

Applicazione di Metodi Machine Learning ad Inibitori SIRT2

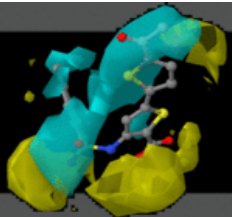


SAPIENZA
UNIVERSITÀ DI ROMA

**Facoltà di Farmacia e Medicina
Corso di Laurea in Farmacia
Tesi Sperimentale in Chimica Farmaceutica
a.a. 2016/2017**

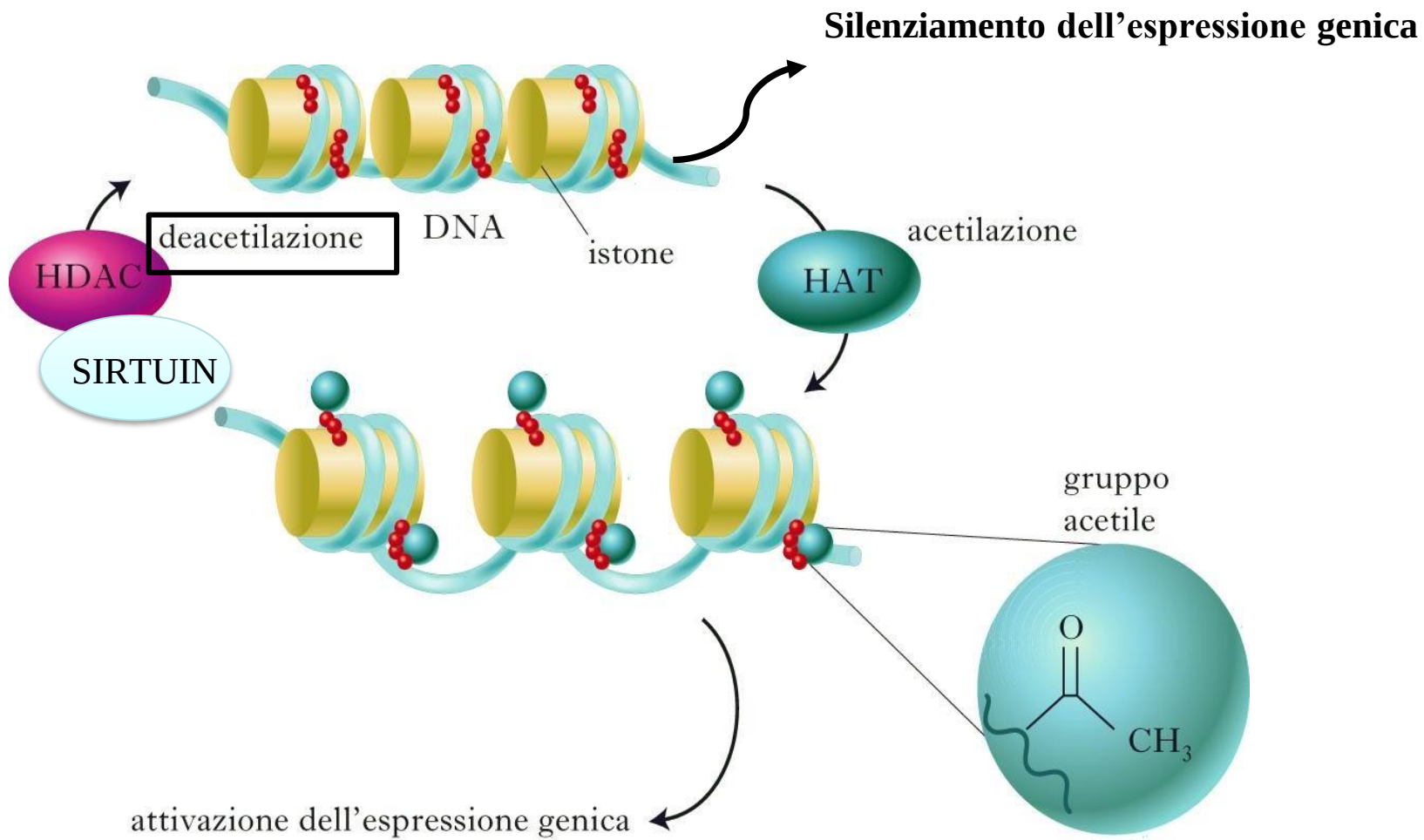
**Mandana Khademi
Matricola: 1228387**

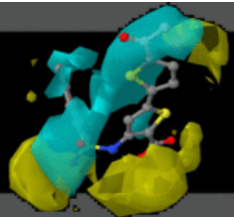
Relatore: Prof. Rino Ragno



Sirtuine e l'Epigenetica

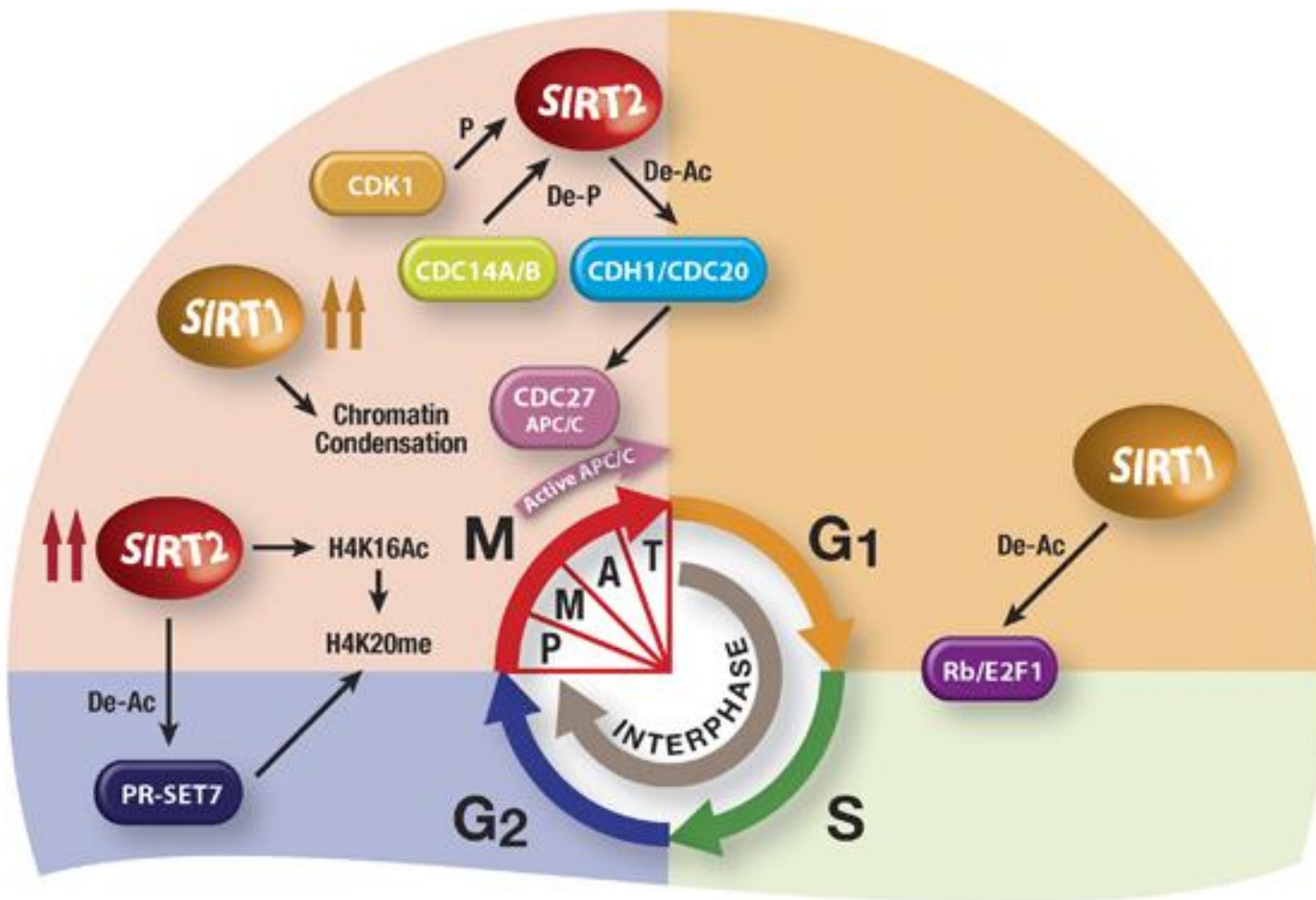
