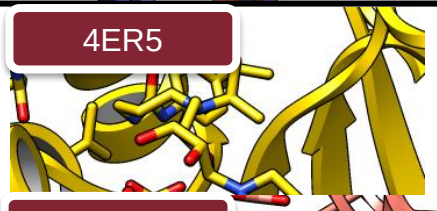


4EKI

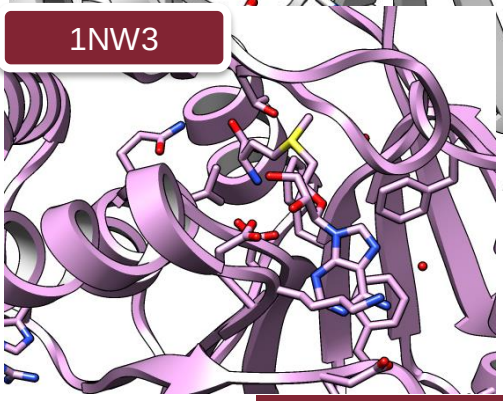
4ER0



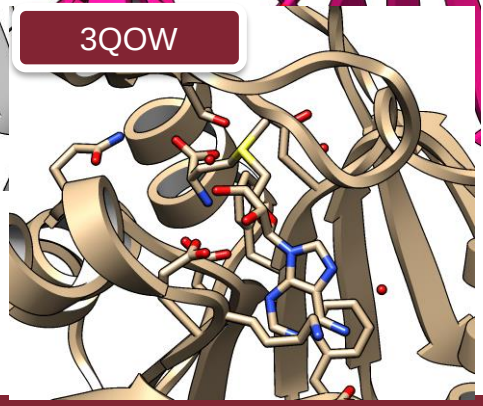
4ER5



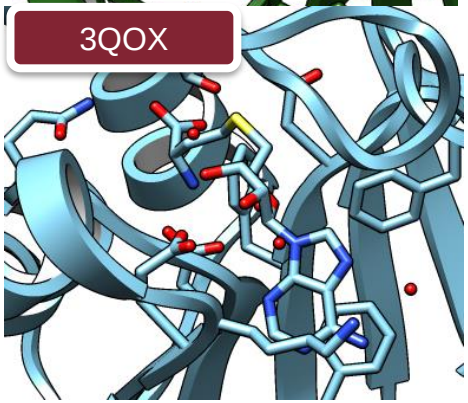
1NW3



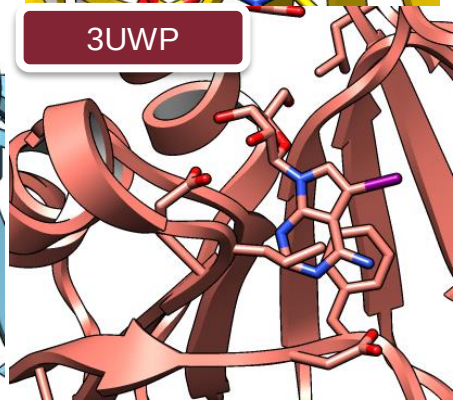
3QOW



3QOX



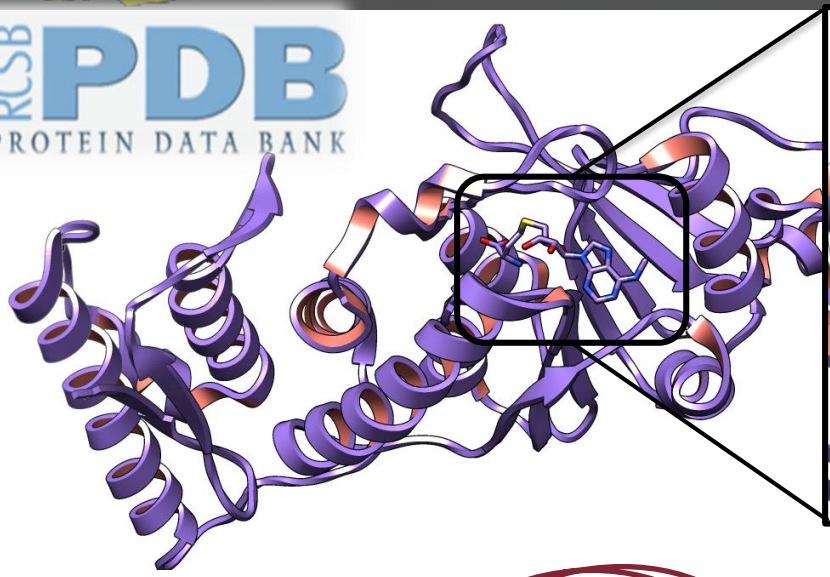
3UWP

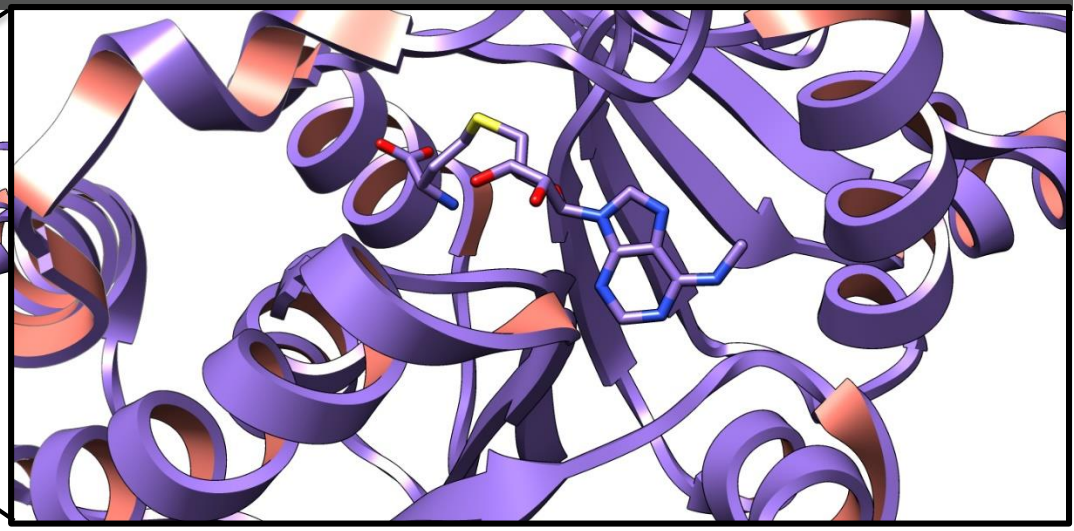


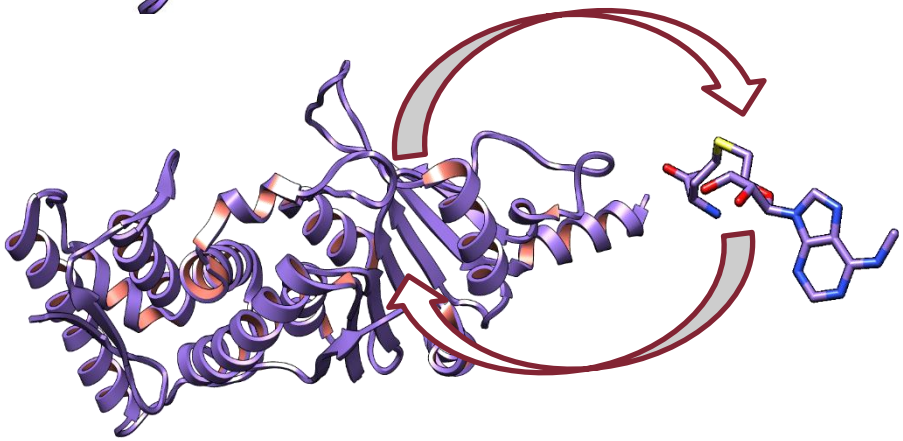
Docking Assessment

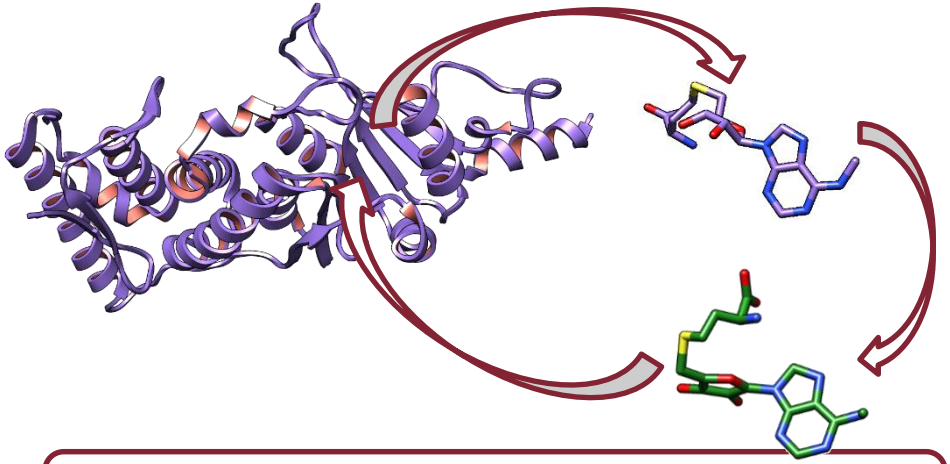
by www.RCMD.it

RCSB PDB PROTEIN DATA BANK







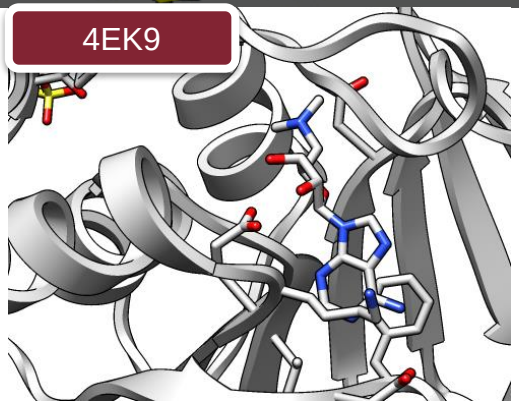


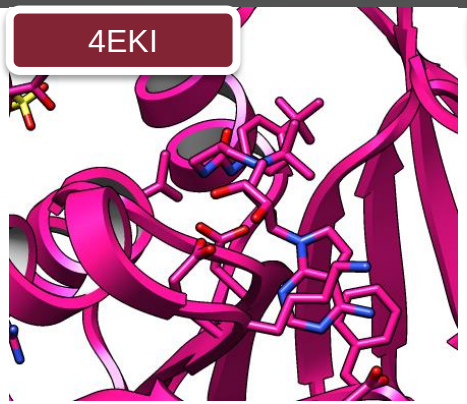
Experimental **C**onformation **R**e-**D**ocking

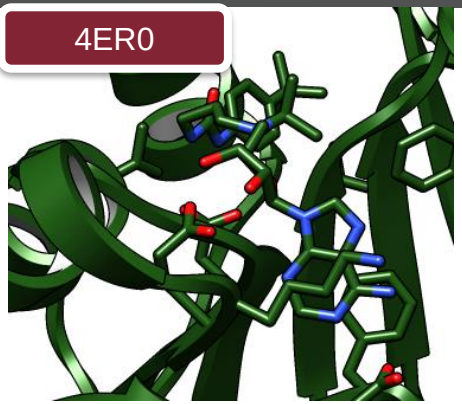
Random **C**onformation **R**e-**D**ocking

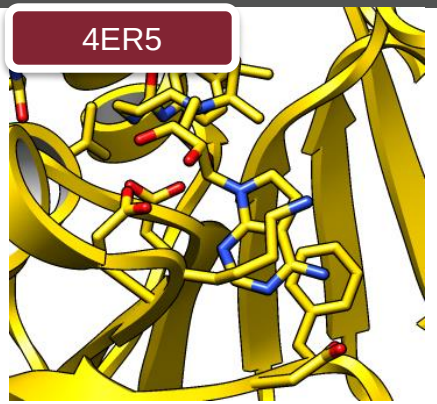
Docking Assessment

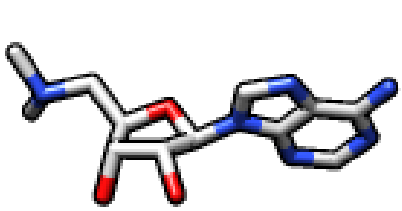
by www.RCMD.it

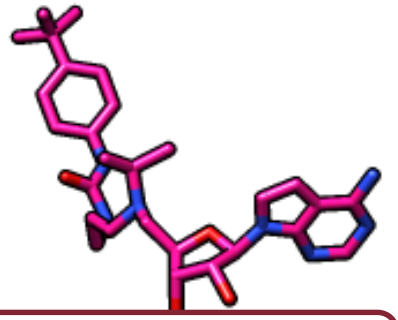




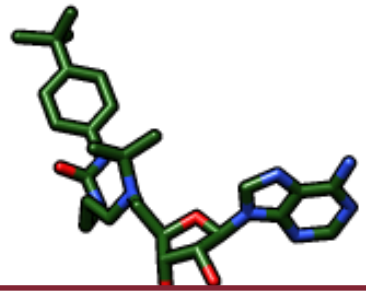


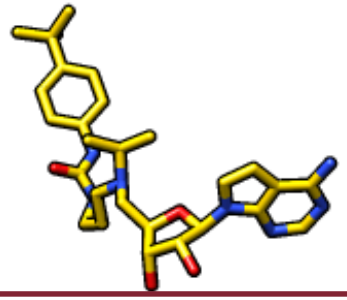






Experimental **C**onformation **C**ross-**D**ocking

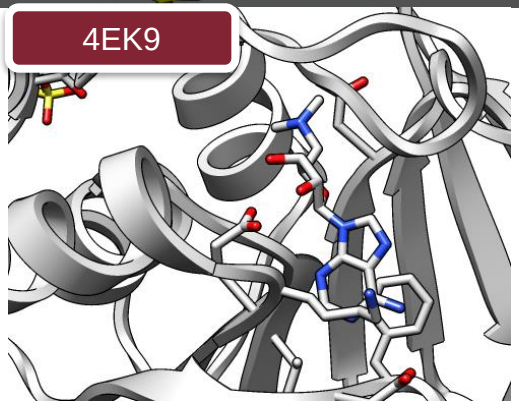


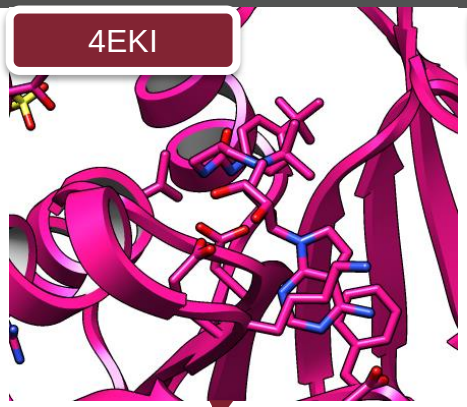


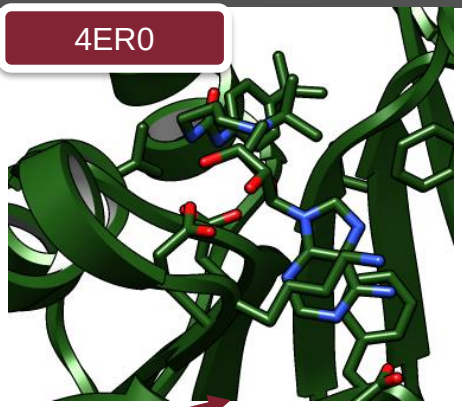
Random **C**onformation **C**ross-**D**ocking

