Applicazione di algoritmi machine learning per lo sviluppo di modelli quantitativi di relazione struttura-attività (QSAR) per HDAC8 umana e di *Schistosoma mansoni*

SAPIENZA
UNIVERSITÀ DI ROMA

Tesi Sperimentale in Chimica Farmaceutica

Maria Stella Delle Chiaie
Matricola 1691945

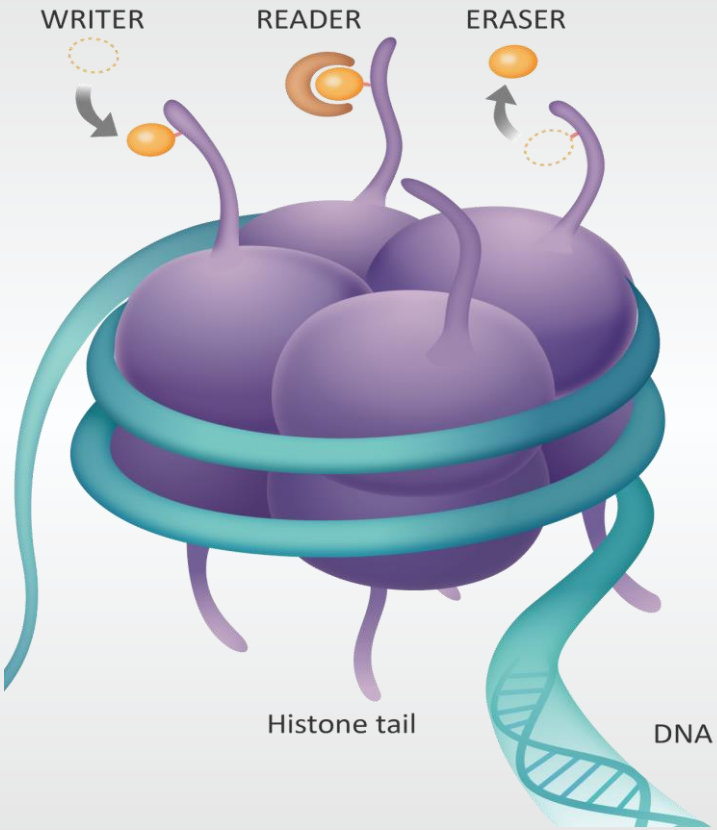
Relatore: Rino Ragno

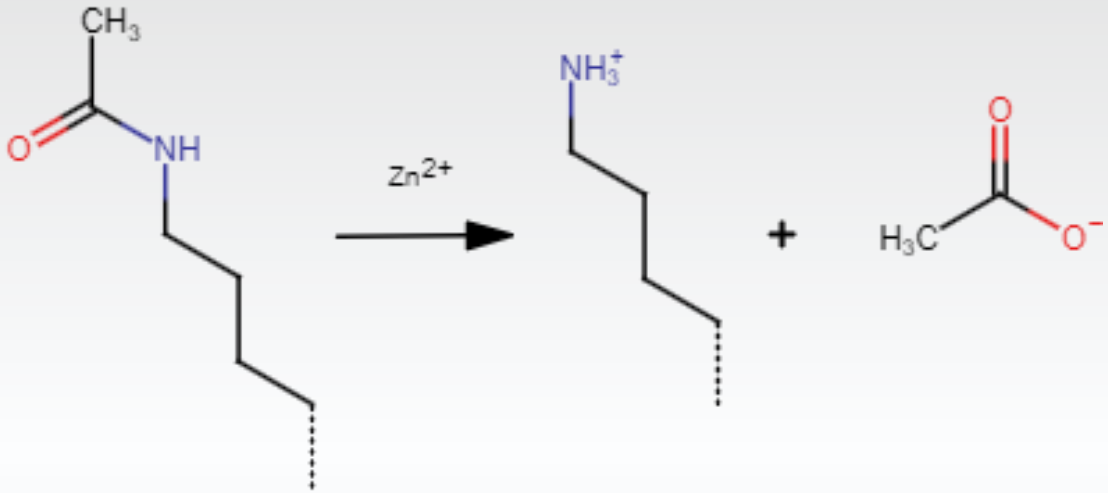
A.A. 2022-2023

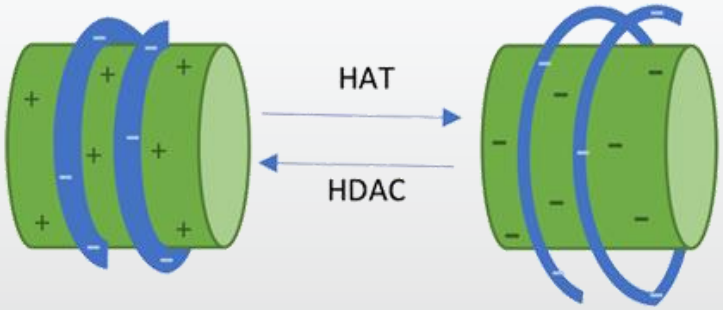
www.3D-QSAR.com

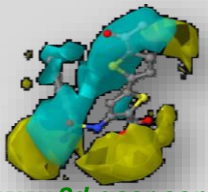
ROME CENTER FOR MOLECULAR DESIGN

MODIFICHE EPIGENETICHE









www.3d-qsar.com

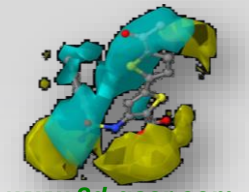
FAMIGLIA DELLE HDAC

Rome Center for Molecular Design
www.rcmd.it

Famiglia HDAC ***Classificate secondo la loro omologia di sequenza***		
Zn^{2+} dipendenti	1° classe	1, 2, 3, 8
	2° classe	4, 5, 6, 7, 9, 10
	4° classe	11
NAD^{+} dipendenti	3° classe (SIRTUINE)	1-7

Perché HDAC8

- Enzima strutturalmente ben caratterizzato
- Deacetila proteine istoniche e non istoniche
- Target infezioni da *Schistosoma mansoni*
- Target nella cura antitumorale
- La sfida attuale: sviluppo potenti inibitori selettivi HDAC8 (minori effetti avversi)