By www.RCMD.it

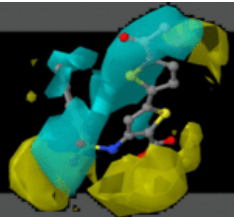




Ruoli di SIRT2

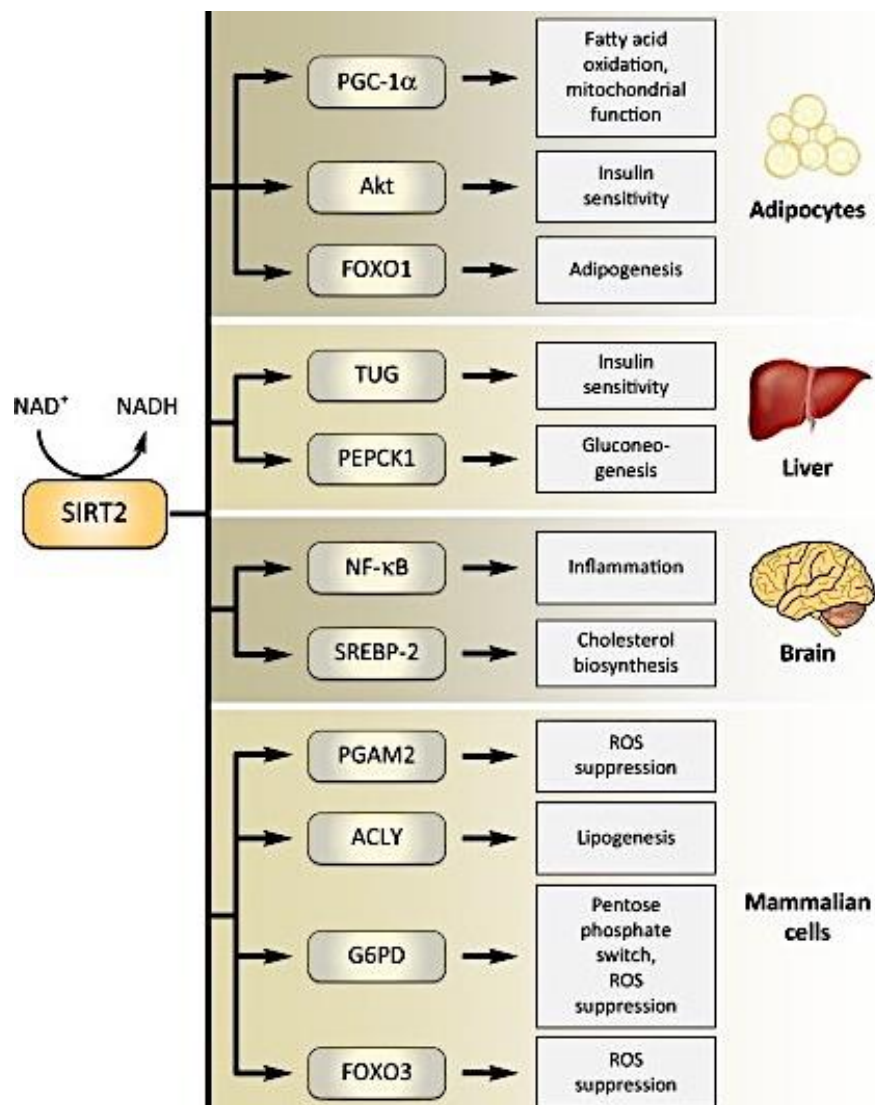
by www.RCMD.it

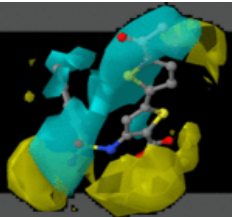




Ruoli di SIRT2

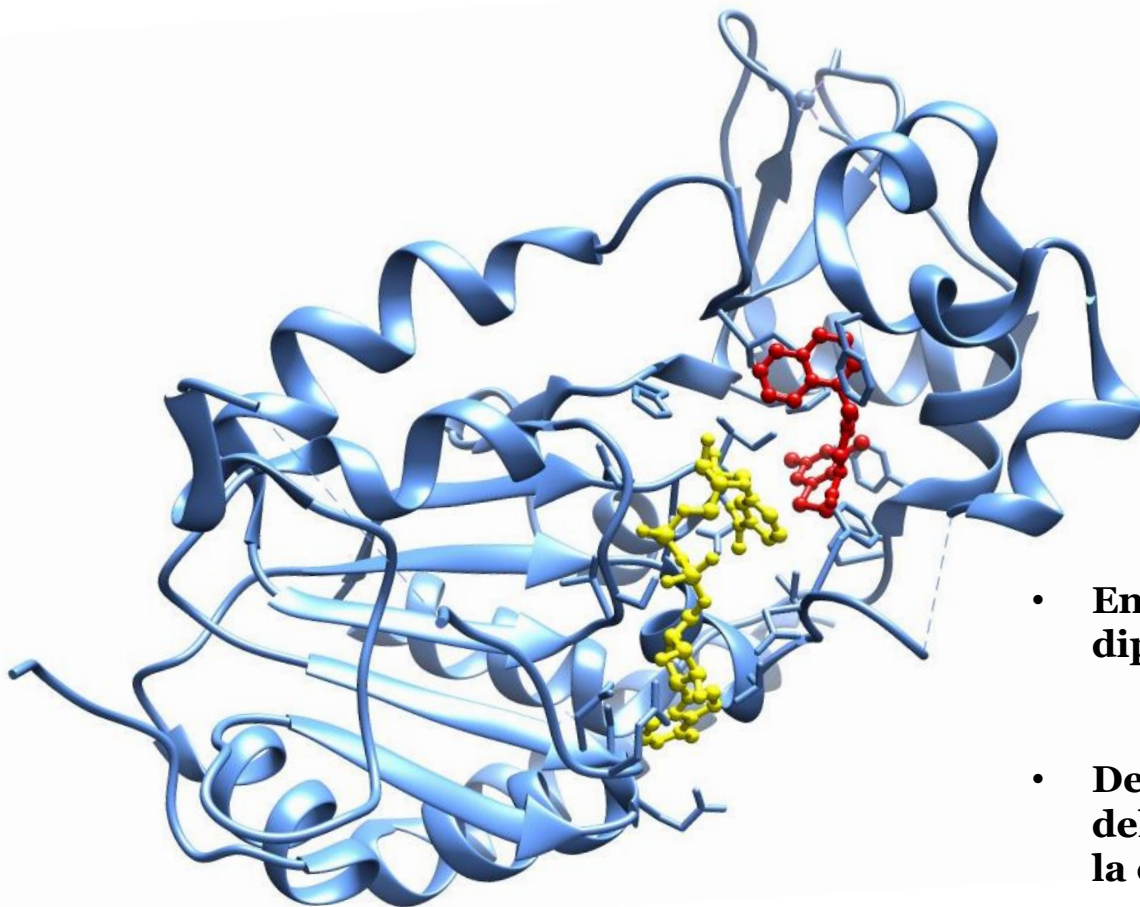
- L'Espressione nei neuroni invecchiati
- Protezione contro lo stress ossidativo
- interazione con p65, FOXO1, FOXO3 e HIF-1 α
- Sensibilità all'insulina
- Autofagia
- Attivazione del fattore NF-kB e MAP-Kinasi