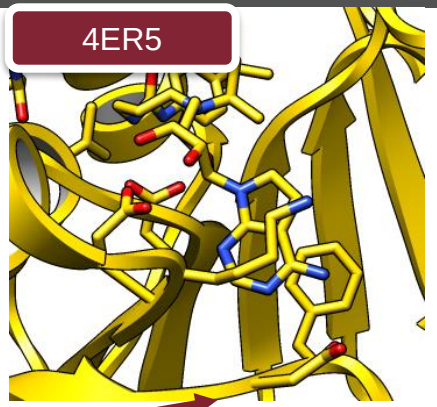
Docking Assessment

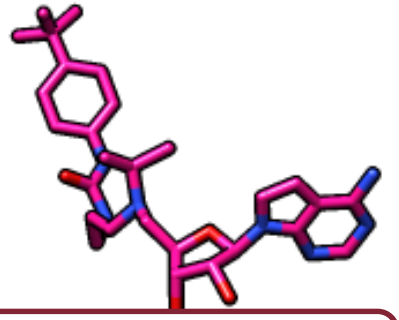
by www.RCMD.it



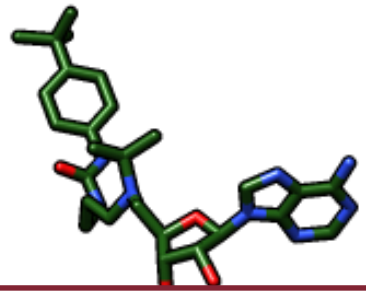


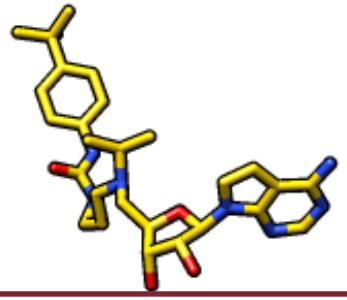






Experimental **C**onformation **C**ross-**D**ocking

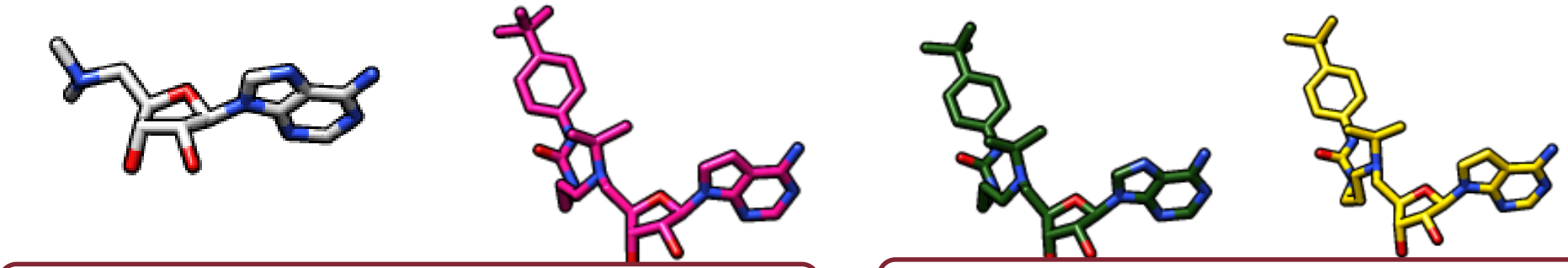




Random **C**onformation **C**ross-**D**ocking

Docking Assessment

by www.RCMD.it

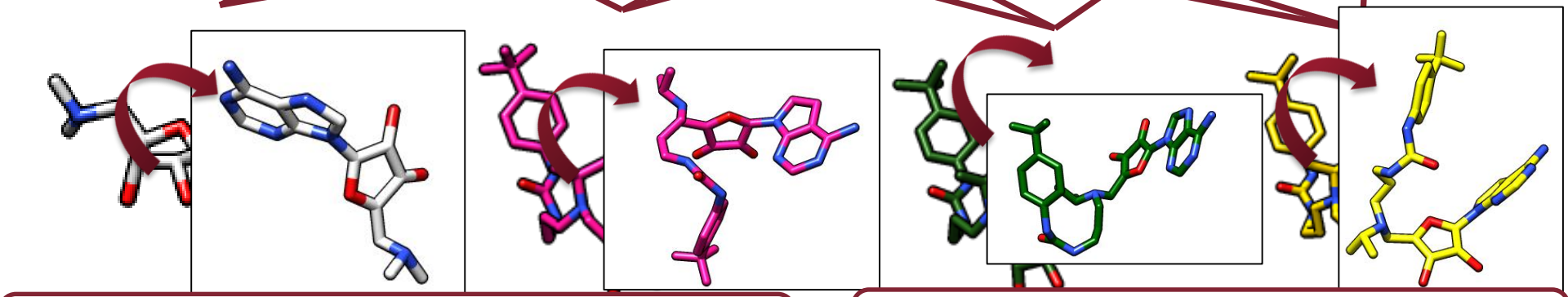


Experimental **C**onformation **C**ross-**D**ocking

Random **C**onformation **C**ross-**D**ocking

Docking Assessment

by www.RCMD.it



Experimental **C**onformation **C**ross-**D**ocking

Random **C**onformation **C**ross-**D**ocking

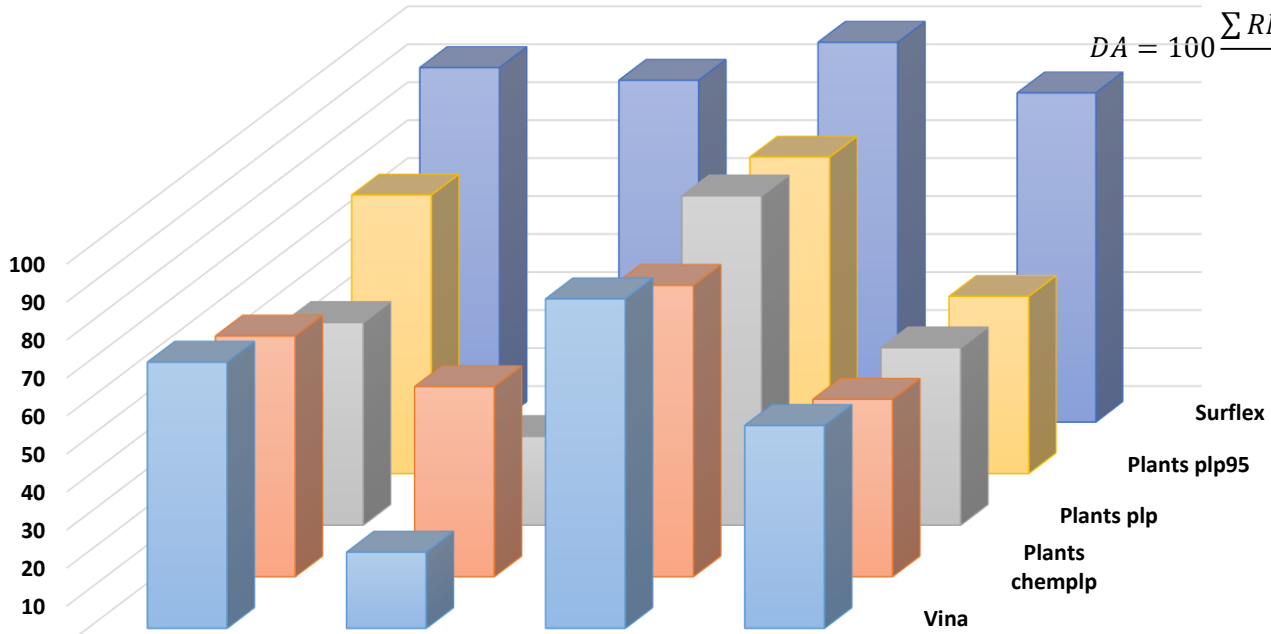
Docking Assessment

by www.RCMD.it

Docking Accuracy

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N} d_i^2}{N}}$$

$$DA = 100\frac{\sum RMSD_{(0-2)} + \sum RMSD_{(2-3)}}{N}$$



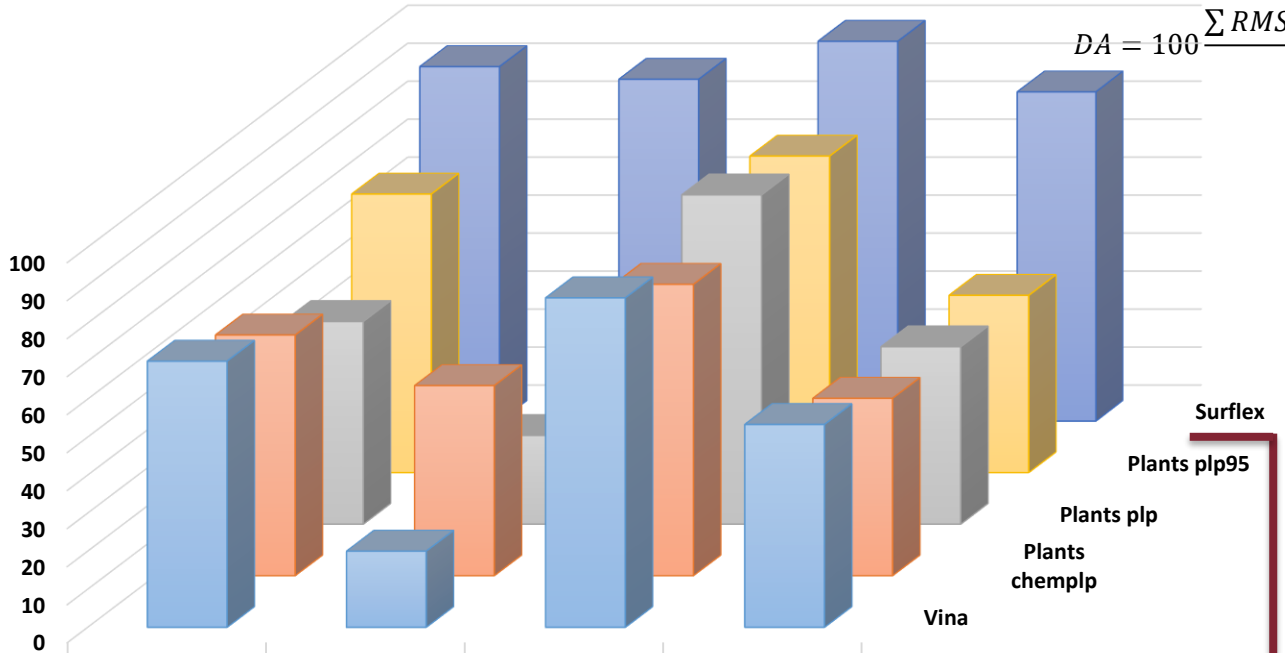
	RCRD	RCCD	ECRD	ECCD
Vina	70	20	86,67	53,33
Plants chemplp	63,33	50	76,67	46,67
Plants plp	53,33	23,33	86,67	46,67
Plants plp95	73,33	40	83,33	46,67
Surflex	93,33	90	100	86,67

RCRD → Random Conformational-ReDocking
RCCD → Random Conformational Cross-Docking
ECRD → Experimental Conformational Re-Docking
ECCD → Experimental Conformational Cross-Docking

Docking Accuracy

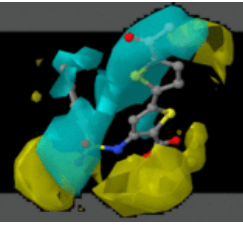
$$RMSD = \sqrt{\frac{\sum_{i=1}^{N} d_i^2}{N}}$$

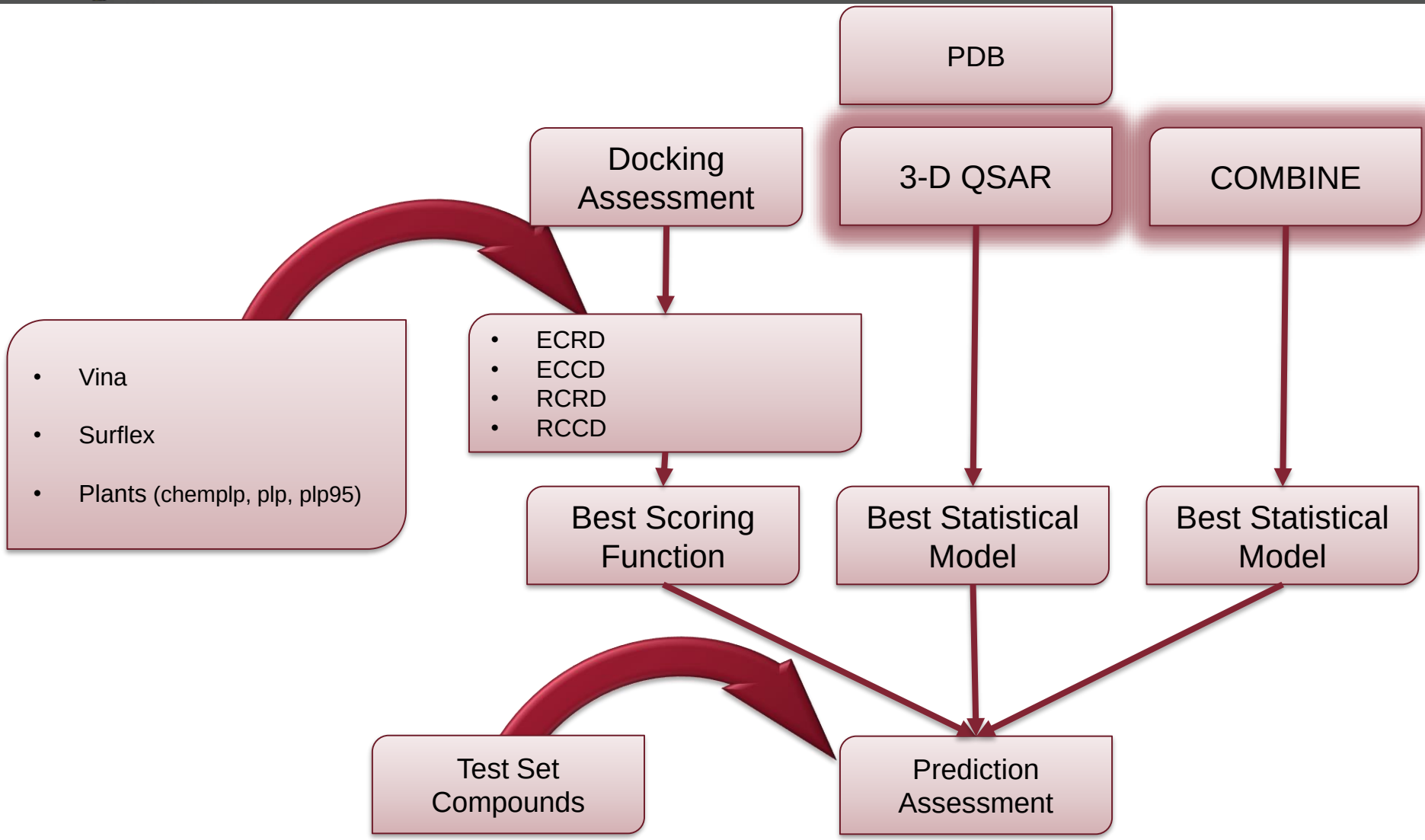
$$DA = 100\frac{\sum RMSD_{(0-2)} + \sum RMSD_{(2-3)}}{N}$$



	RCRD	RCCD	ECRD	ECCD
Vina	70	20	86,67	53,33
Plants chemplp	63,33	50	76,67	46,67
Plants plp	53,33	23,33	86,67	46,67
Plants plp95	73,33	40	83,33	46,67
Surflex	93,33	90	100	86,67

Workflow




PDB
Docking Assessment
3-D QSAR
COMBINE
• Vina
• Surflex
• Plants (chemplp, plp, plp95)
• ECRD
• ECCD
• RCRD
• RCCD
Best Scoring Function
Best Statistical Model
Best Statistical Model
Test Set Compounds
Prediction Assessment

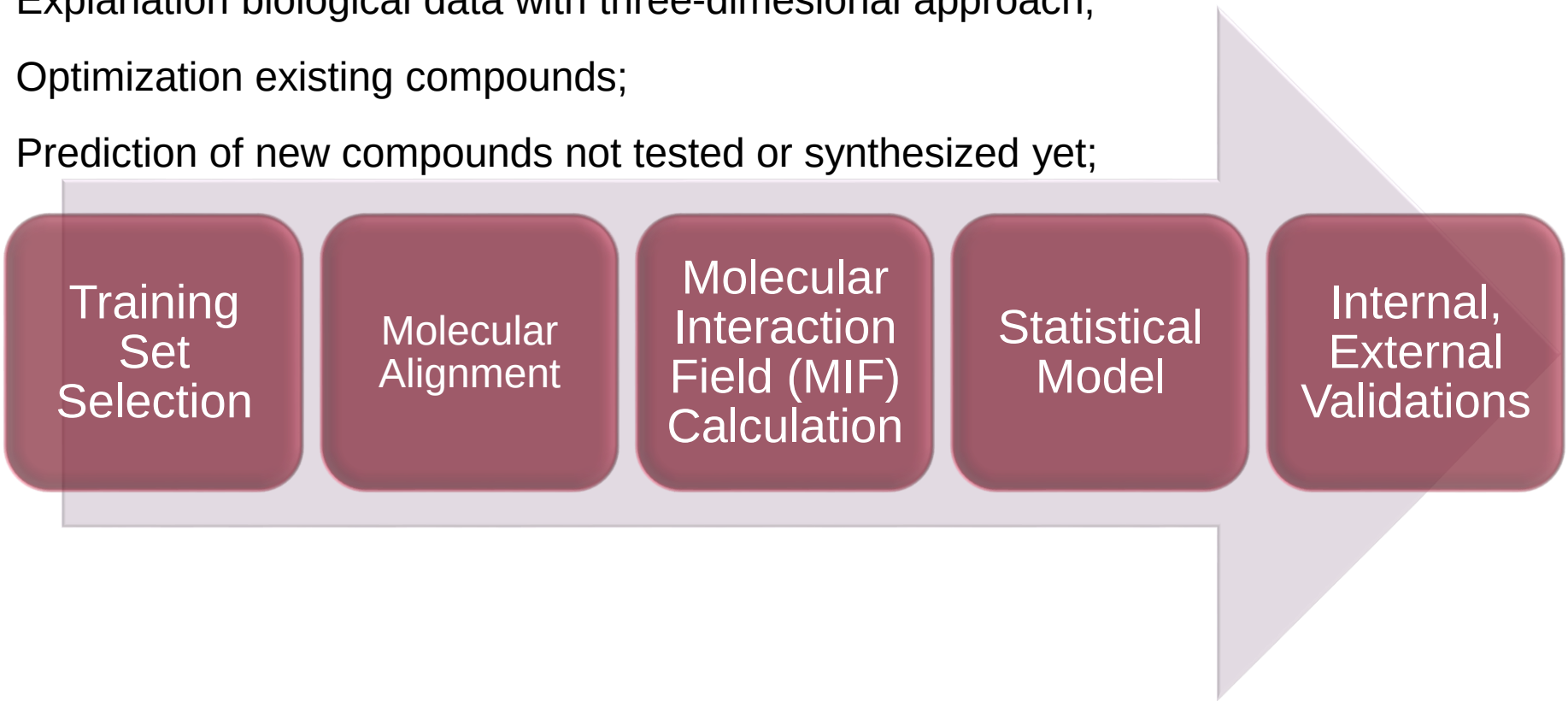
(3-D QSAR)

3-D Quantitative **S**tructure-**A**ctivity **R**elationships

Explanation biological data with three-dimesional approach;

Optimization existing compounds;

Prediction of new compounds not tested or synthesized yet;



(3-D QSAR)

3-D Quantitative **S**tructure-**A**ctivity **R**elationships

Explanation biological data with three-dimesional approach;

Optimization existing compounds;

Prediction of new compounds not tested or synthesized yet;

Training Set Selection → Molecular Alignment → Molecular Interaction Field (MIF) Calculation → Statistical Model → Internal, External Validations

(3-D QSAR)

3-D Quantitative **S**tructure-**A**ctivity **R**elationships

Explanation biological data with three-dimesional approach;

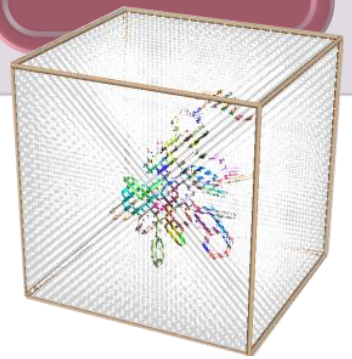
Optimization existing compounds;

Prediction of new compounds not tested or synthesized yet;

Training Set Selection

Molecular Alignment

Molecular Interaction Field (MIF) Calculation



Statistical Model

Internal, External Validations

(3-D QSAR)

3-D **Q**uantitative **S**tructure-**A**ctivity **R**elationships

Explanation biological data with three-dimesional approach;

Optimization existing compounds;

Prediction of new compounds not tested or synthesized yet;

Training Set Selection → Molecular Alignment → Molecular Interaction Field (MIF) Calculation → Statistical Model → Internal, External Validations

$$Y=\beta_0 + \beta_1 x_1 + \beta_a x_a + \ldots.. + \beta_n x_n$$

(3-D QSAR)

3-D Quantitative **S**tructure-**A**ctivity **R**elationships

Explanation biological data with three-dimesional approach;

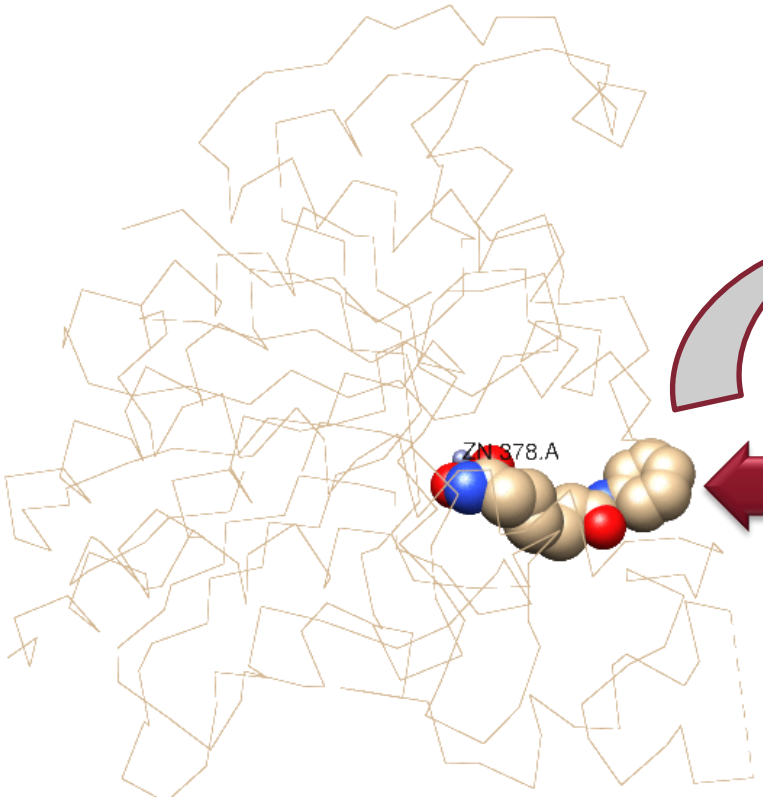
Optimization existing compounds;

Prediction of new compounds not tested or synthesized yet;

Training Set Selection → Molecular Alignment → Molecular Interaction Field (MIF) Calculation → Statistical Model → Internal, External Validations

$$q^2 = 1 - \frac{\sum(y_{exp} - y_{pred})^2}{\sum y_{exp} - \bar{y})^2}$$

$$SDEP = \sqrt{\sum \frac{(y_{exp} - y_{pred})^2}{n-1}}$$



SAHA

$$\Delta U = \sum_{i=1}^{n_l}\sum_{j=1}^{n_r} u_{ij}^{\mathrm{VDW}} + \sum_{i=1}^{n_l}\sum_{j=1}^{n_r} u_{ij}^{\mathrm{ELE}}$$

Journal of
Medicinal Chemistry

Subscriber access provided by UNIVERSITA DI ROMA LA SAPIENZA

Prediction of Drug Binding Affinities by Comparative Binding Energy Analysis

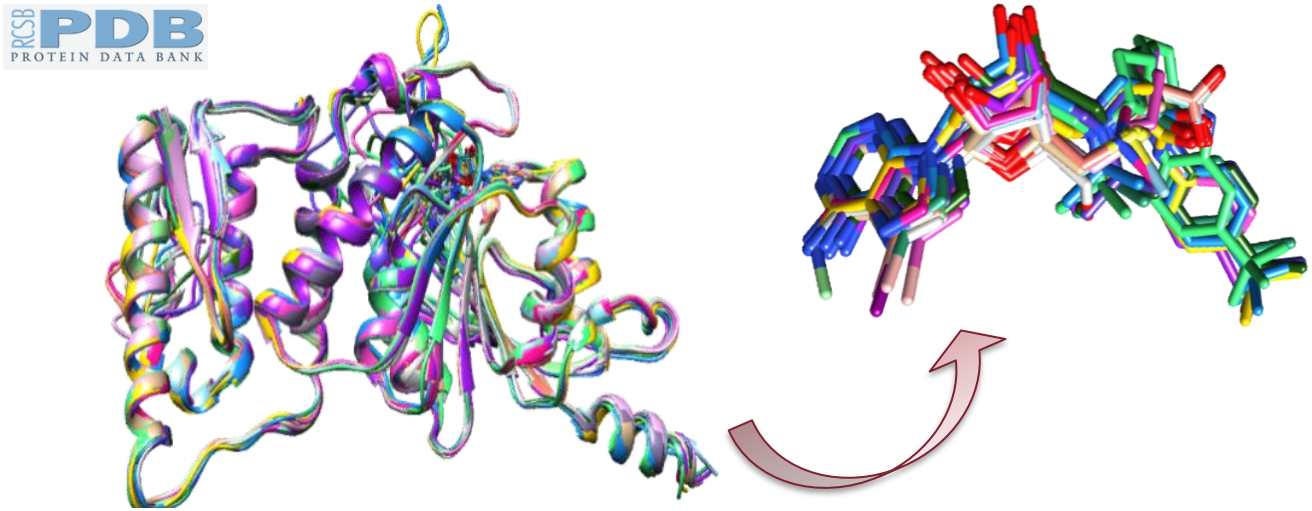
Angel R. Ortiz, M. Teresa Pisabarro, Federico Gago, and Rebecca C. Wade

J. Med. Chem., 1995, 38 (14), 2681-2691• DOI: 10.1021/jm00014a020 • Publication Date (Web): 01 May 2002

	Res1	Res2	Res3	.....	Res-n		
Cplx 1	X1	X2	X3	.....	Xn		Y1
Cplx 2	X1	X2	X3	.....	Xn		Y2
Cplx 3	X1	X2	X3	.....	Xn		Y3
...	.....						...
...	.....						...
...	.....						...
...	.....						...
...	.....						...
...	.....						...
...	.....						...
...	.....						...
...	.....						...
Cplx m	X1	X2	X3	.....	Xn		Ym