www.3d-qsar.com

SCOPO DEL PROGETTO SPERIMENTALE DI TESI

realizzare dei modelli predittivi su una serie di inibitori noti in letteratura per l'enzima HDAC8 umano e di *Schistosoma mansoni*.

Il modello si configura come uno strumento per effettuare uno studio di virtual screening per individuare potenziali nuovi inibitori di *h*HDAC8 e/o di *sm*HDAC8 da acquistare e/o sintetizzare.

METODOLOGIA PROGETTO SPERIMENTALE DI TESI

ChEMBL

python™

jupyter

Pretrattamento

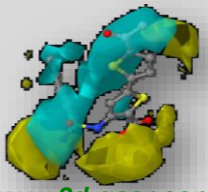
Cut-Off

Iperparametri

Modello Finale

DATA CLEANING

- Outliers
- Saggio ≠ B
- Attività ≠IC_{50}
- Unità di misura ≠ nM
- No smiles
- Studio duplicati
- Studio stereoisomeri
- Fps e Descs

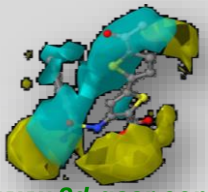


www.3d-qsar.com

HDAC8 umano

Rome Center for Molecular Design
www.rcmd.it

Molecole totali	**3204**
Outliers	3166
Analizzate con saggio ≠ B	3058
Con standard type ≠ IC_{50}	2253
Con unità di misura ≠ nM	2157
Senza smiles riportate	2151
Duplicati	534
Duplicati raggruppati	142
Raggruppamenti con σ pIC_{50} > 1	134
Molecole senza duplicati	1617
Con aggiunta dei raggruppati	**1751**