SIRT2

by www.RCMD.it



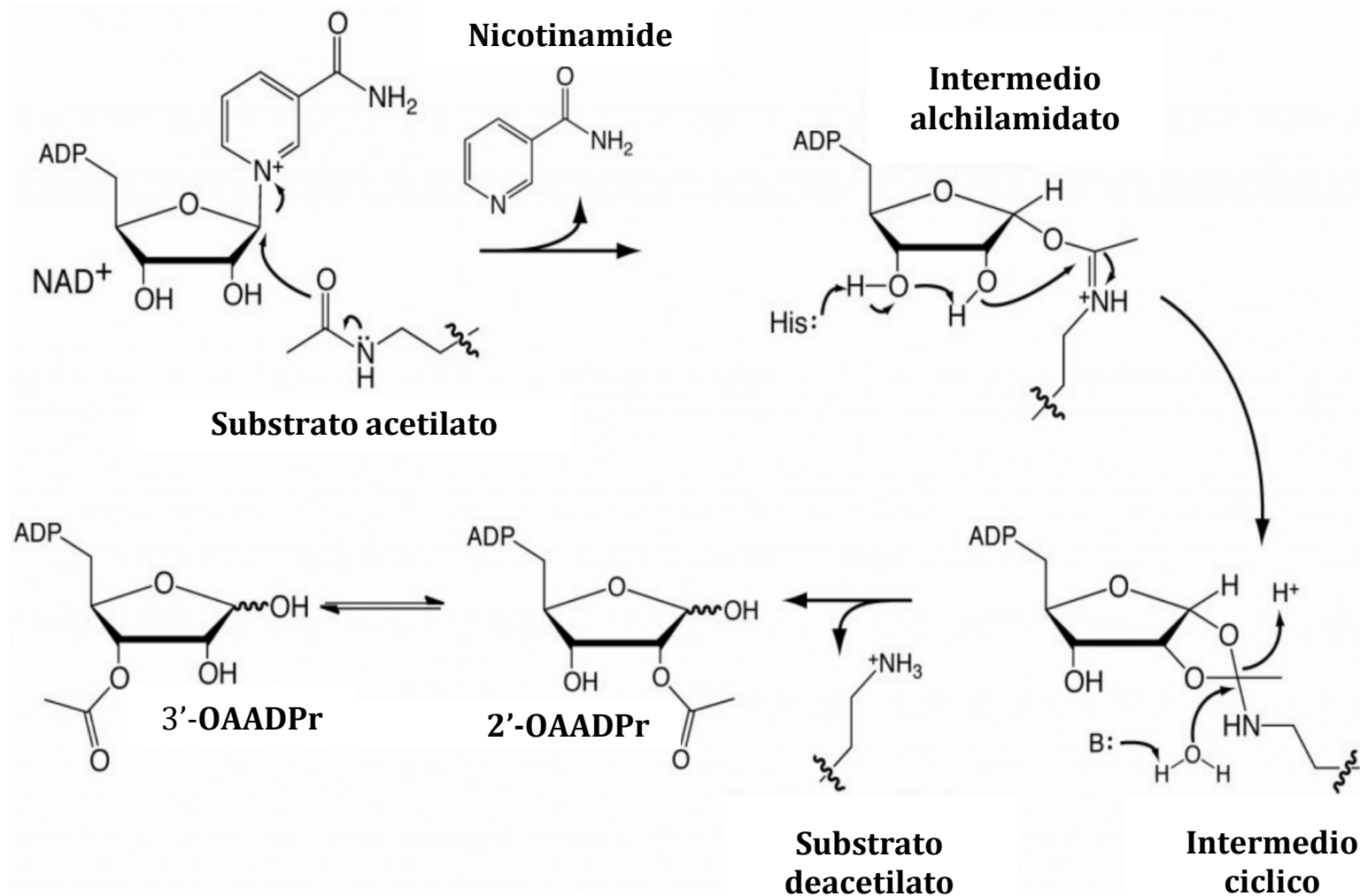
- Enzima lisina- deacetilasi NAD⁺ dipendente
- Deacetilazione delle porzioni ϵ -NH₂ della lisina degli substrati peptidici e la coda N-terminale Istonica

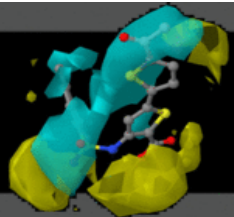
PDB Id : 4RMG in complex with SirReal2 and NAD⁺