PLS

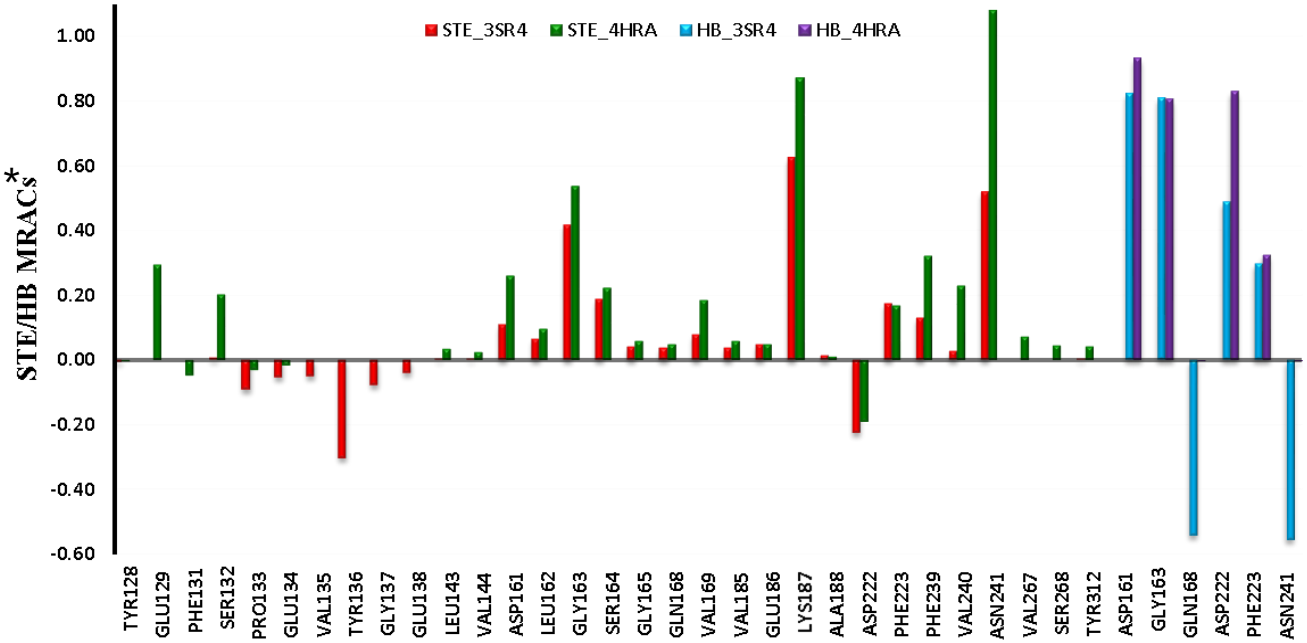
COMBINE Model



OA Hydrogen binding acceptor oxygen
N Amidic nitrogen

Mod	Probe	r^2	q^2_{L5O}	$SDEP_{L5O}$	PC	GRID	Cut Off	Zeroing	Min STD
1	A	0.83	0.70	0.994	1	2.45	4	0.008	0.005
2		0.83	0.70	0.995	1	2.45	4	0.010	0.005
3	D	0.96	0.70	0.992	3	2.10	5	0.002	0.050
4	OA	0.97	0.79	0.837	3	2.30	4	0.003	0.015
5	N	0.97	0.79	0.834	3	2.30	5	0.002	0.015
6		0.96	0.78	0.836	3	2.30	5	0.002	0.015
7		0.84	0.60	1.145	2	2.10	3	0.005	0.050
8	d	0.85	0.79	0.822	1	1.95	1	0.001	0.015

#	Molecule-Residue Interaction Field Combination	PC	Fitting		LOO		L5O		LHO	
			r^2	SDEC	q^2	SDEP	q^2	SDEP	q^2	SDEP
7s	STE.HB	4	0.96	0.36	0.80	0.81	0.78	0.85	0.69	1.01

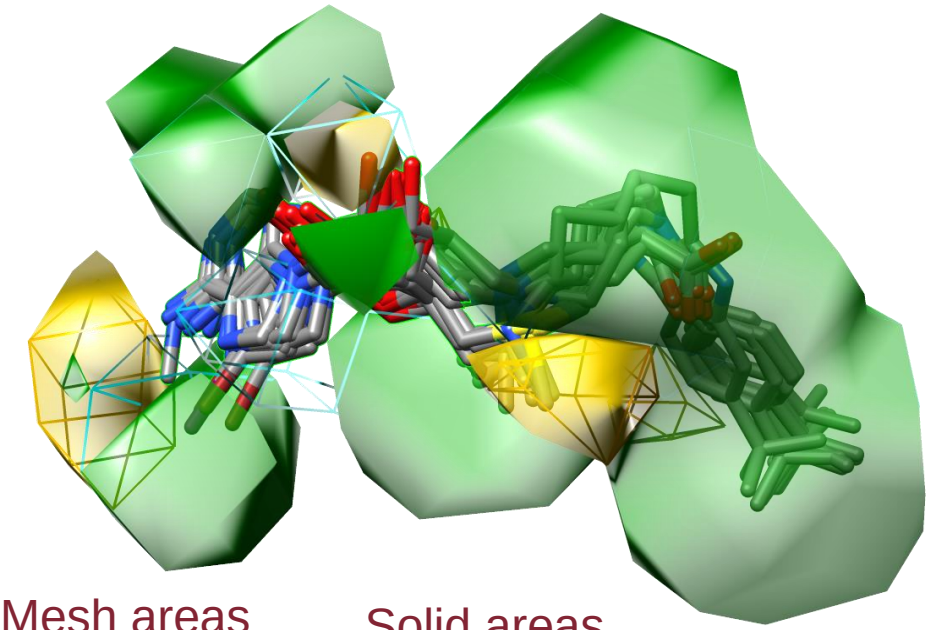


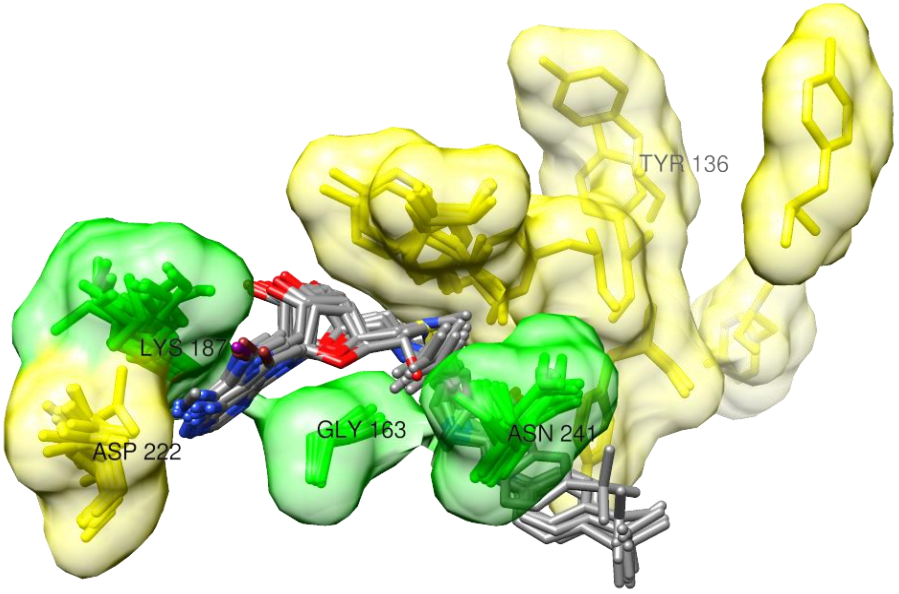
*STE/HB molecule-residue activity contribution

Average activity contribution plot

3-D QSAR: OA and N probes

COMBINE: STE.HD molecular interaction fields



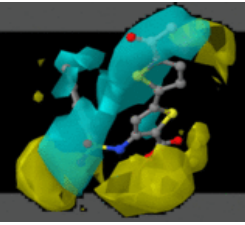


Mesh areas

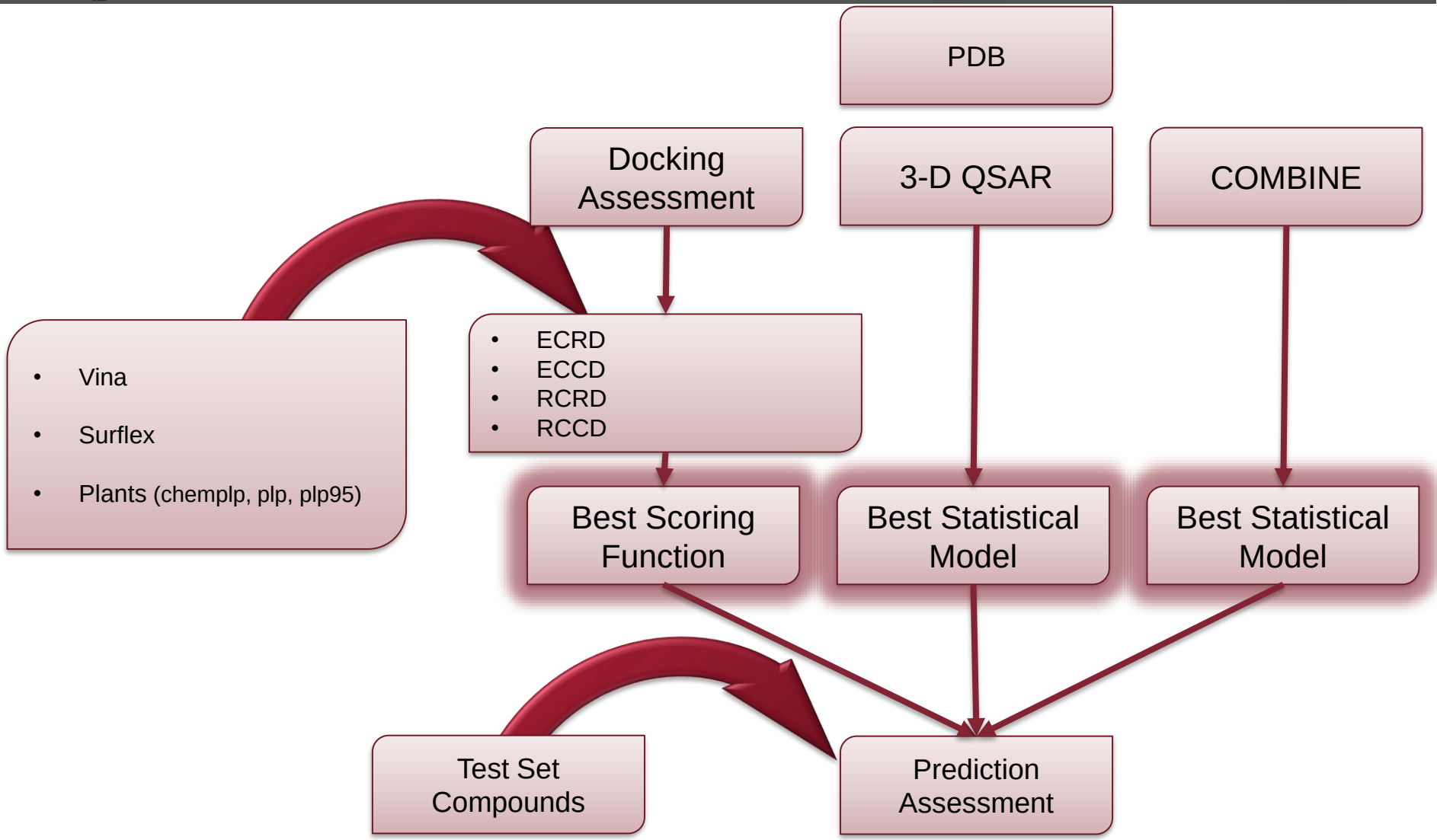
Solid areas

Solid areas

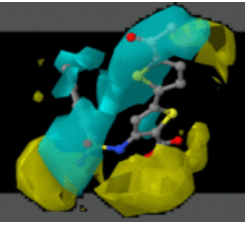
Overflow



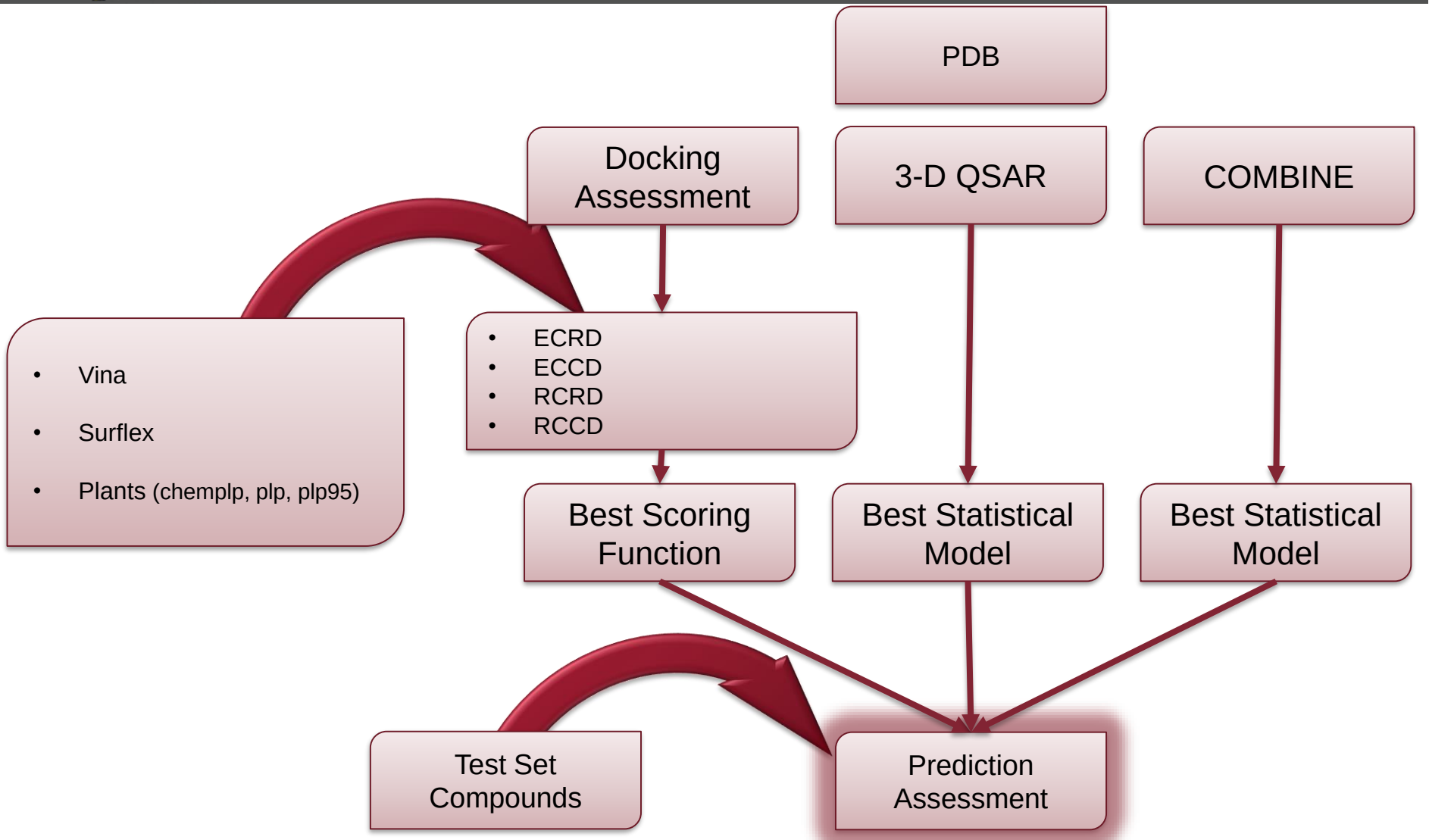
by www.RCMD.it



Overflow



by www.RCMD.it

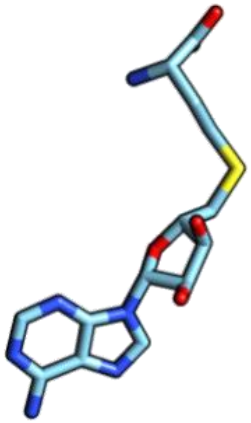


Modelled External Test Set

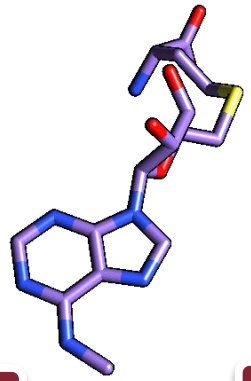
by www.RCMD.it

PDB ID	Activity (pIC_{50}[a] o pK_i[b])
1TS	4.00[b]
2TS	4.34[b]
3TS	4.66[b]
4TS	4.66[b]
5TS	4.70[b]
6TS	4.74[b]
7TS	4.80[a]
8TS	4.80[b]
9TS	4.92[b]
10TS	5.72[b]
11TS	5.96[b]
12TS	5.96[b]
13TS	6.07[b]
14TS	6.92[a]
15TS	6.26[b]
16TS	6.96[a]
17TS	7.09[b]
18TS	7.15[b]
19TS	7.24[b]
20TS	7.42[a]
21TS	8.40[b]
22TS	8.89[b]
23TS	8.96[b]
24TS	9.10[b]

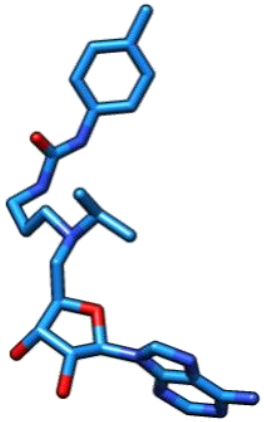
3QOX



3SR4

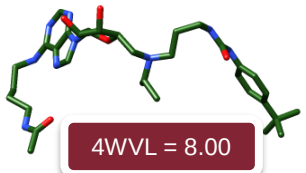


4EQZ

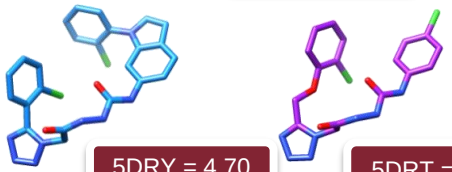


4EKG

Crystallized External Test Set

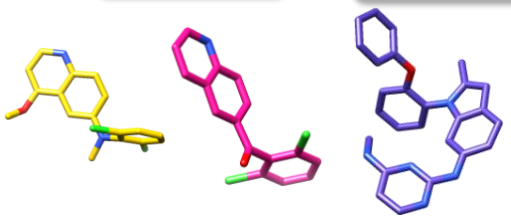


4WVL = 8.00



5DRY = 4.70

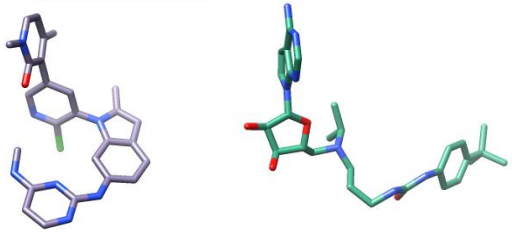
5DRT = 5.36



5DTR = 4.41

5DTQ = 3.86

5DT2 = 6.82



5DSX = 8.00

4ER3 = 9.53

ACS chemical biology
Letters
pubs.acs.org/acschemicalbiology

Structure-Guided DOT1L Probe Optimization by Label-Free Ligand Displacement

ACS Medicinal Chemistry Letters
Letter
pubs.acs.org/acsmedchemlett

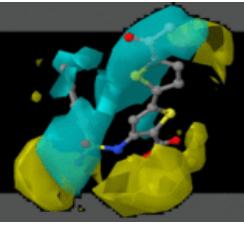
Optimization of a Fragment-Based Screening Hit toward Potent DOT1L Inhibitors Interacting in an Induced Binding Pocket

ACS Medicinal Chemistry Letters
Letter
pubs.acs.org/acsmedchemlett

Discovery of Novel Dot1L Inhibitors through a Structure-Based Fragmentation Approach

Chao Chen,[†] Hugh Zhu,[†] Frédéric Stauffer,[‡] Giorgio Caravatti,[‡] Susanne Vollmer,[‡] Rainer Machauer,[‡] Philipp Holzer,[‡] Henrik Möbitz,[‡] Clemens Scheufler,[‡] Martin Klumpp,[‡] Ralph Tiedt,[‡] Kim S. Beyer,[‡] Keith Calkins,[‡] Daniel Guthy,[‡] Michael Kiffe,[‡] Jeff Zhang,[†] and Christoph Gaul*[,‡]

[‡]Novartis Institutes for Biomedical Research, 4002 Basel, Switzerland
[†]Novartis Institutes for Biomedical Research, Shanghai 201203, China



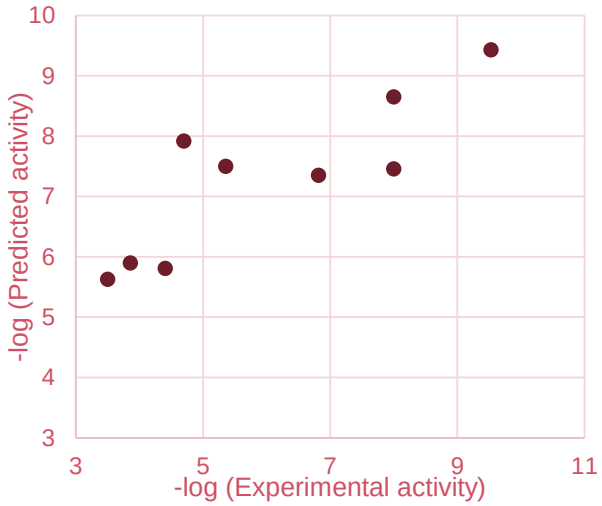
Statistical Results

by www.RCMD.it

Model	Probe	r^2	q^2_{L5O}	$SDEP_{L5O}$	PC	GRID	Cut Off	Zeroing	Min STD
1	A	0.83	0.70	0.994	1	2.45	4	0.008	0.005
2	C	0.83	0.70	0.995	1	2.45	4	0.010	0.005
3	HD	0.96	0.70	0.992	3	2.10	5	0.002	0.050
4	OA	0.97	0.79	0.837	3	2.30	4	0.003	0.015
5	N	0.97	0.79	0.834	3	2.30	5	0.002	0.015
6	NA	0.96	0.78	0.836	3	2.30	5	0.002	0.015
7	e	0.84	0.60	1.145	2	2.10	3	0.005	0.050
8	d	0.85	0.79	0.822	1	1.95	1	0.001	0.015