www.3d-qsar.com

HDAC8 S.M

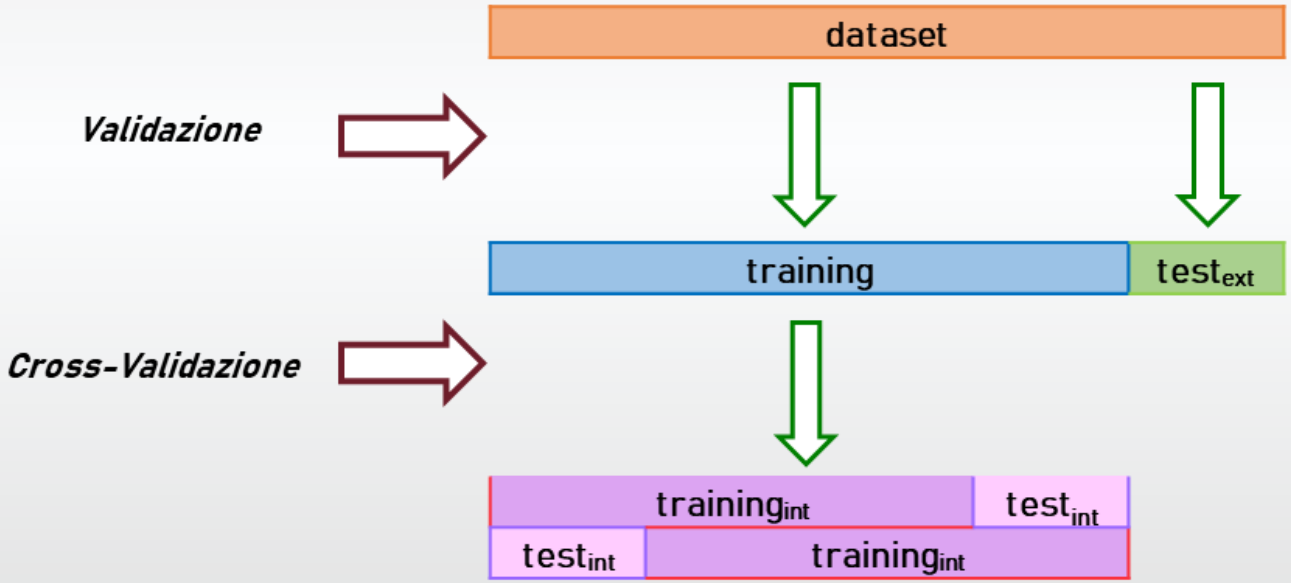
Rome Center for Molecular Design
www.rcmd.it

Molecole totali	**483**
Outliers	482
Analizzate con saggio ≠ B	482
Con standard type ≠ IC_{50}	258
Con unità di misura ≠ nM	256
Senza smiles riportate	256
Duplicati	96
Duplicati raggruppati	45
Raggruppamenti con σ pIC_{50} > 1	45
Molecole senza duplicati	160
Con aggiunta dei raggruppati	**205**