Meccanismo Catalitico

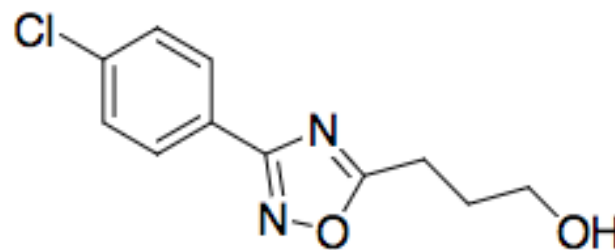
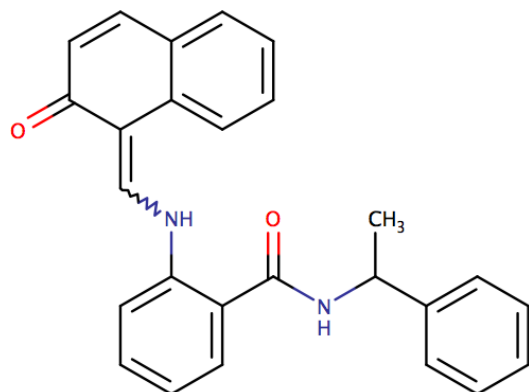
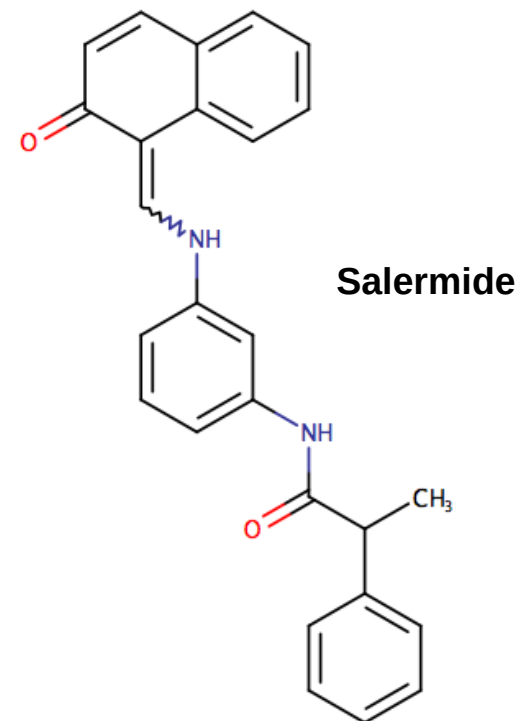
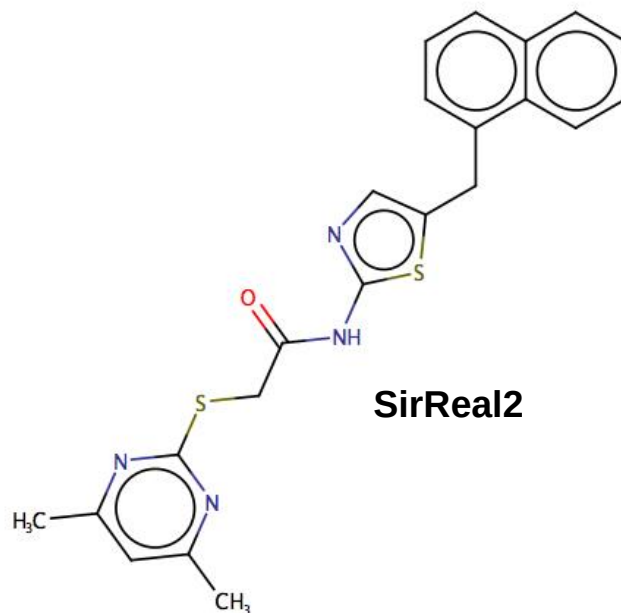
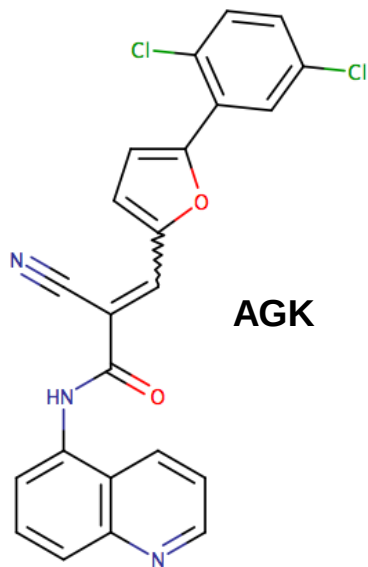
by www.RCMD.it