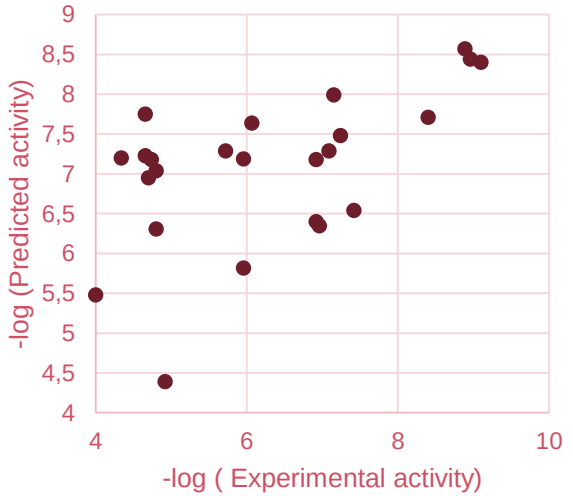
Cristallized External Test Set

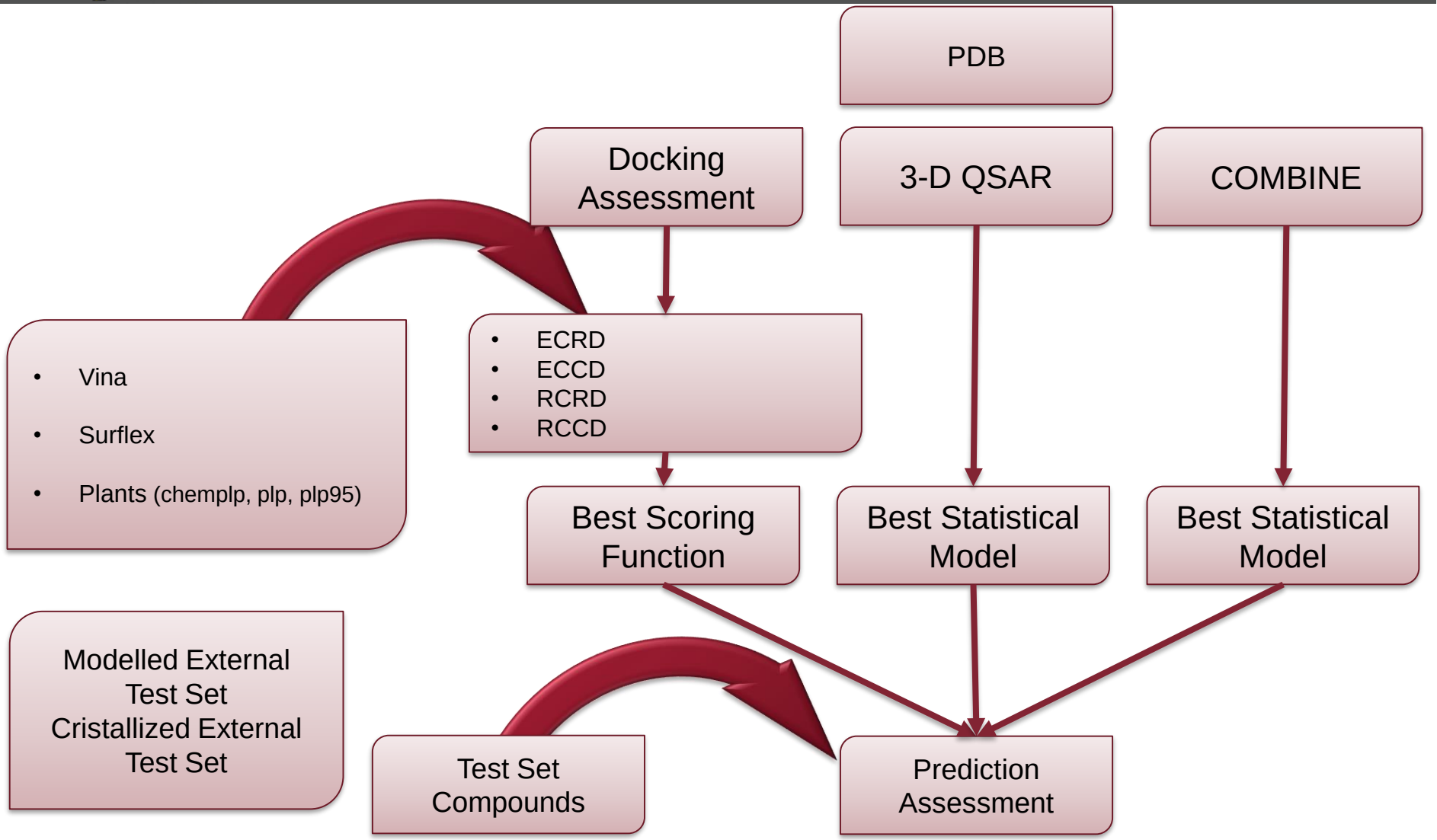
Predicted activity OA probe

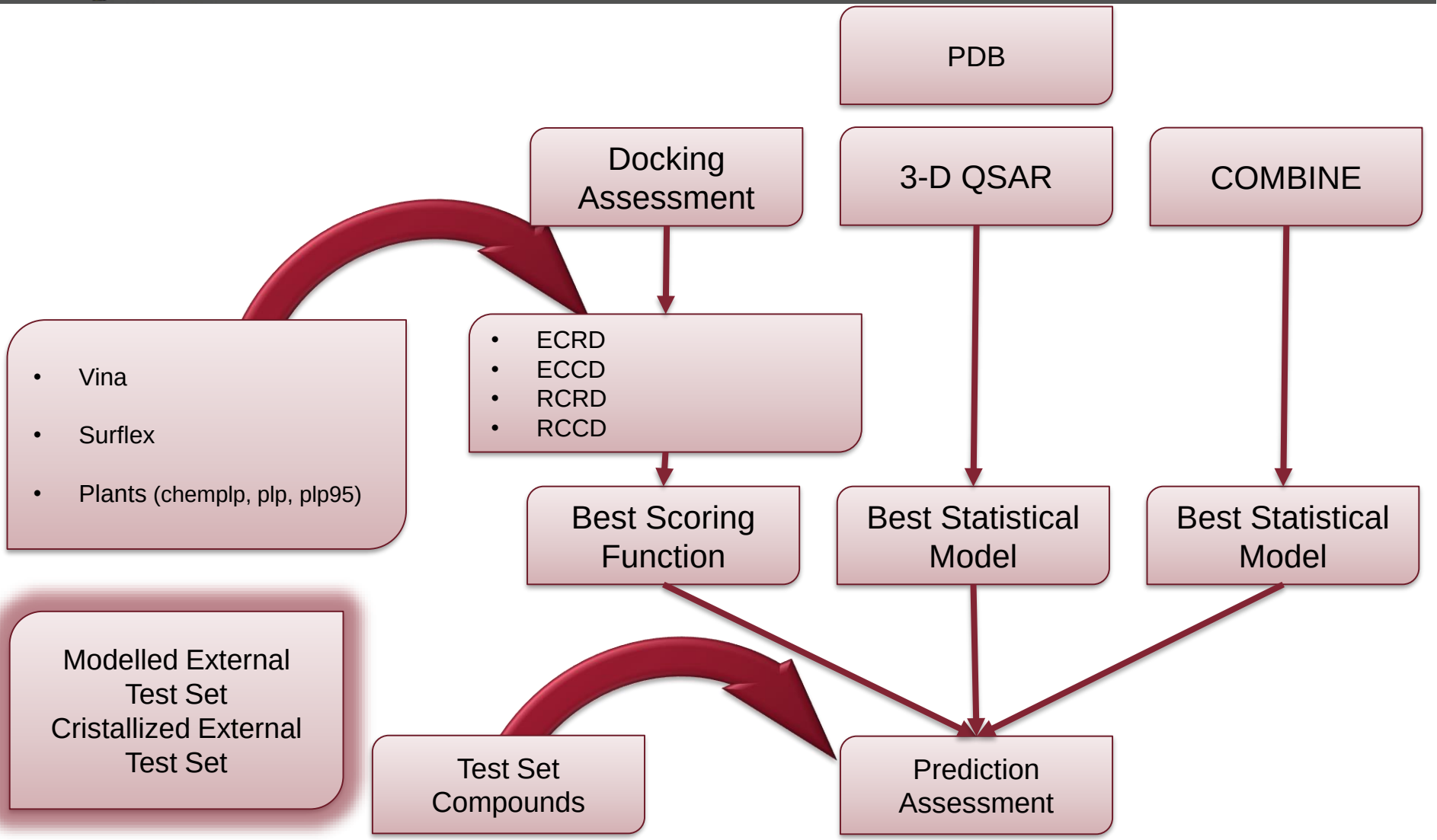


Modelled External Test Set

Predicted activity OA probe




PDB
Docking Assessment
3-D QSAR
COMBINE
Vina
Surflex
Plants (chemplp, plp, plp95)
ECRD
ECCD
RCRD
RCCD
Best Scoring Function
Best Statistical Model
Best Statistical Model
Modelled External Test Set
Cristallized External Test Set
Test Set Compounds
Prediction Assessment


PDB
Docking Assessment
3-D QSAR
COMBINE
Vina
Surflex
Plants (chemplp, plp, plp95)
ECRD
ECCD
RCRD
RCCD
Best Scoring Function
Best Statistical Model
Best Statistical Model
Modelled External Test Set
Cristallized External Test Set
Test Set Compounds
Prediction Assessment

Journal of Medicinal Chemistry

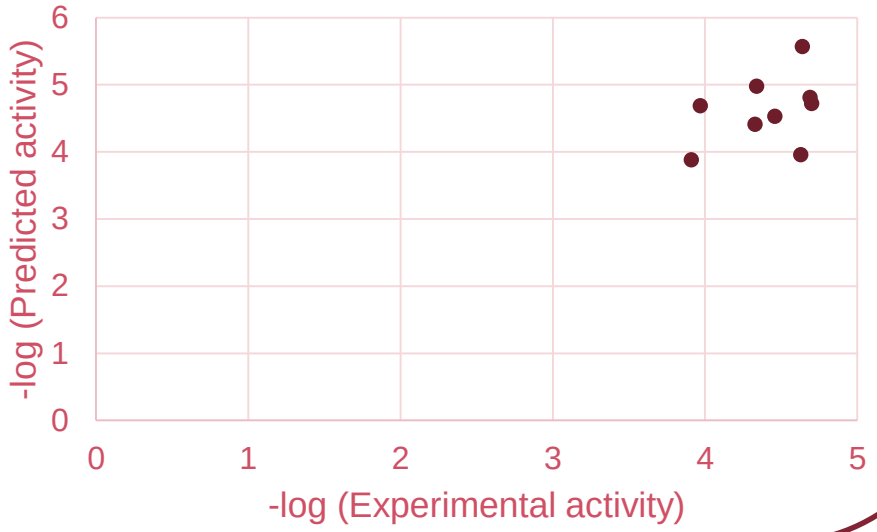
Article

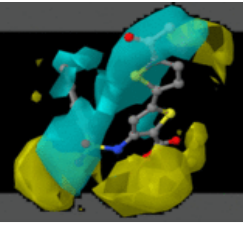
pubs.acs.org/jmc

Discovery of Novel Disruptor of Silencing Telomeric 1-Like (DOT1L) Inhibitors using a Target-Specific Scoring Function for the (*S*)-Adenosyl-L-methionine (SAM)-Dependent Methyltransferase Family

Yulan Wang,[#,†,‡,¶] Linjuan Li,[#,§,¶] Bidong Zhang,[#,‡] Jing Xing,[#,‡] Shijie Chen,[#,‡] Wei Wan,[#,‡] Yakai Song,[#,II] Hao Jiang,[#,‡] Hualiang Jiang,[#,‡,§] Cheng Luo,[*,#,‡] and Mingyue Zheng[*,#,‡]

Prediction Set





Overflow

by www.RCMD.it

87 compounds
Prof Mai's Lab

PREDICTION SET

Validated

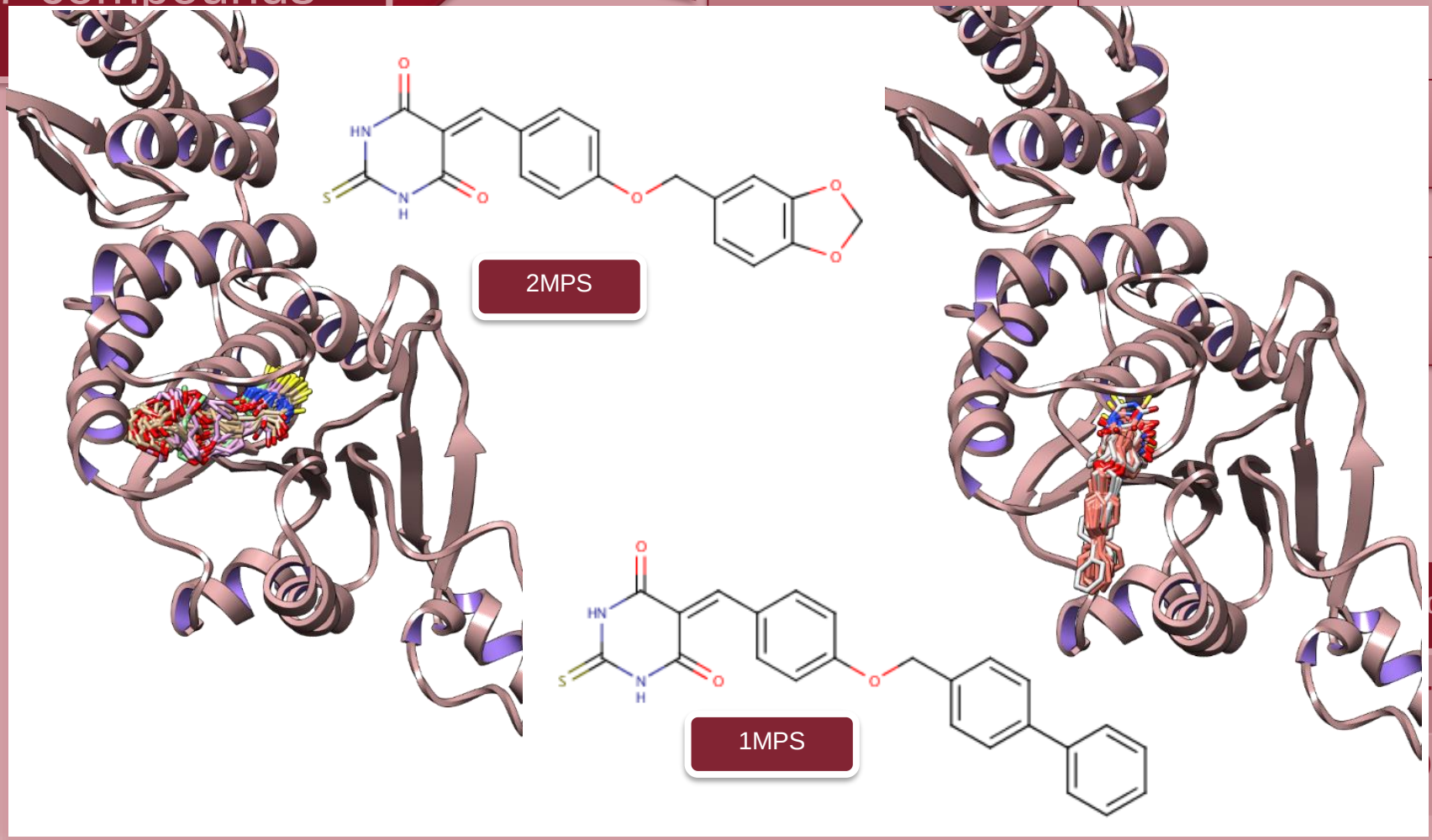
Docking | 3-D QSAR | COMBINE

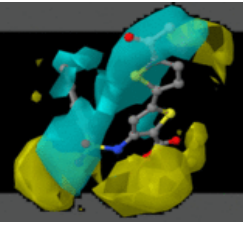
Best Scoring Function:Surflex | Best Statistical Model:OA, N | Best Statistical Model:STE.HB

#	Probe OA	Probe N	COMBINE	Average	pIC_{50}
1MPS	6.42	6.63	4.66	5.90	**5.14**
2MPS	6.99	6.99	4.63	6.20	**5.10**

87 compounds

PREDICTION SET





Conclusion

- Using a Docking assessment was identified a best docking software for DOT1L target;
- 3-D QSAR and COMBINE models with high robustness were built for DOT1L target;
- Using Docking, 3-D QSAR and COBINE protocol a very comprehensive SB and LB study was developed;
- Using a presented protocol was identified a new chemical scaffold as starting point to synthesize new promising compounds;
- This protocol was accepted on Journal of Computer Aided Molecular Design (JCAM) *(Sabatino, M.; Rotili, D., Patsilinakos A.; Forgione, M.; Tomaselli, D.; Alby, F.; Arimondo, P.; Mai, A.; Ragno, R., Disruptor of Telomeric Silencing 1-Like (DOT1L): Disclosing a New Class of Non-nucleoside Inhibitors by Means of Ligand-Based and Structure-Based Approaches.)*

Essenziali Oils

Essential Oils are organic complex mixtures extracted from *aromatic* plants thanks physical process.

High percentage of:

- Terpenes
- Oxigenated compounds

To extract the essential oils are available several techniques:

- Ultrasound
- Steam distillation, Hydro-distilation
- Solvent extraction, Supercritical fluids extraction

Essenziali Oils

Essential Oils are organic complex mixtures extracted from *aromatic* plants thanks physical process.