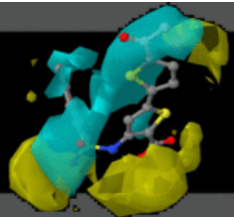




Inibitori SIRT2

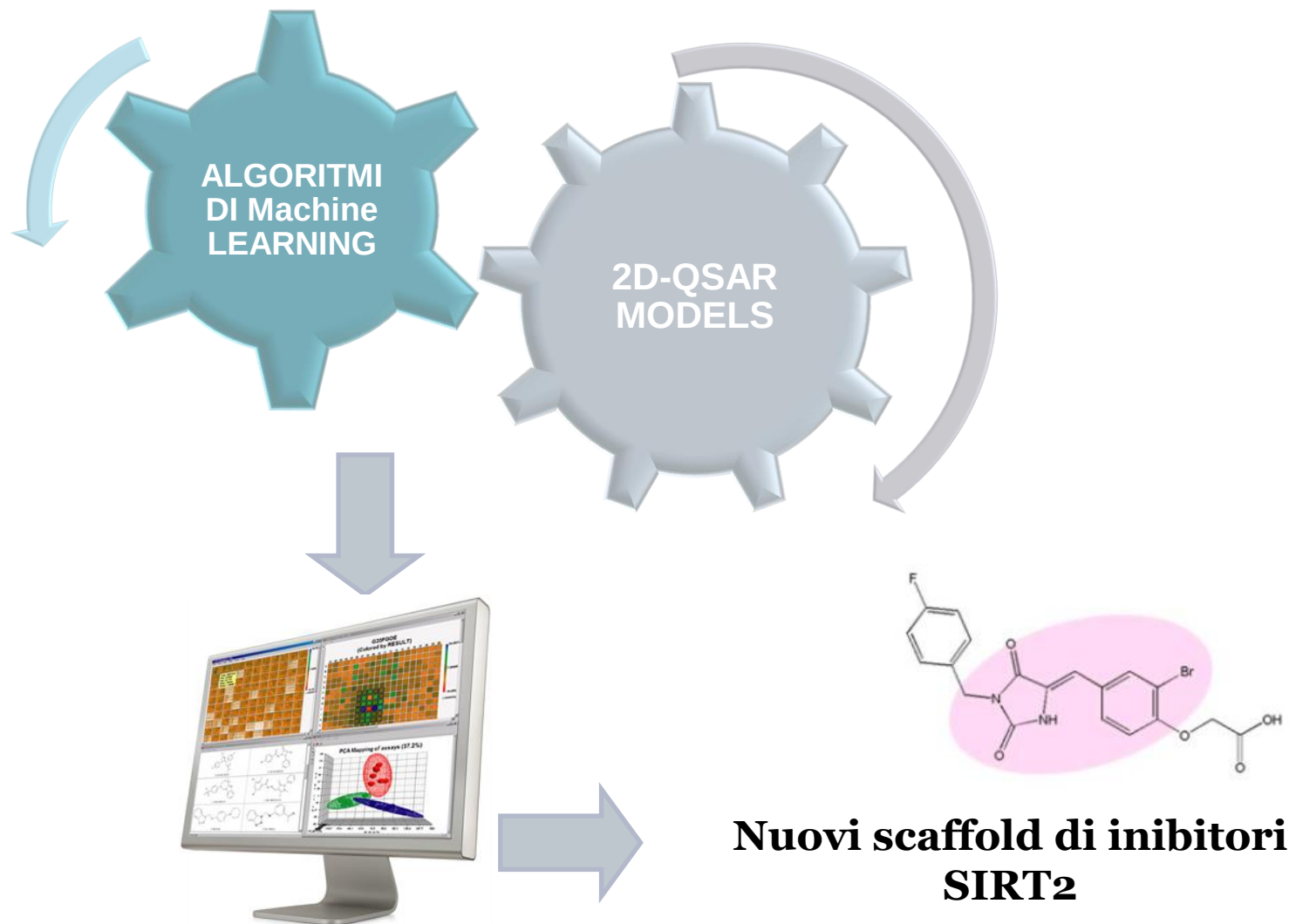
by www.rcmd.it

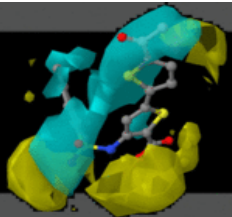




Scopo della tesi

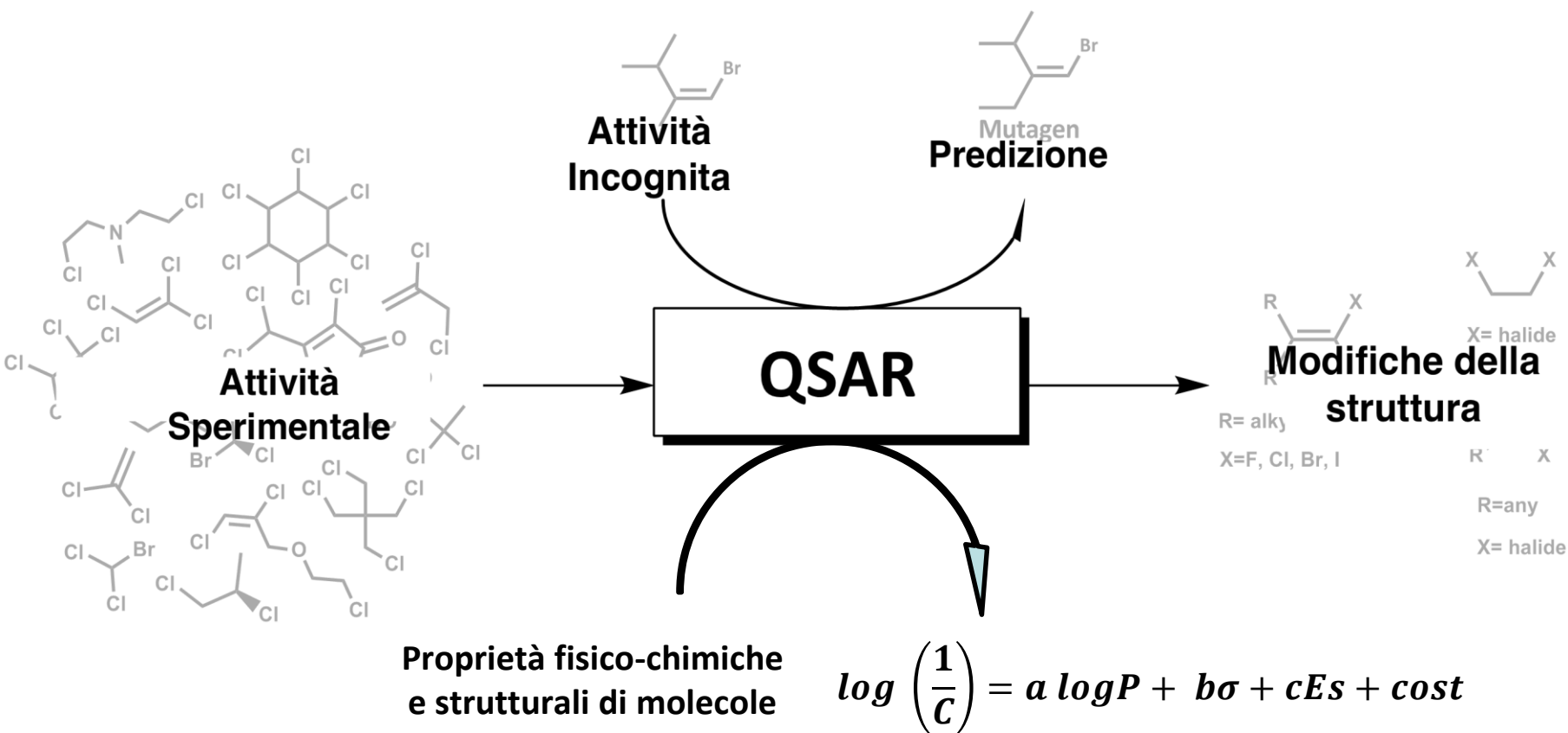
by www.RCMD.it





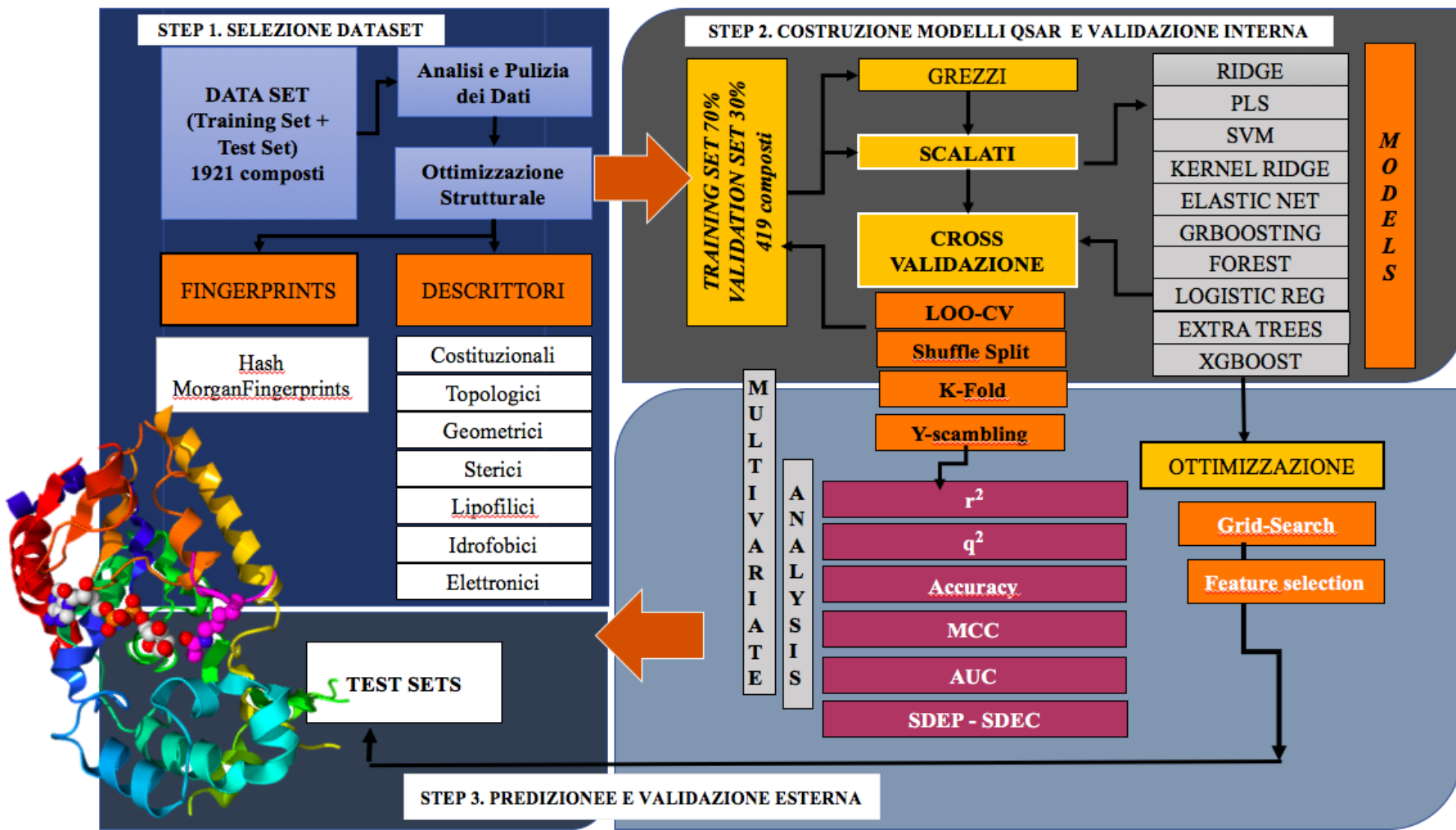
Quantitative Structure–Activity Relationships

by **www.RCMD.it**



Procedura Sperimentale

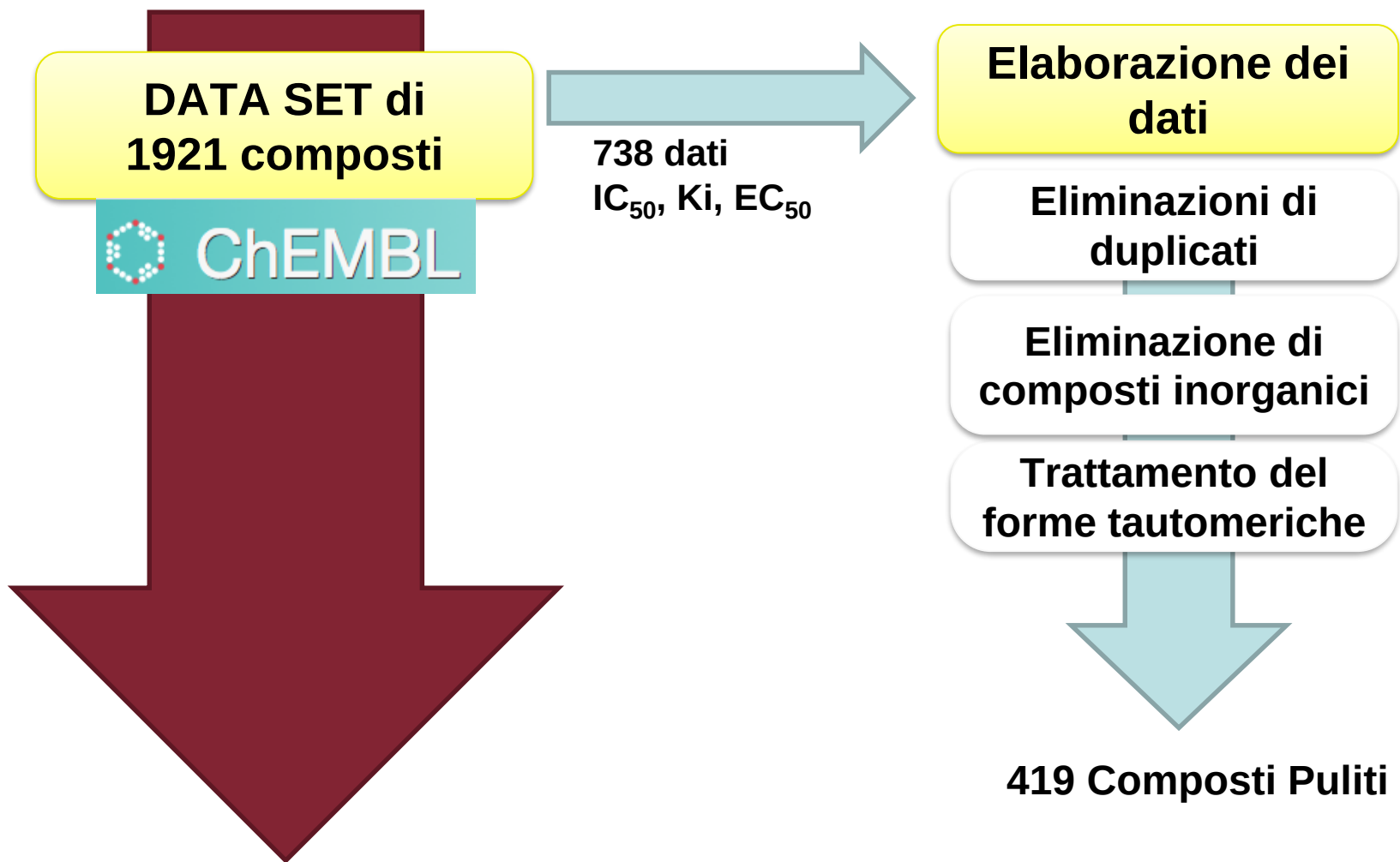
by **RCMD**.it





Step1: Pre-processing

by www.RCMD.it





Generazione dei descrittori

by **www.RCMD.it**

DATA SET di
composti

Generazione dei
descrittori

Descrittori molecolari

predittori

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & \dots & x_{1p} \\ x_{21} & x_{22} & & & \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ x_{nr1} & & & & x_{nrp} \end{bmatrix}$$

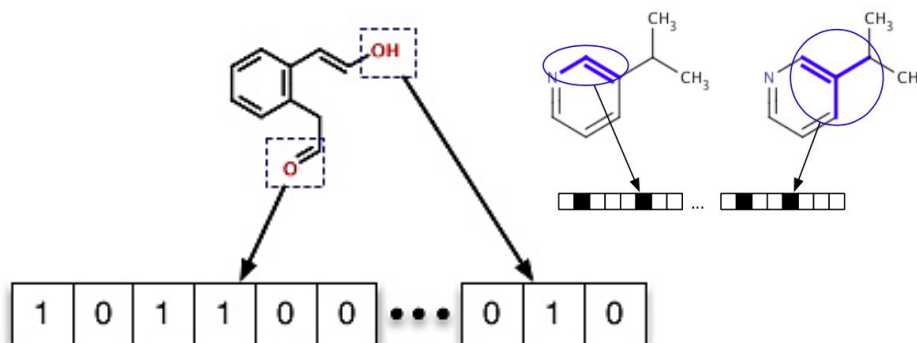
Variabili indipendenti

una o più risposte
da modellare

$$Y = \begin{bmatrix} y_{1r} & y_{12} & \dots & \dots & y_{1R} \\ y_{2r} & y_{22} & & & \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ y_{nr} & & & & y_{nR} \end{bmatrix}$$