High percentage of:

Terpenes

Oxigenated compounds

To extract the essential oils are available several techniques:

Ultrasound

Steam distillation, Hydro-distilation

Solvent extraction, Supercritical fluids extraction

Steam distillation

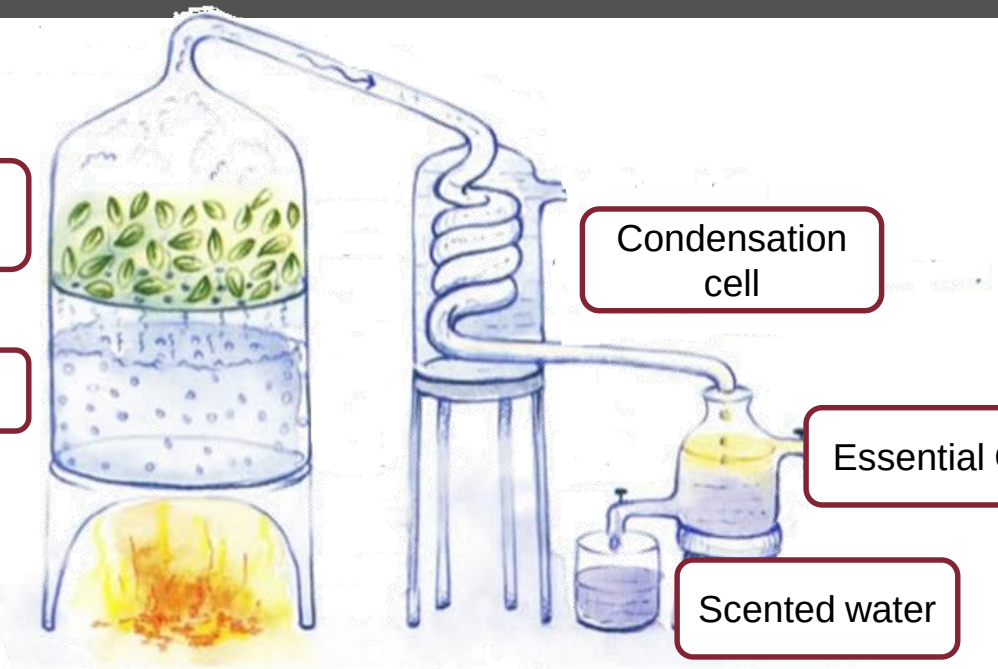
Aromatic Plants

Water

Condensation cell

Essential Oils

Scented water

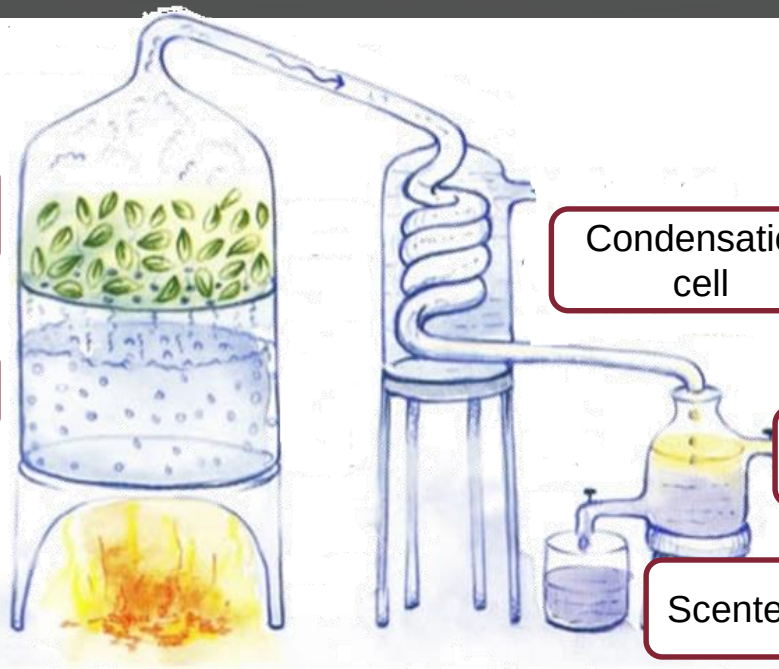


1. *Calamintha nepeta* L.; **2**. *Ridolfia segetum* Mor.; **3**. *Sideritis purpurea* Tal. Ex Ben.; **4**. *Melissa alttissima* Sm.; **5.** *Mentha suaveolens*

Steam distillation

Aromatic Plants

Water



Condensation cell

Essential Oils

Scented water

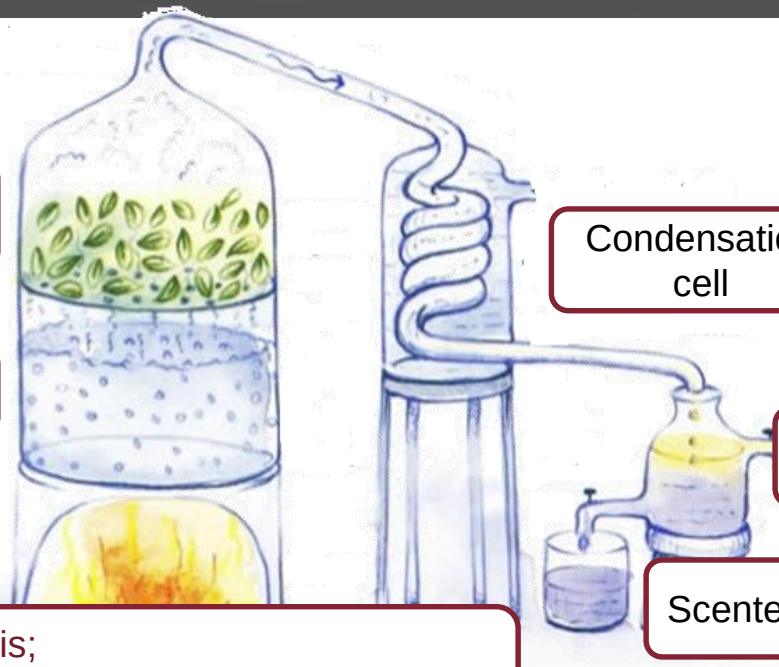
1. *Calamintha nepeta* L.; **2**. *Ridolfia segetum* Mor.; **3**. *Sideritis purpurea* Tal. Ex Ben.; **4**. *Melissa alttissima* Sm.; **5.** *Mentha suaveolens*

Month
- June
- July
- August
- September
- October

Sampling temp
- 1h
- 2h
- 3h
- 6h
- 12h
- 24h

Kind of plants
- Fresh
- Dry

Steam distillation



Aromatic Plants

Water

Condensation cell

Essential Oils

Scented water

1. *Calamintha nepeta* L.; **2**. *Ridolfia segetum* Mor.; **3**. *Sideritis purpurea* Tal. Ex Ben.; **4**. *Melissa alttissima* Sm.; **5.** *Mentha suaveolens*

-GC/MS analysis;
(*Prof.ssa Garzoli, Prof. Pepi- Roma*)

-Anticandidal;
(Prof. Mafredini - Ferrara)

-Antibacterial;
(Prof.ssa Oliva - Roma)

-Antibiofilm;
(Prof.ssa Selan - Roma)

-Antiviral;
(Prof.ssa Palamara - Roma)

Month
June
July
August
September
October

Sampling temp
1h
2h
3h
6h
12h
24h

Kind of plants
Fresh
Dry

Steam distillation

Aromatic Plants

Water

Condensation cell

Essential Oils

Scented water

1. *Calamintha nepeta* L.; **2**. *Ridolfia segetum* Mor.; **3**. *Sideritis purpurea* Tal. Ex Ben.; **4**. *Melissa alttissima* Sm.; **5.** *Mentha suaveolens*

-GC/MS analysis;
(*Prof.ssa Garzoli, Prof. Pepi- Roma*)

-Anticandidal;
(Prof. Mafredini - Ferrara)

-Antibacterial;
(Prof.ssa Oliva - Roma)

-Antibiofilm;
(Prof.ssa Selan - Roma)

-***Antiviral;***
(Prof.ssa Palamara - Roma)

Month
June
July
August
September
October

Sampling temp
1h
2h
3h
6h
12h
24h

Kind of plants
Fresh
Dry

Starting Data

From the 90 Essential oils extracted from three different plants with the explained protocol, were chosen 37 for the Antiviral tests.

Calamintha Nepeta

- July
  - 1 e 2 hours
  - Mix 2 e 5
- August
  - 1, 2, 3 e 6 hours
  - Mix 1,3 e 5
- September
  - 1,2,3 e 6 hours
  - Mix 1,3 e 5
- October
  - 1, 2 e 6 hours
  - Mix 1,3 e 5

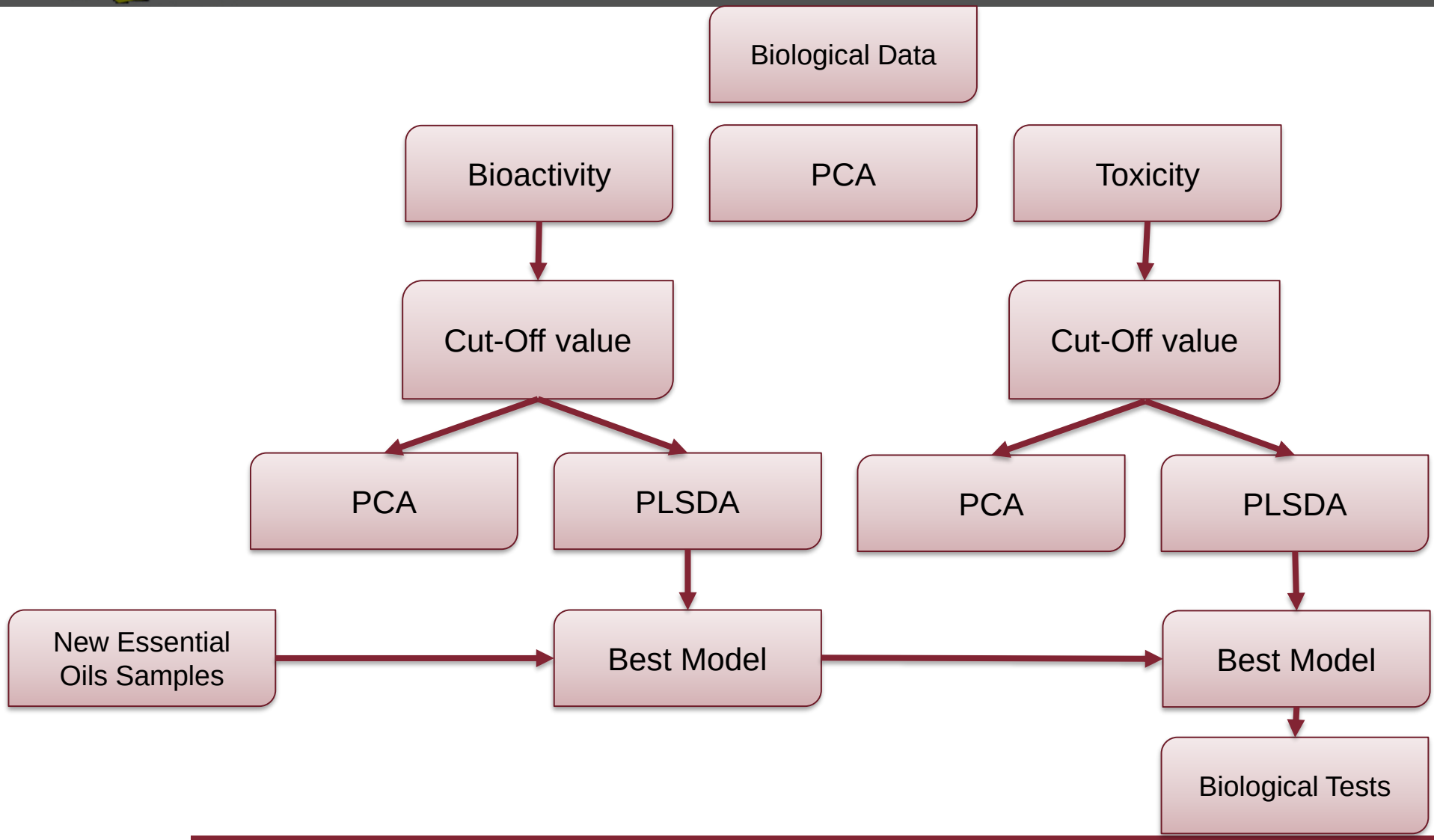
Foeniclum Vulgare

- Augsut
  - 2 e 6 hours
- September
  - 1, 2 e 6 hours
- October
  - 1, 3, 6 e 24 hours
  - Mix 3

Ridolfia segetum

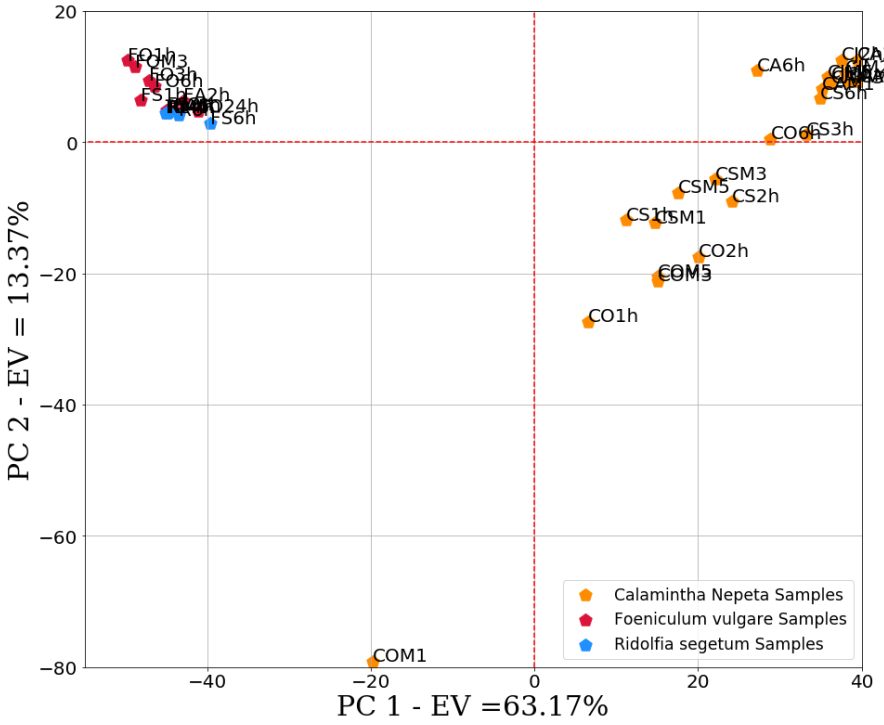
- July 6 e 24 hours
- Mix 4 e 6

Workflow


Biological Data
Bioactivity
PCA
Toxicity
Cut-Off value
Cut-Off value
PCA
PLSDA
PCA
PLSDA
New Essential Oils Samples
Best Model
Best Model
Biological Tests

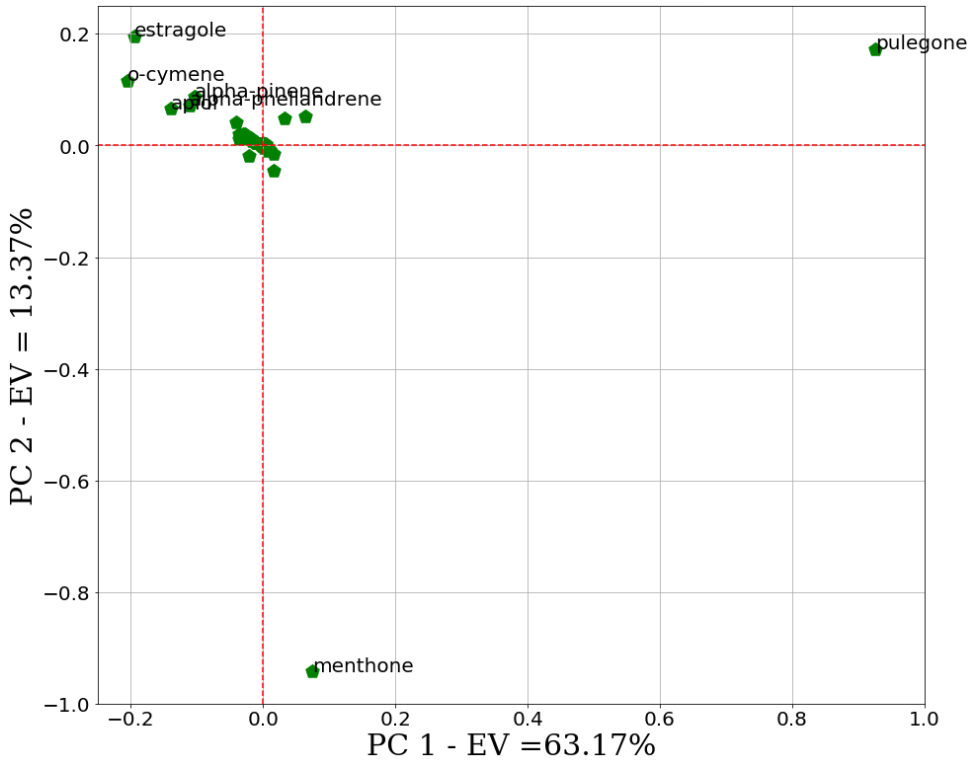
PCA

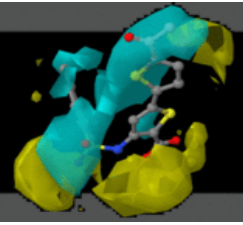
- Two different cluster;
- Foeniculum Vulgare more similar to Ridolfia than Calamintha nepeta;
- Calamintha Nepeta rich in Pulegone;



Score Plot

Loading Plot



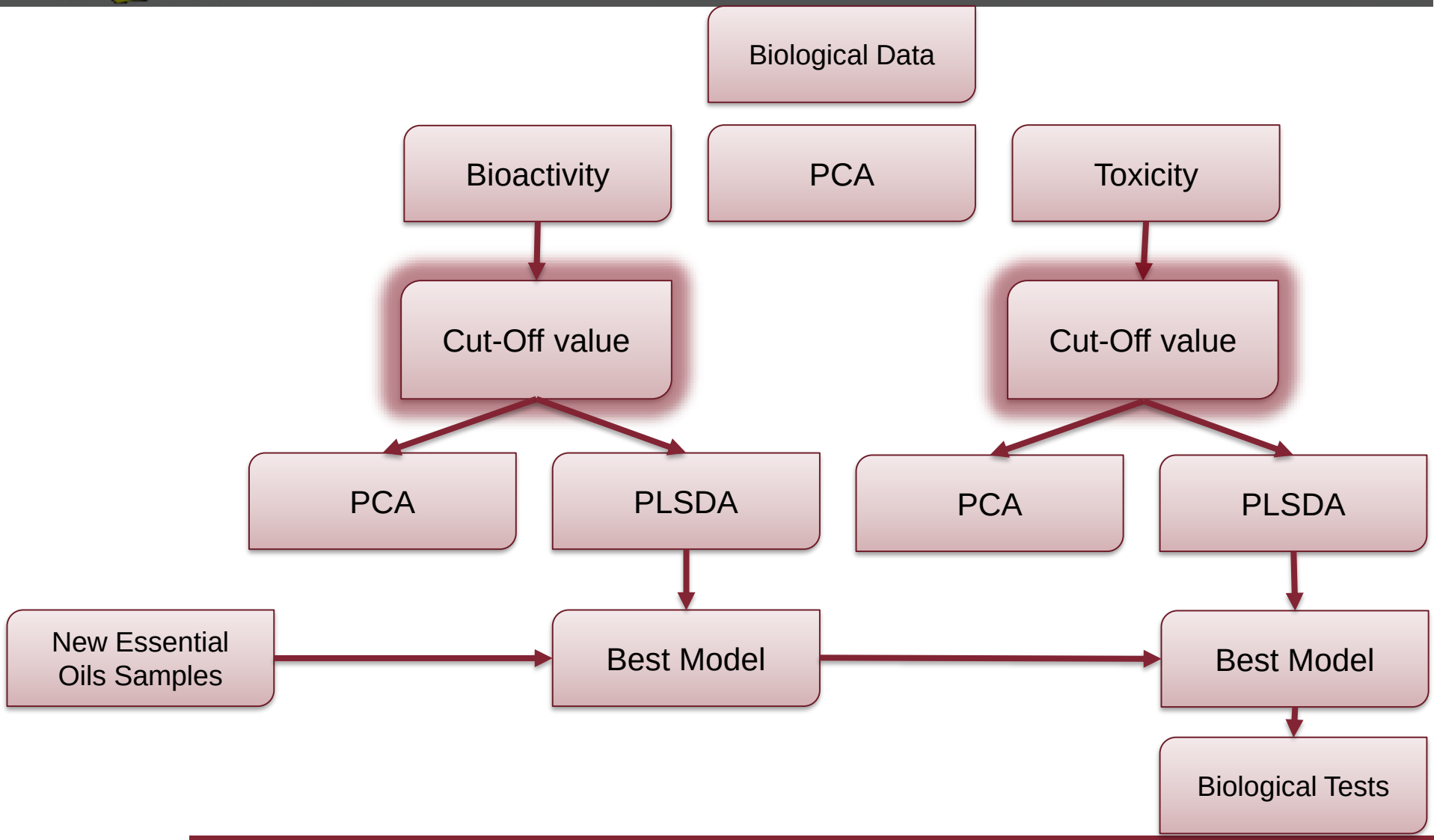


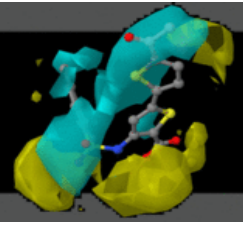
Biological Data

by www.RCMD.it

Sample	IC50 (mg/ml)	CC50 (mg/ml)	Sample	IC50 (mg/ml)	CC50 (mg/ml)	Sample	IC50 (mg/ml)	CC50 (mg/ml)
CS1	0,14	2,5	CJM1	0,328	2,50	***FO3***	***0,6465***	***1,02***
CS2	0,2063	2,9	CJM5	0,1353	3,14	***FO6***	***0,54025***	***1,694***
CS3	0,231	2,68	CA1	0,1617	2,84	***FO24***	***0,1765***	***0,355***
CS6	0,213	2,66	CA2	0,1517	4,71	***FOM3***	***0,582***	***1,7***
CSM1	0,174	2,00	CA3	0,175	2,50	***RJ6***	***0,431***	***0,405***
CMS3	0,236	2,12	***CA6***	***0,142***	***1,14***	***RJ24***	***0,18075***	***0,36***
CSM5	0,1225	1,93	CAM1	0,202	2,65	***RJM4***	***0,295***	***0,9***
CO1	0,21	2,50	CAM3	0,1537	3,00	***RJM6***	***0,2***	***0,33***
CO2	0,332	4,60	***CAM5***	***0,4395***	***1,10***			
CO6	0,174	2,22	***FA2***	***0,1495***	***1,45***			
COM1	0,282	2,76	***FA6***	***0,1605***	***1,815***			
COM3	0,3125	1,99	***FS1***	***0,1945***	***1,005***			
COM5	0,3645	2,46	***FS2***	***0,192***	***0,31***			
CJ1	0,2725	2,15	***FS6***	***0,192***	***1,513***			
CJ2	0,135	4,67	***FO1***	***0,7235***	***1,513***			

Overflow


Biological Data
Bioactivity
PCA
Toxicity
Cut-Off value
Cut-Off value
PCA
PLSDA
PCA
PLSDA
New Essential Oils Samples
Best Model
Best Model
Biological Tests