Variabili dipendenti

Morgan Hash Fingerprints 2048 bits





Preparazione del TrainingSet

by www.RCMD.it

DATA SET di
composti

Generazione dei
descrittori

Scalatura dei Descrittori
rdkit

$$x' = \frac{x - \bar{x}}{\sigma}$$



Costruzione del Modello

by **www.RCMD.it**

**DATA SET di
composti**

**Generazione dei
descrittori**

**Divisione dati in Training,
Test e Validation Set**

**Techniche Machine
Learning**

Ottimizzazione

Grid Search

**Ricerca di Migliori
Descrittori
'PEOE_VSA14' e 'Balaban J'**

```
model1 = svm.SVC(kernel='poly',  
C=1, degree = 4, tol= 0.01,  
gamma=0.1, probability=True )
```

Classificazione

Regressione

```
model2=SGDRegressor(loss='squared_  
loss', penalty='l2', alpha=0.0001, l1_rati  
o=0.15, fit_intercept=True)
```



Predizione e validazione interna *RCMD* by www.rcmd.it

**DATA SET di
composti**

**Generazione dei
descrittori**

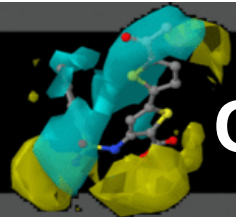
**Divisione dati in Training,
Test e Validation Set**

**Techniche Machine
Learning**

**Validazione Interna ed
esterna.
Selezione dei migliori
Modelli**

**Modelli QSAR
predittivo**

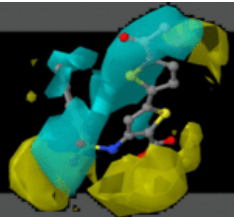
$$\left\{ \begin{array}{l} R^2(y, \hat{y}) = 1 - \frac{\sum_{i=0}^{n_{\text{samples}}-1} (y_i - \hat{y}_i)^2}{\sum_{i=0}^{n_{\text{samples}}-1} (y_i - \bar{y})^2} \\ \text{accuracy}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} 1(\hat{y}_i = y_i) \\ \text{MCC} = \frac{tp \times tn - fp \times fn}{\sqrt{(tp + fp)(tp + fn)(tn + fp)(tn + fn)}} \end{array} \right\}$$



Classificazione con FingerPrints

by www.RCMD.it

	Desc	Modello	Scalato	Accuracy CV	rocAUC	MCC CV	Accuracy ext	MCC ext	n_iter
1	Fing	SVC	No	0.88	0.90	0.71	0.99	1	100
2	Fing	Extra Trees	No	0.92	0.90	0.60	0.97	0.70	
3	Fing	Random forest	No	0.93	0.91	0.66	0.99	0.70	
4	Fing	GBM	No	1.0	0.9	0.65	1	0.7	
5	Fing	LogisticR	No	0.81	0.9	0.63	0.97	0.67	



Classificazione con Descrittori

by www.RCMD.it

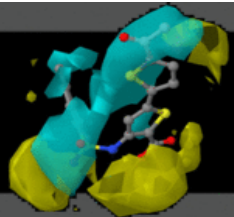
	Desc	Modello	Scalato	Accuracy CV	rocAUC	MCC CV	Accuracy ext	MCC ext	n_iter
1	Desc	LogisticR	si	0.86	0.87	0.62	0.93	0.65	100
2	Desc	SVC	si	0.8	0.90	0.62	0.98	0.62	
3	Desc	ExtraTrees	si	0.82	0.90	0.66	0.94	0.82	
4	Desc	Random Forest	si	0.81	0.91	0.63	0.98	1	
5	Desc	GradientBM	si	1.0	0.9	0.65	1	1	



Regressione con Fingerprints

by www.RCMD.it

		Modello	Scalato	q^2	q^2 Y-s	r^2 pred	SDEP ext	N_iter
1	Fing	PLS	No	0.71	-13	0.66	0.083	100
2	Fing	Kernel Ridge	No	0.68	-10	0.788	1.07	
3	Fing	SVM (rbf)	No	0.75	-23	0.58	0.23	
4	Fing	Random Forest	No	0.69	-8.2	0.68	0.032	
5	Fing	ExtraTrees	No	0.74	-7.1	0.7	0.083	
6	Fing	GradientBM	No	0.78	-6.7	0.655	0.23	



Regressione con Descrittori

by www.RCMD.it

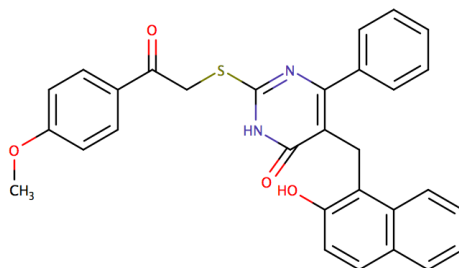
		Modello	Scalato	q^2	q^2 Y-s	r^2 pred	SDEP ext	N_iter
1	Desc	PLS	Si	0.56	-9.5	0.61	0.97	100
2	Desc	Ridge	Si	0.62	-12	0.46	0.84	
3	Desc	SVM (rbf)	Si	0.75	-0.3	0.69	0.6	
4	Desc	Extra Trees	Si	0.67	-0.23	0.46	1.03	
5	Desc	GradientBM	Si	0.61	-0.37	0.62	0.84	
6	Desc	Random forest	Si	0.66	-0.20	0.55	0.97	



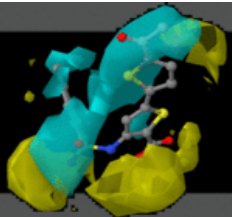
Validazione esterna dei modelli

by **RCMD**.it

Test Set Esterno
Composto da
128 composti da varie
pubblicazioni



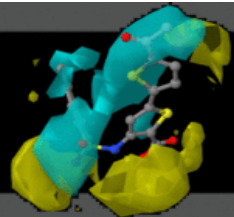
	pIC50_true	pIC50_pred	Predizione
	5.66	5.68	Attivo



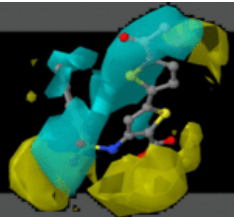
Conclusione e Obiettivi Futuri

by **RCMD**.it

- ◎ **Modelli generati risultano validi per la robusta attività predittive**
- ◎ L'Obiettivo futuro sarà l'utilizzo dei modelli per lo Screening di librerie commerciali
- ◎ Arricchimento del training set con I nuovi dati inseriti nel *ChEMBL* e *pubChem* e dai risultati dello Screening Virtuale in modo da ottenere i modelli più predittivi



Grazie per l'attenzione



Risultati

Risultati dei modelli di classificazione SIRT1- SIRT2 selettivi con la Validazione esterna

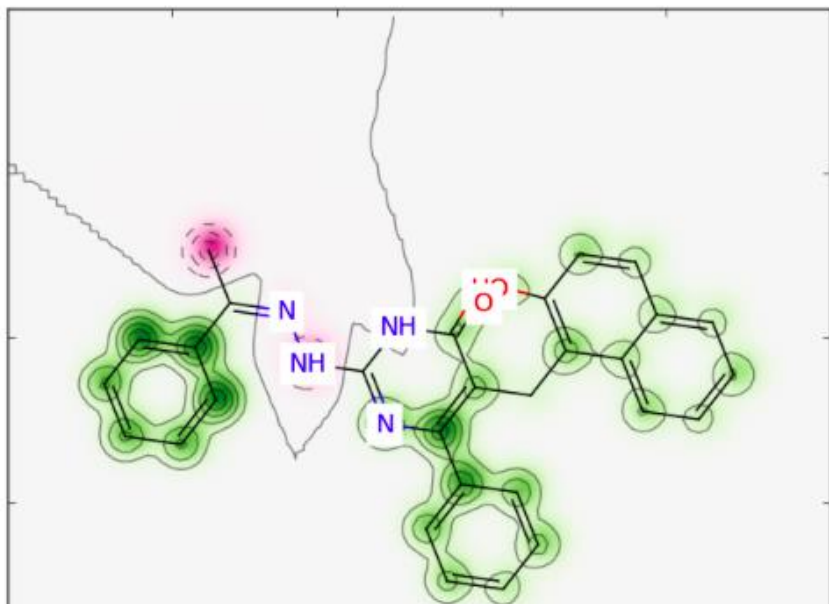
	Desc	METHOD	MCC CV	Mean accuracy	AUC Roc	MCC ext	n_iter
1	Fing	LG	0.95	0.99	0.8	0.98	100
2	Fing	GBC	0.98	0.975	0.74	0.95	
3	Fing	RFC	0.97	0.994	0.78	0.984	
4	Fing	EXC	0.96	0.993	0.62	0.84	
5	Fing	SVM	0.92	0.993	0.79	0.98	
6	desc	GBC	0.99	0.98	0.88	0.982	
7	desc	RFC	0.91	0.95	0.68	0.95	
8	desc	EXC	0.95	0.95	0.76	0.87	
9	desc	SVM	0.92	0.94	0.63	0.93	
10	desc	LG	0.98	0.98	0.81	0.86	



Risultati di predizione

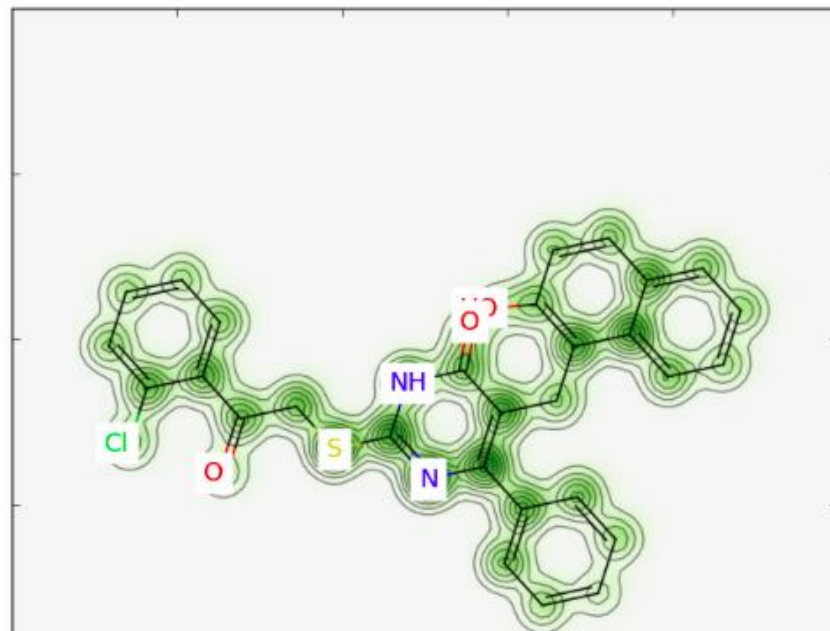
by **www.RCMD.it**

MC3191
(1, 2048)
Experimental Outcome: 617 5.68403
Name: ACTIVITY_MEAN, dtype: float64
Predicted Outcome [True]
Probability of being Not Active: 0.08924825311520837
Probability of being Active: 0.9107517468847915



Composto di Test esterno
MC3191

MC3185
(1, 2048)
Experimental Outcome: 611 6.522879
Name: ACTIVITY_MEAN, dtype: float64
Predicted Outcome [True]
Probability of being Not Active: 0.08890005152971861
Probability of being Active: 0.9110999484702813



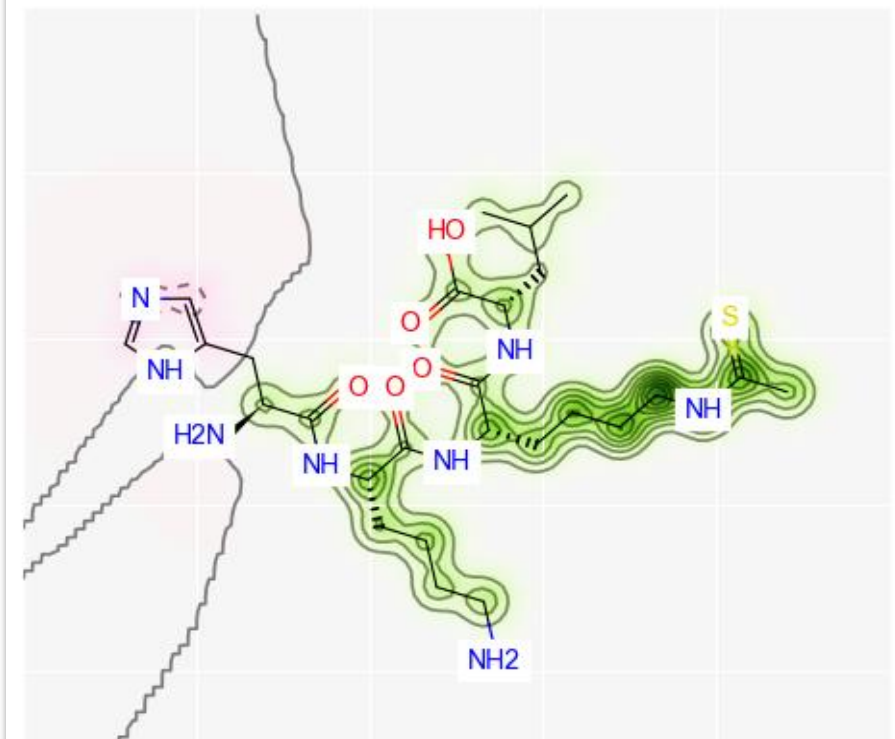
Composto di Test esterno
MC3185



Similarity Map Sirt2 vs Sirt1

by www.rcmd.it

Experimental SIRT1 Outcome: 6.318758762624412
 Experimental SIRT2 Outcome: 4.7011469235902945
 Predicted CLASS [True]
 Probability of beeing SIRT2 selective: 0.12968195141132188
 Probability of beeing SIRT1 selective: 0.8703180485886781



Experimental SIRT1 Outcome: 7.508638306165731
 Experimental SIRT2 Outcome: 8.318758762624412
 Predicted CLASS [True]
 Probability of beeing SIRT2 selective: 0.49461999432186066
 Probability of beeing SIRT1 selective: 0.5053800056781393