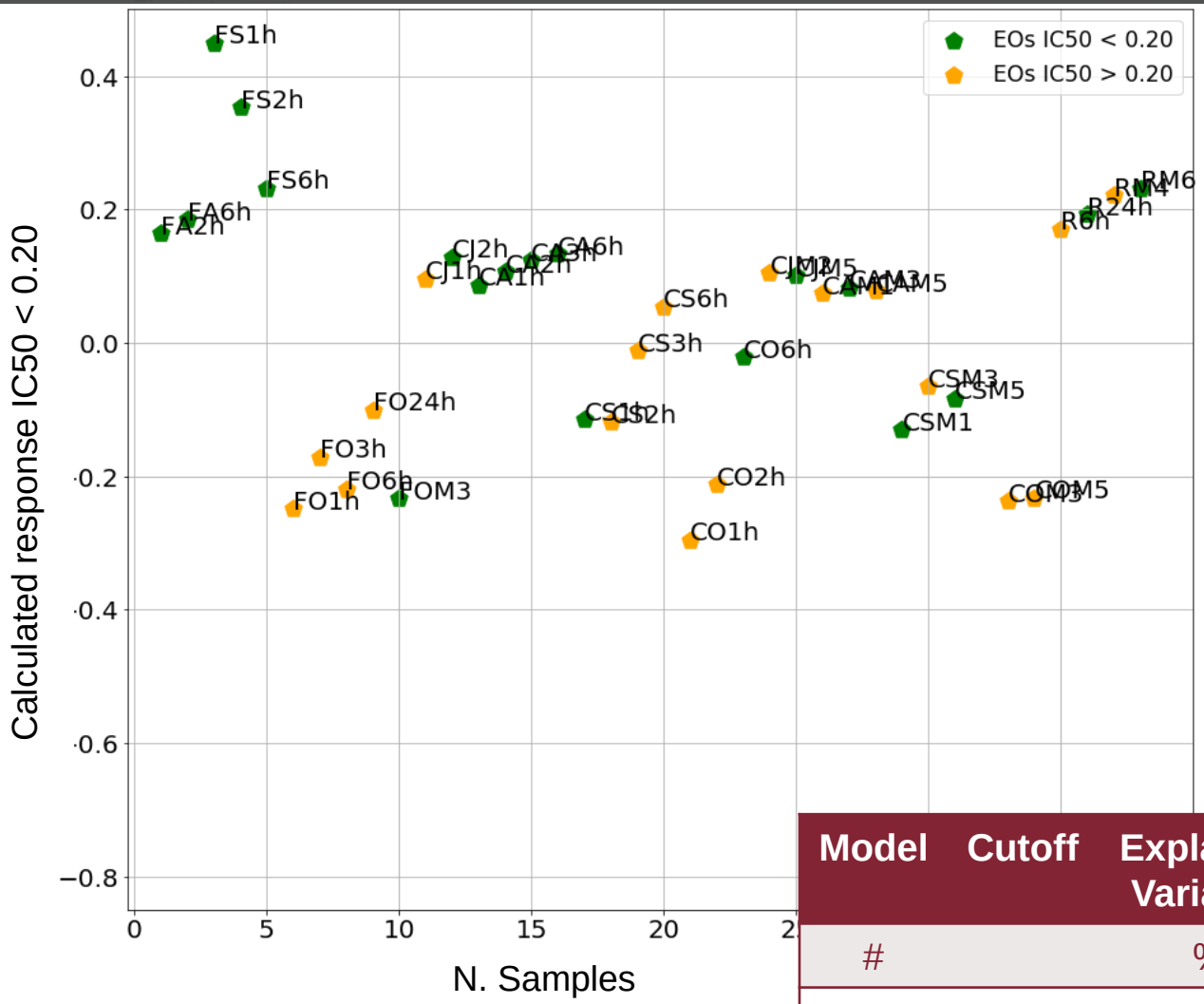


Biological Data

by www.RCMD.it

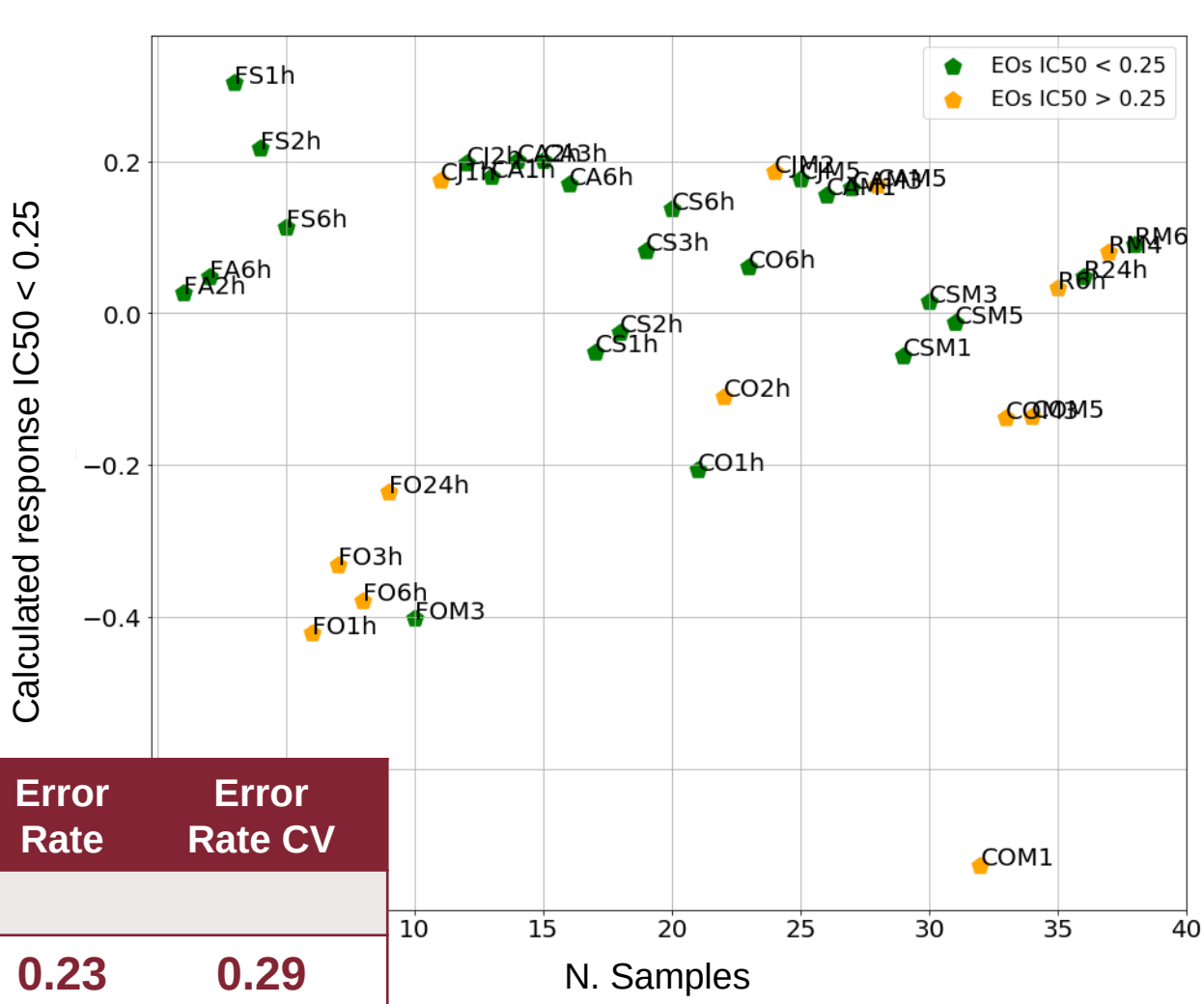
Sample	IC50 (mg/ml)	CC50 (mg/ml)	Sample	IC50 (mg/ml)	CC50 (mg/ml)	Sample	IC50 (mg/ml)	CC50 (mg/ml)
CS1	0,14	2,5	CJM1	0,328	2,50	***FO3***	***0,6465***	***1,02***
CS2	0,2063	2,9	CJM5	0,1353	3,14	***FO6***	***0,54025***	***1,694***
CS3	0,231	2,68	CA1	0,1617	2,84	***FO24***	***0,1765***	***0,355***
CS6	0,213	2,66	CA2	0,1517	4,71	***FOM3***	***0,582***	***1,7***
CSM1	0,174	2,00	CA3	0,175	2,50	***RJ6***	***0,431***	***0,405***
CMS3	0,236	2,12	***CA6***	***0,142***	***1,14***	***RJ24***	***0,18075***	***0,36***
CSM5	0,1225	1,93	CAM1	0,202	2,65	***RJM4***	***0,295***	***0,9***
CO1	0,21	2,50	CAM3	0,1537	3,00	***RJM6***	***0,2***	***0,33***
CO2	0,332	4,60	***CAM5***	***0,4395***	***1,10***	***MEDIAN***	***0,20***	***2,06***
CO6	0,174	2,22	***FA2***	***0,1495***	***1,45***			
COM1	0,282	2,76	***FA6***	***0,1605***	***1,815***			
COM3	0,3125	1,99	***FS1***	***0,1945***	***1,005***			
COM5	0,3645	2,46	***FS2***	***0,192***	***0,31***			
CJ1	0,2725	2,15	***FS6***	***0,192***	***1,513***			
CJ2	0,135	4,67	***FO1***	***0,7235***	***1,513***			

PLSDA on Biological Data

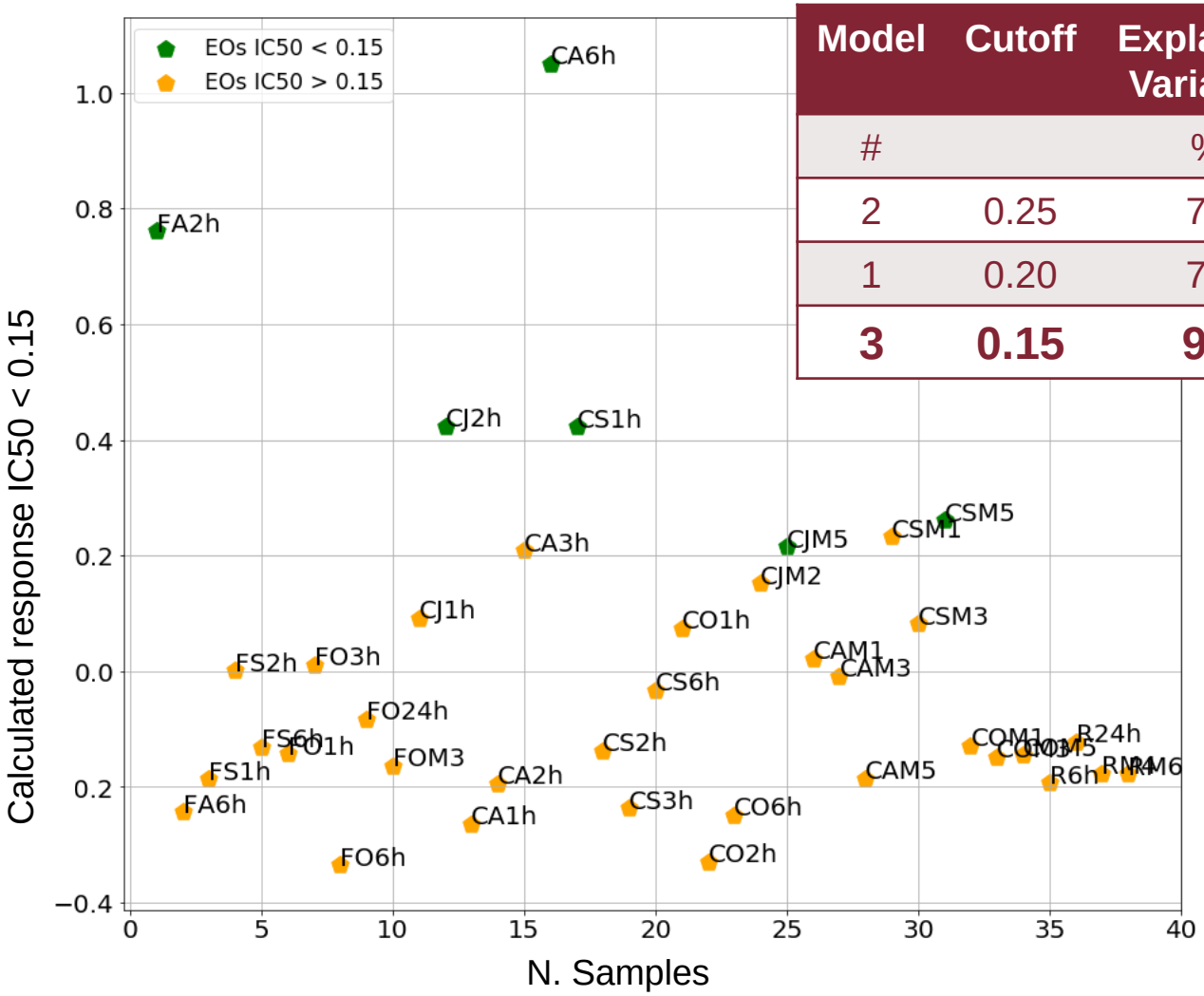


Model	Cutoff	Explained Variance	PC	Error Rate	Error Rate CV
#		%	#		
1	0.20	75	2	0.32	0.39

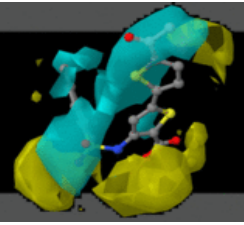
PLSDA on Biological Data



Model	Cutoff	Explained Variance	PC	Error Rate	Error Rate CV
#		%	#		
2	**0.25**	**75**	**2**	**0.23**	**0.29**
1	0.20	75	2	0.32	0.39



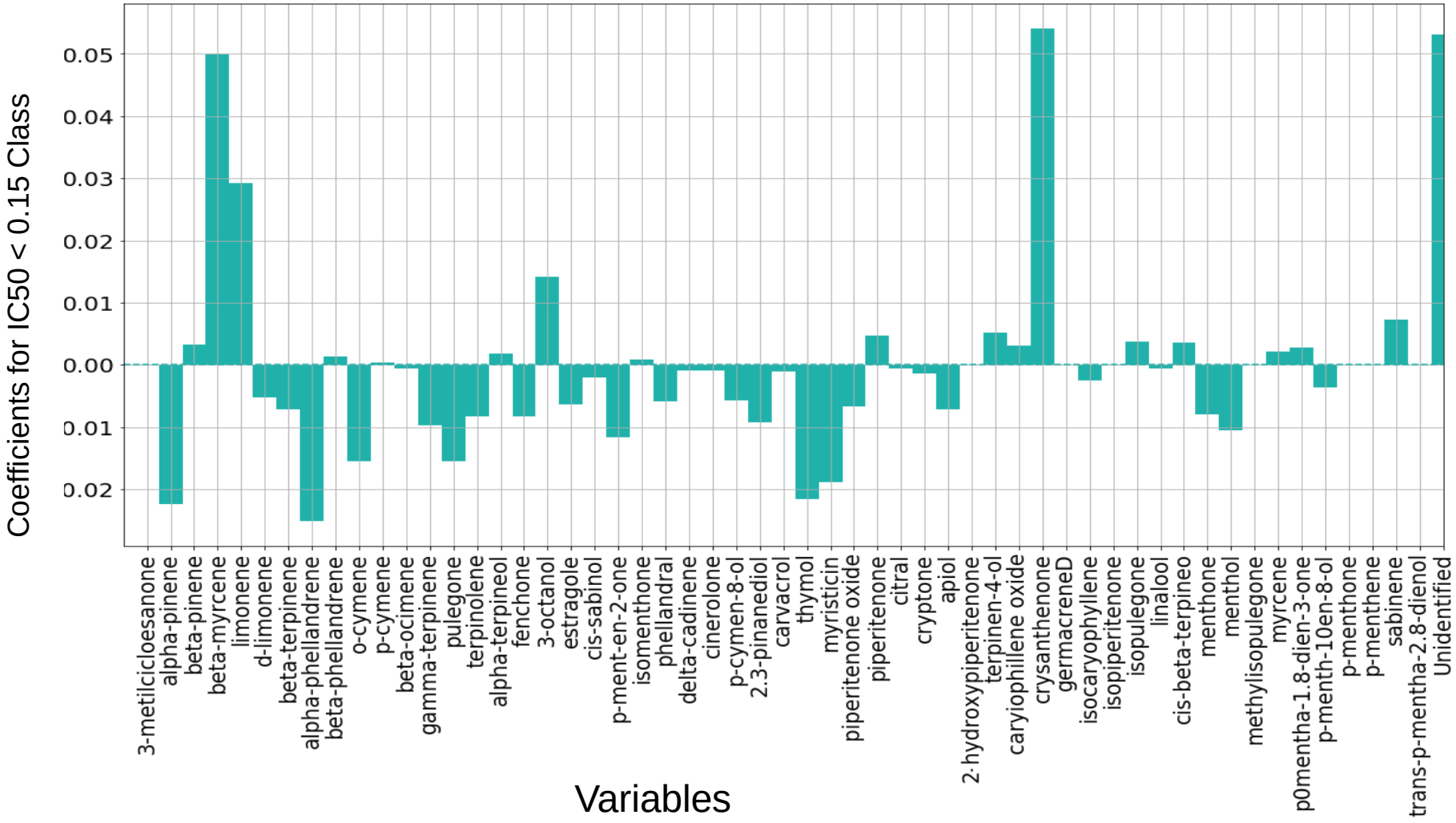
Model	Cutoff	Explained Variance	PC	Error Rate	Error Rate CV
#		%	#		
2	0.25	75	2	0.23	0.29
1	0.20	75	2	0.32	0.39
3	**0.15**	**98**	**7**	**0.03**	**0.15**

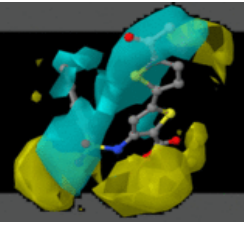


Variable Coefficients

-Biological Model

by www.RCMD.it

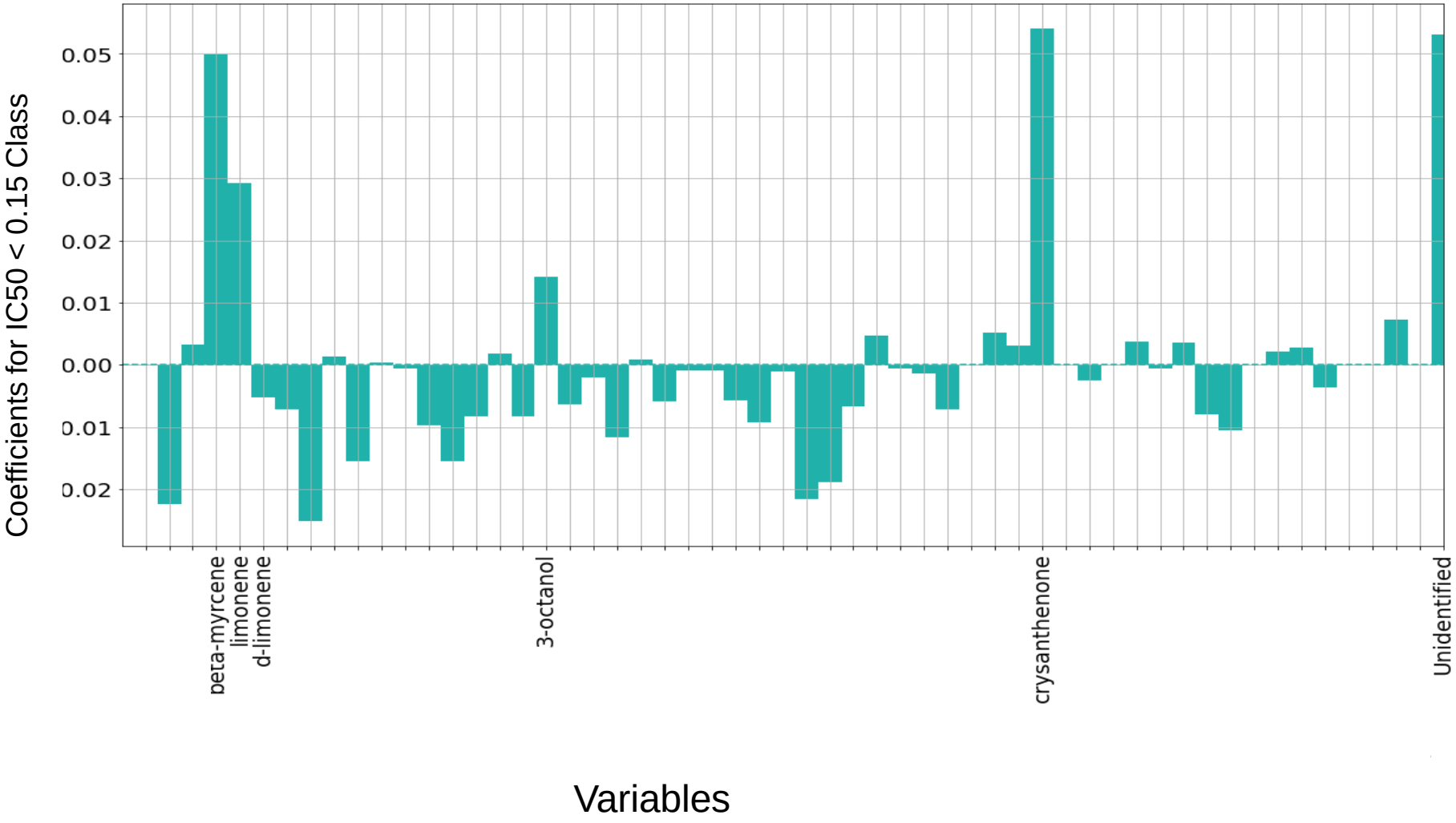




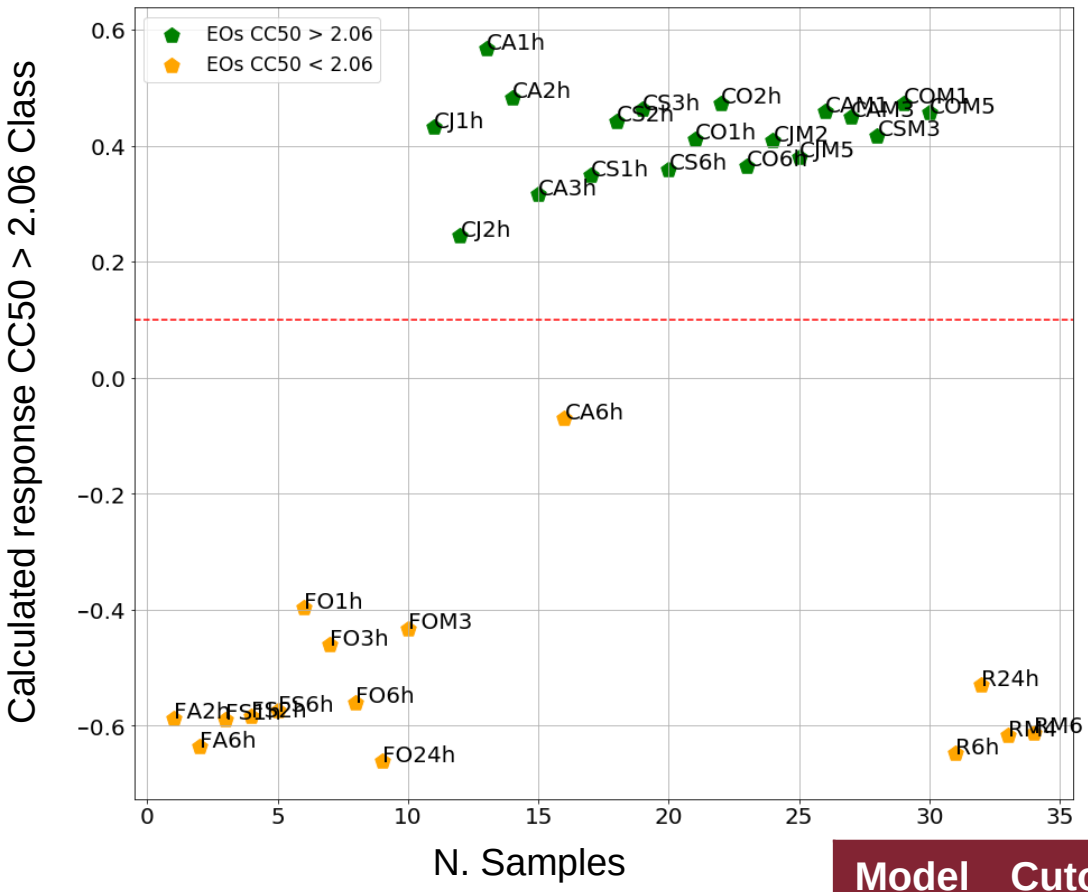
Variable Coefficients

-Biological Model

by www.RCMD.it



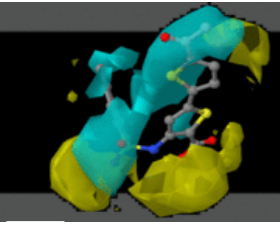
PLSDA on Toxicological Data



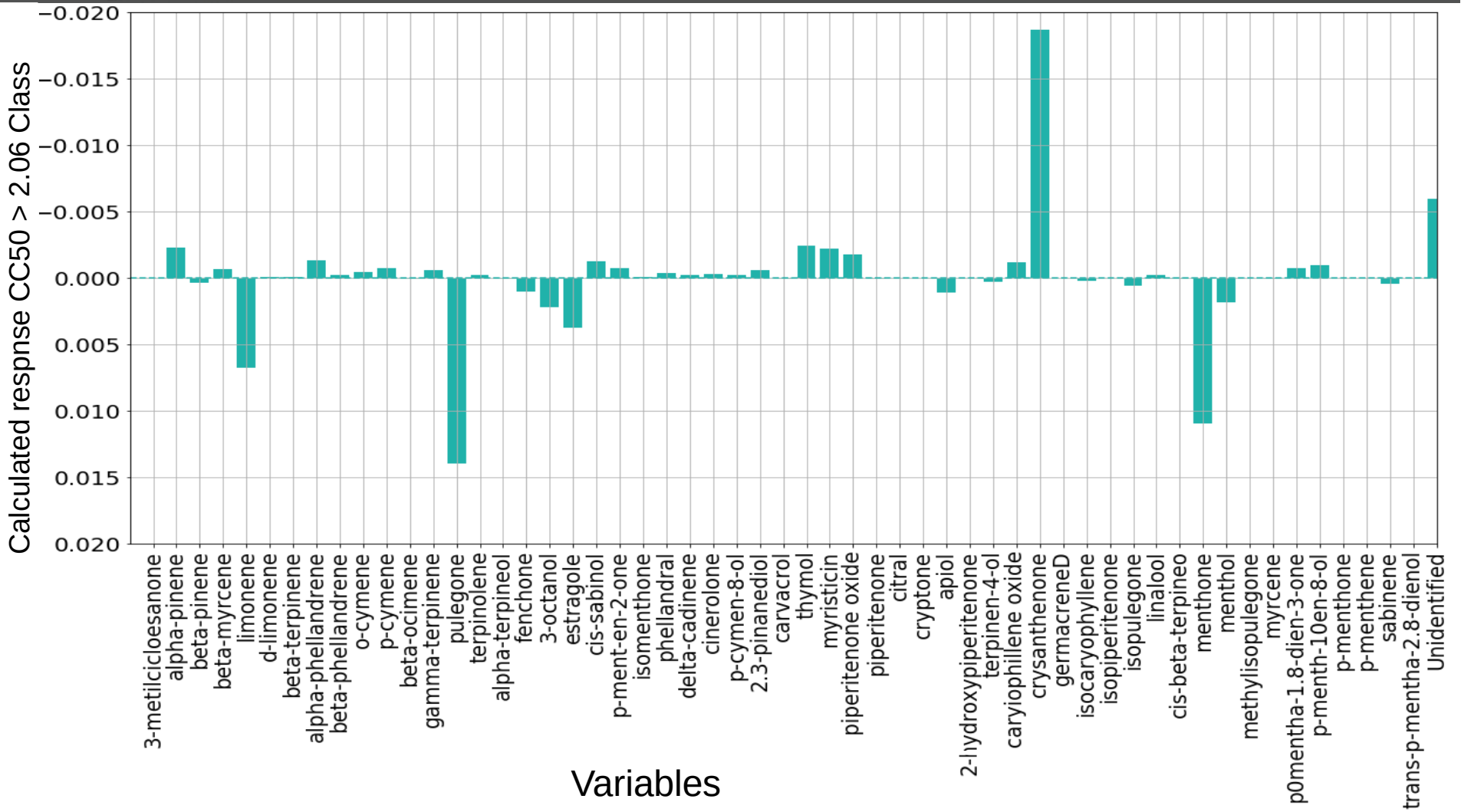
Model	Cutoff	Explained Variance	PC	Error Rate	Error Rate CV
#		%	#		
1	2	78	3	0	0.03

Variable Coefficients

-Toxicological Model

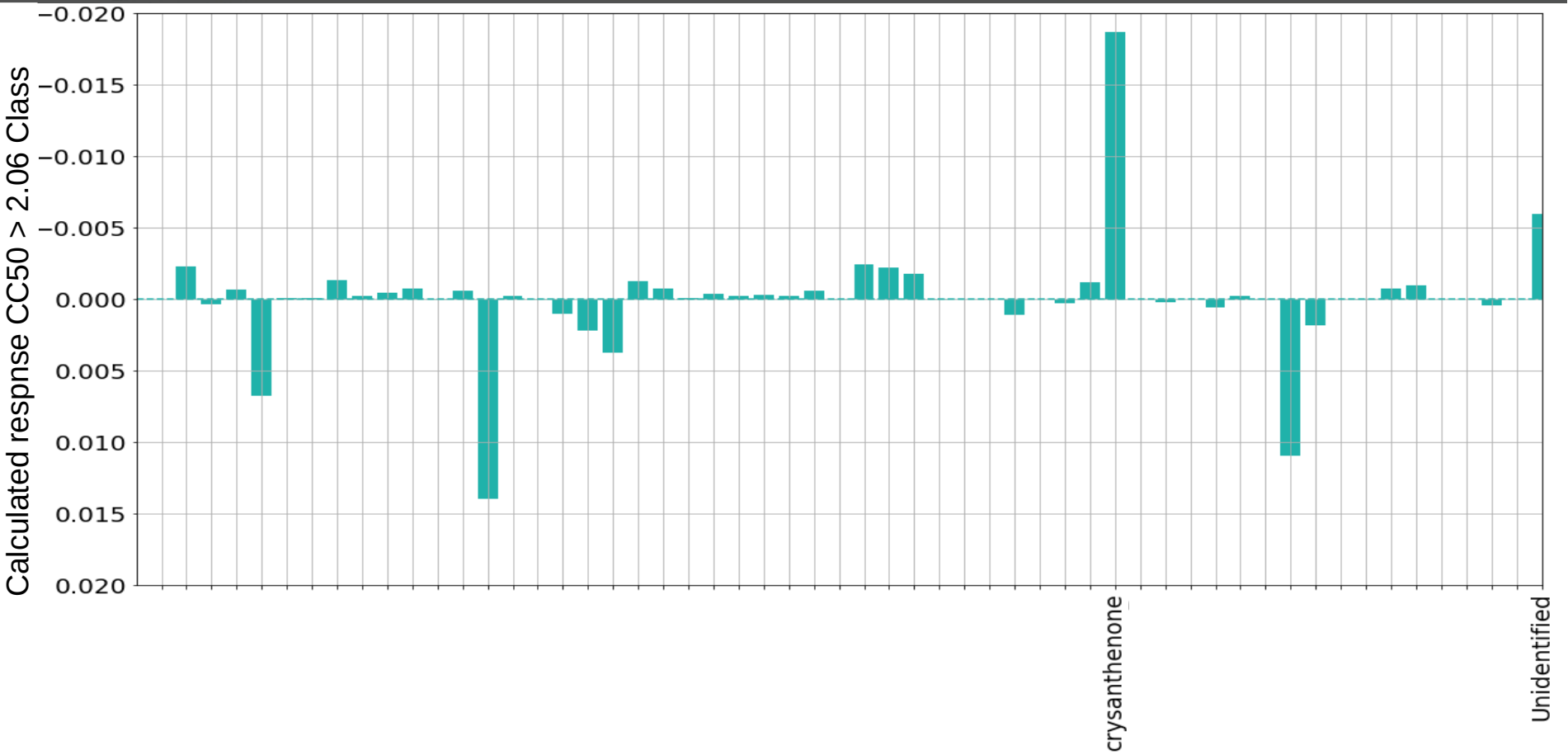


by www.RCMD.it

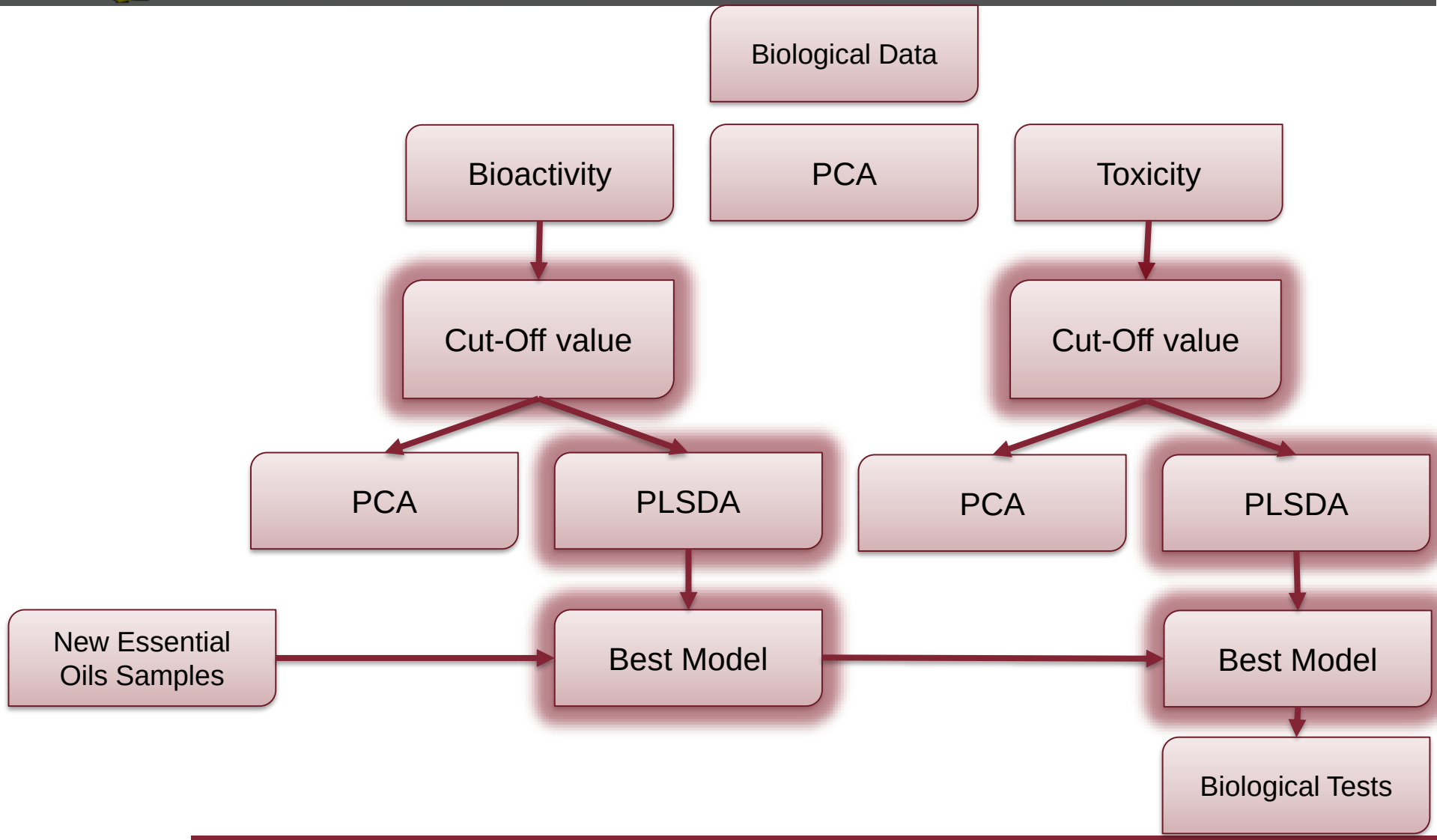


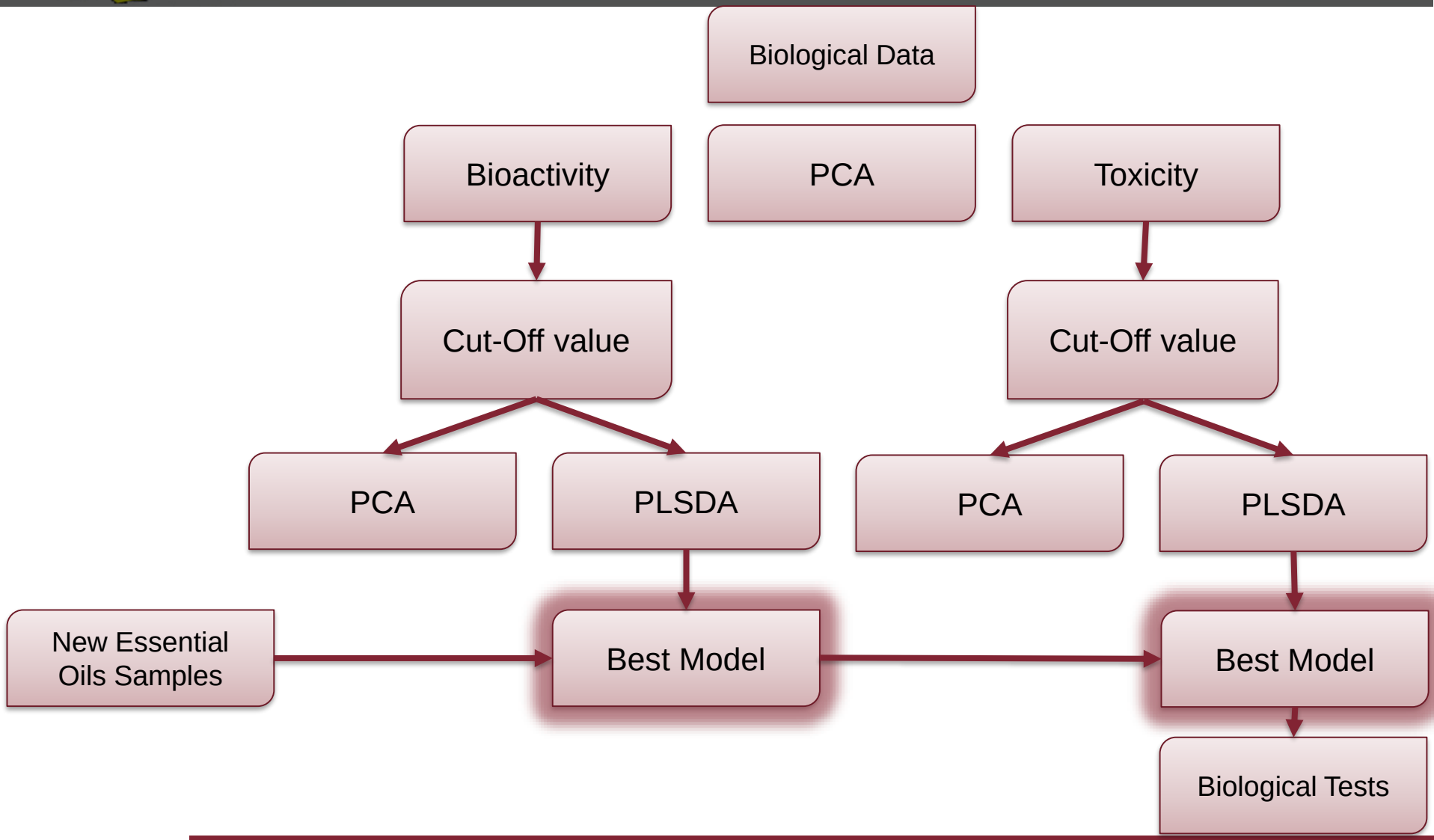
Variable Coefficients

-Toxicological Model



Overflow


Biological Data
Bioactivity
PCA
Toxicity
Cut-Off value
Cut-Off value
PCA
PLSDA
PCA
PLSDA
New Essential Oils Samples
Best Model
Best Model
Biological Tests


Biological Data
Bioactivity
PCA
Toxicity
Cut-Off value
Cut-Off value
PCA
PLSDA
PCA
PLSDA
New Essential Oils Samples
Best Model
Best Model
Biological Tests

Prediction Samples Set

by www.RCMD.it

FA1h	FAM1	FSM5	CA12h	CSM2	RJM1
FA3h	FAM2	FOM1	CA24h	CSM4	RJM2
FA12h	FAM3	FOM2	CS12h	COM2	RJM3
FA24h	FAM4	FOM4	CS24h	COM4	
FS3h	FAM5	FOM5	CJM1	RJ1h	
FS12h	FSM1	CJ3h	CJM3	RJ2h	
FS24h	FSM2	CJ6h	CJM4	RJ3h	
FO2h	FSM3	CJ12h	CAM2	RJ12h	
FO12h	FSM4	CJ24h	CAM4	RJ30h	

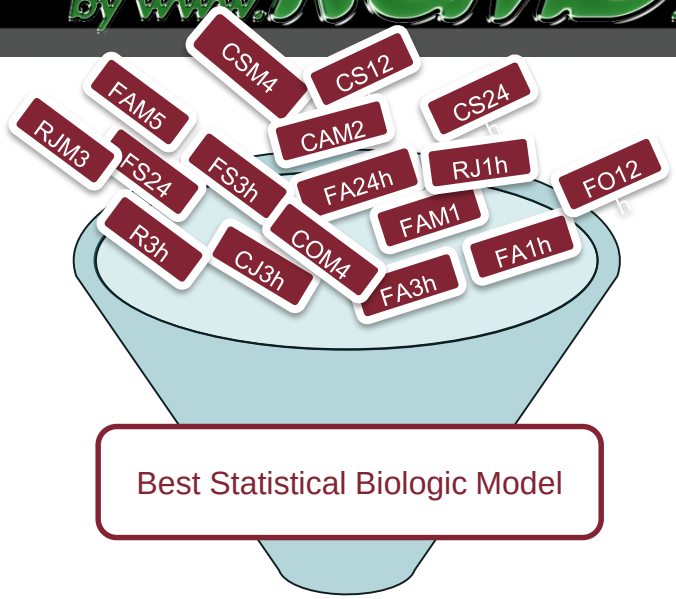
Calamintha Nepeta

Foeniclum Vulgare

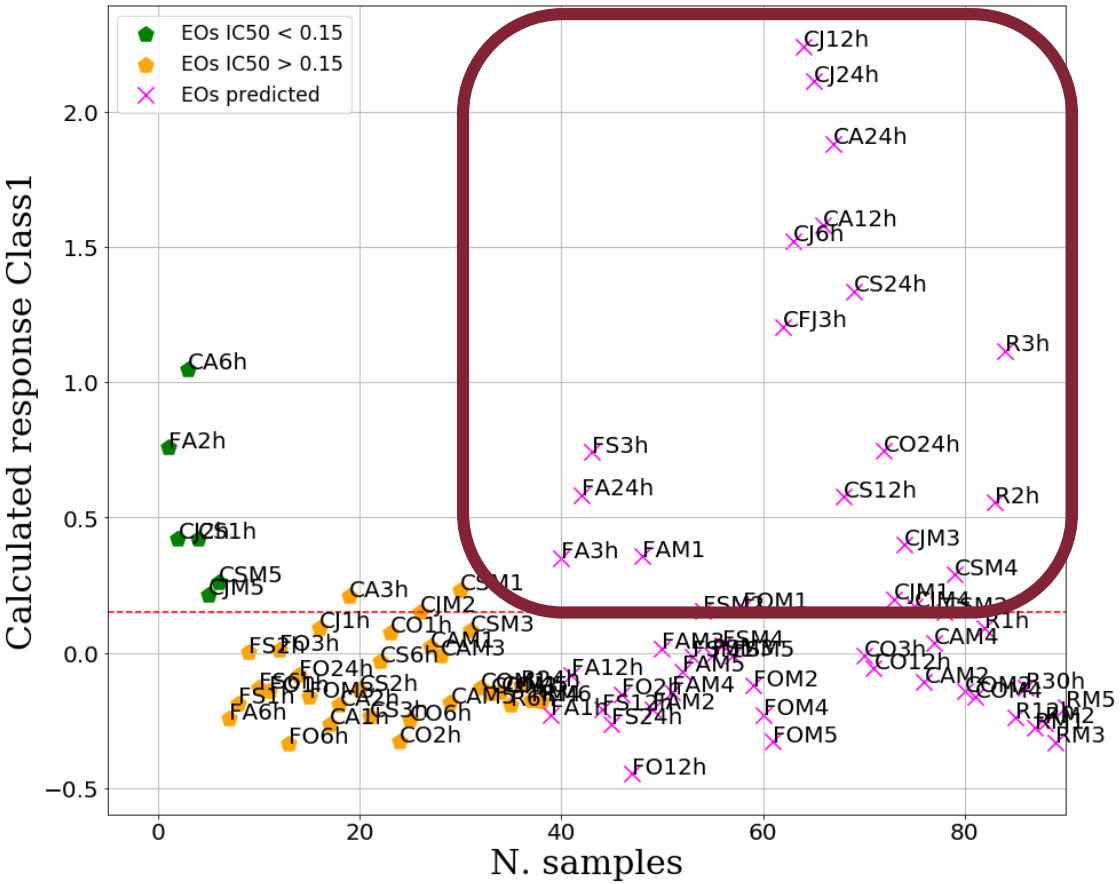
Ridolfia segetum

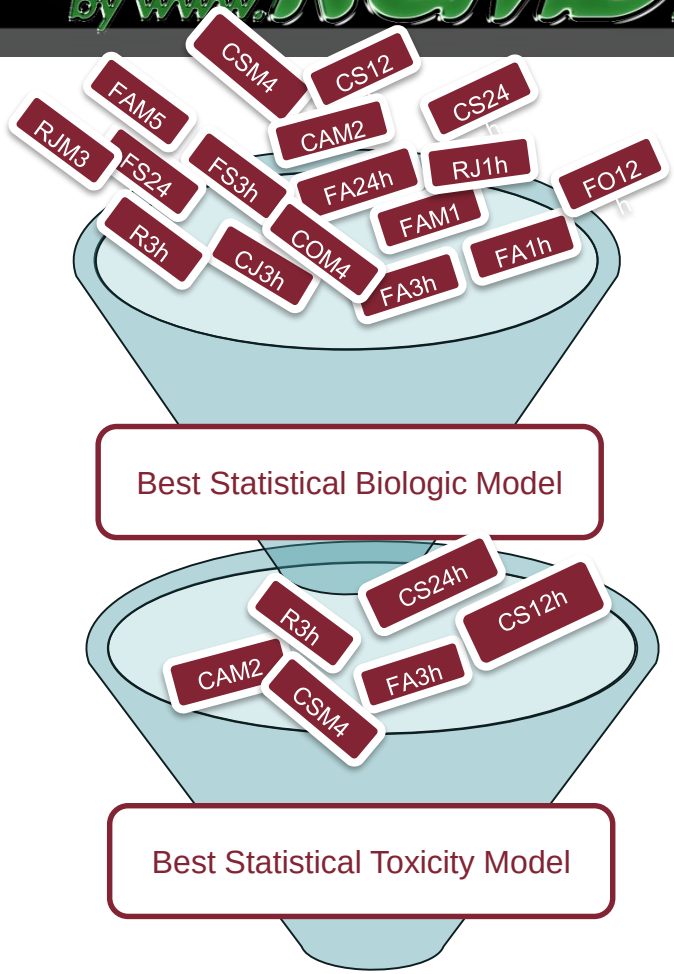
Biological Activity Prediction

by www.RCMD.it

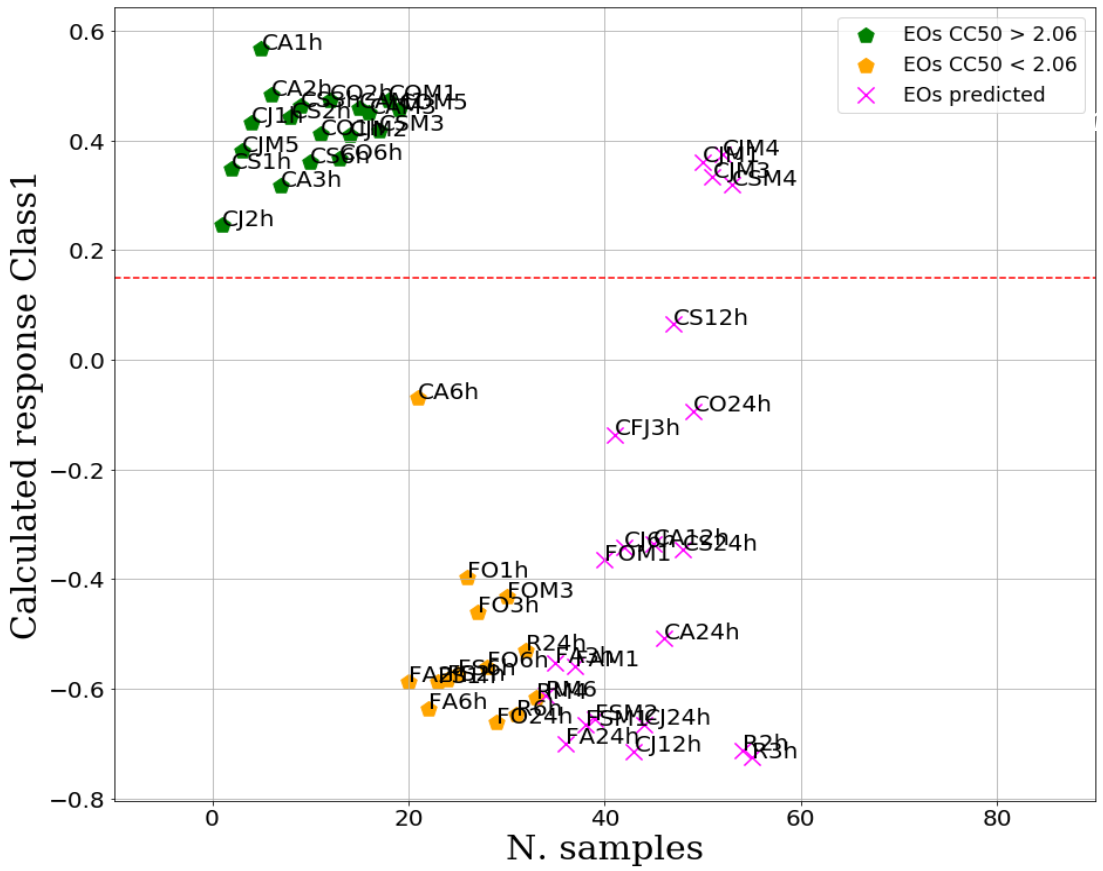


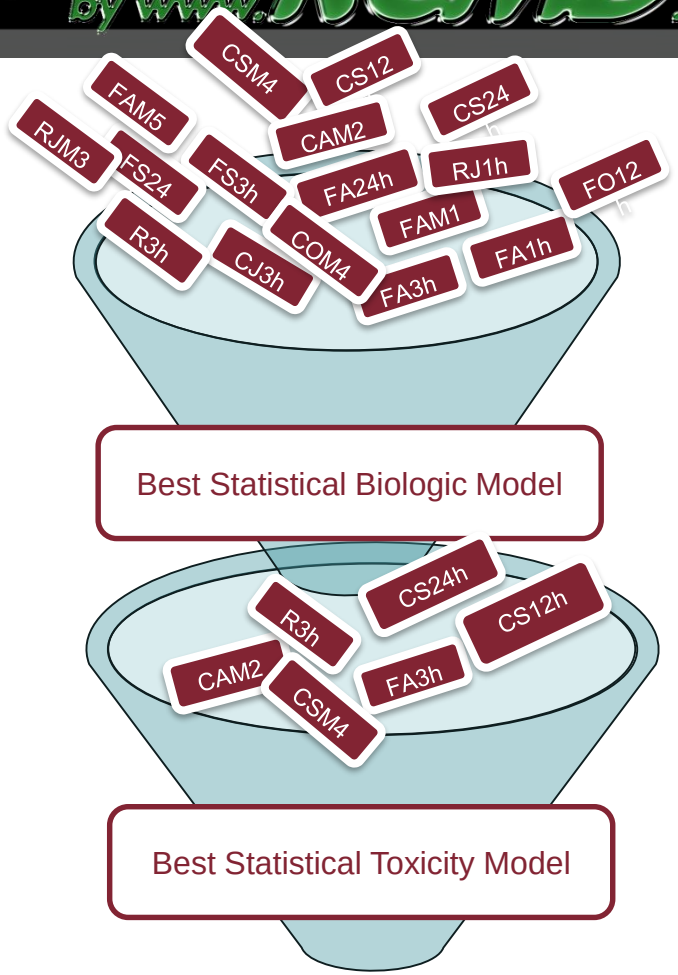
Biological Activity Prediction

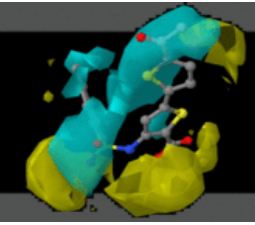




Toxicity Prediction

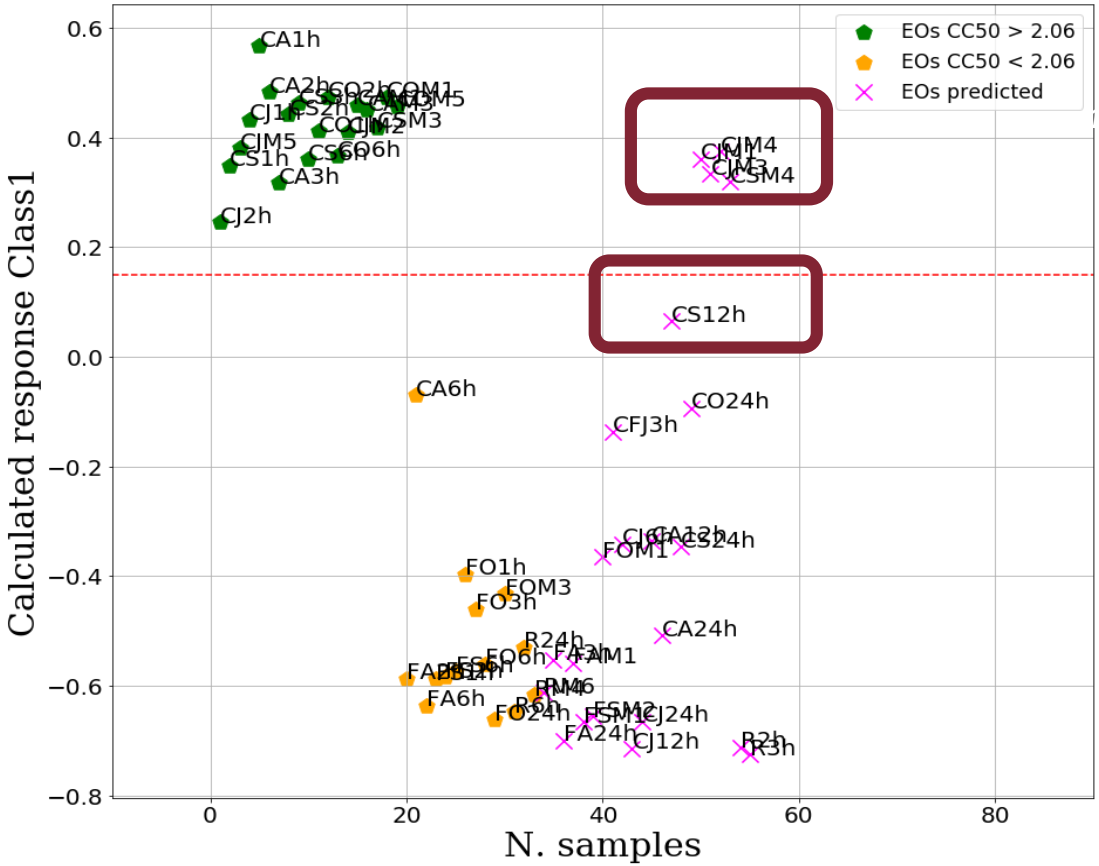


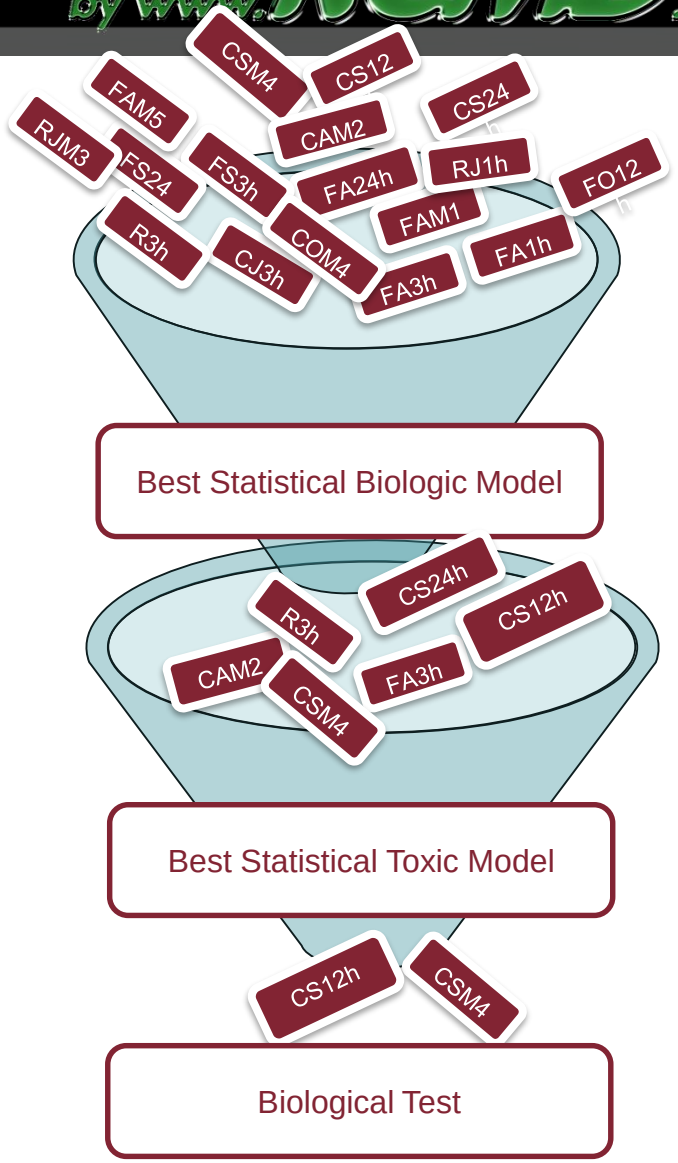


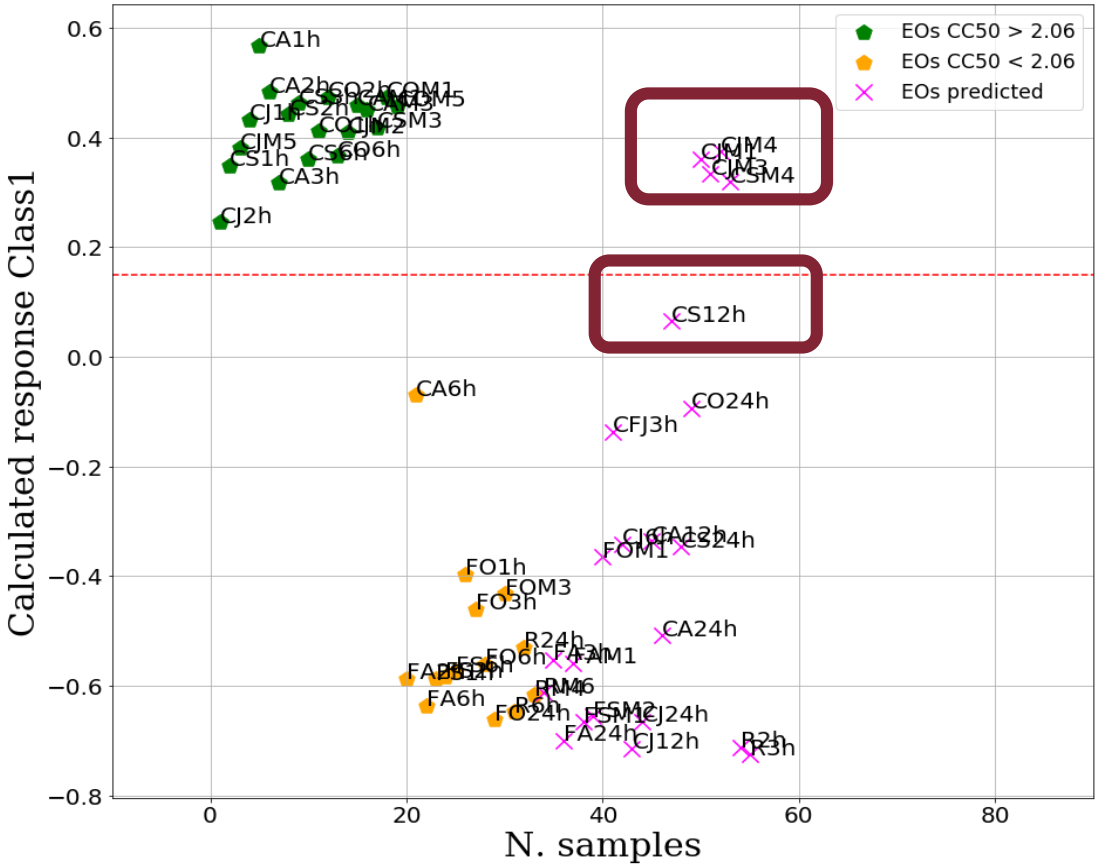


Toxicity Prediction

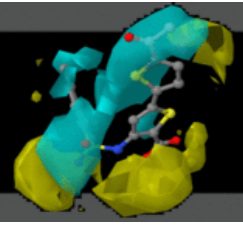
by www.RCMD.it







Sample	IC50 (mg/ml)	CC50 (mg/ml)
CJM1	0.063	>3.0
CJM3	0.116	>3.0
CJM4	0.124	>3.0
CS12	0.143	2.50
CSM4	0.460	0.52



Conclusion

- It was showed as Machine Learning techniques are usefull to study and analyse also chemical complex mixtures;
- Using Machine Learning techniques was developed a scientific protocol;
- Applying this presented protocol were identified a Essential Oils with promising biological activity against Herpes Simplex Virus;

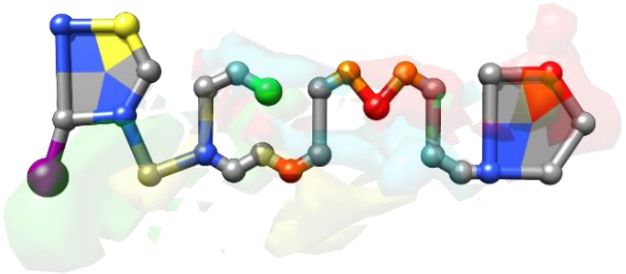
Rational Design of Potential Novel Pharmaceutical Compounds Through Computational Approaches and Machine Learning Techniques.

SAPIENZA
UNIVERSITÀ DI ROMA

PhD School of Pharmaceutical Sciences XXX Cycle
Department of Chemistry and Drug Technologies
Faculty of Pharmacy and Medicine
"Sapienza" University of Rome

Manuela Sabatino

Acknowledgements



Prof. Rino Ragno
Alexandros Patsilinakos (PhD student)
Lorenzo Antonini (PhD student)
Valentina Generoso
Nikola Dino Capocchiano
Carta Ylenia
Vellucci Sonia
Zonetti Pamela

Prof. Stefano Manfredini
dr.ssa Alessandra Oliva
dr. Mijat Bozovic

Prof. Mai
dr. Sergio Valente
dr. Dante Rotili
dr.ssa Giulia Stazi
Annalisa Romnelli (PhD student)

Prof.ssa Barbara Arimondo

Prof.ssa Rosa Diliddo
dr. Sandra Shrenk
dr. Carola Cenzi

Prof.ssa Teresa Palamara
dr.ssa Giovanna De Chiara
dr.ssa Elena Marcocci
dr. Marco Fabiani