CONCETTI CHIAVE

- **FASE DI ALLENAMENTO**
- **FASE DI PREDIZIONE**

VALIDAZIONE ESTERNA e CROSS-VALIDAZIONE



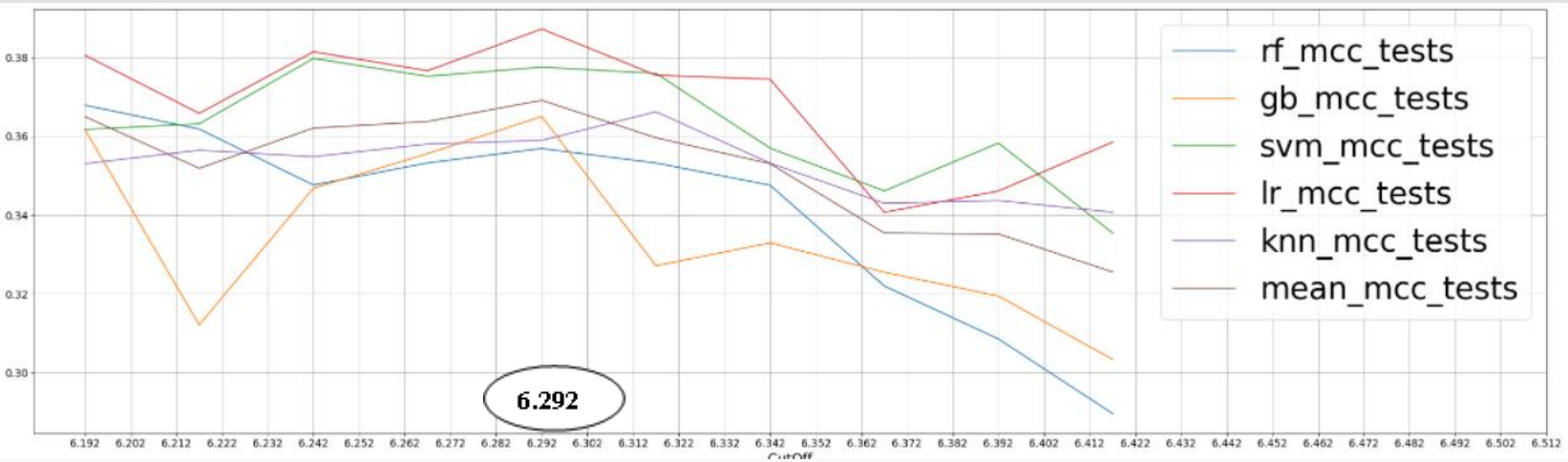
CONCETTI CHIAVE

MCC = Coefficiente di Correlazione di Metthews

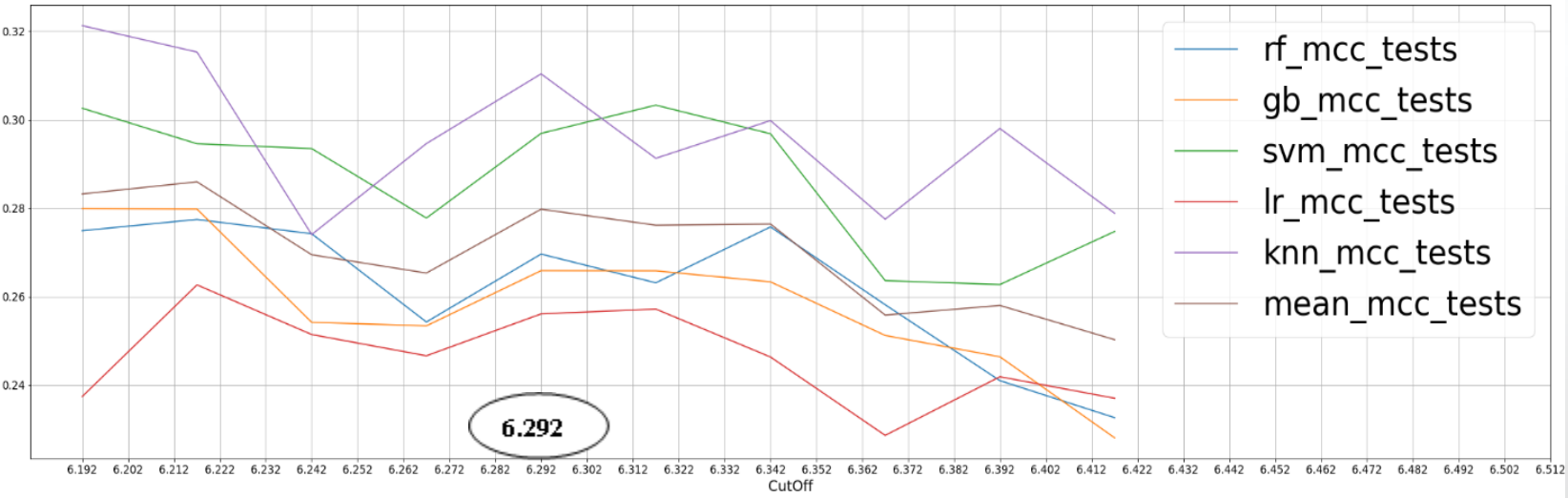
$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}$$

TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative

RICERCA CUT-OFF (HDAC8 umano)

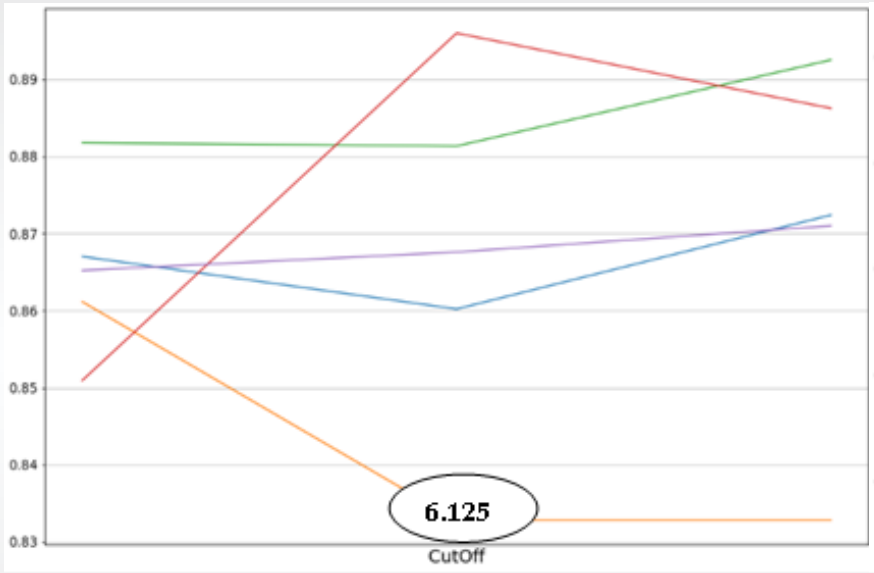


fps

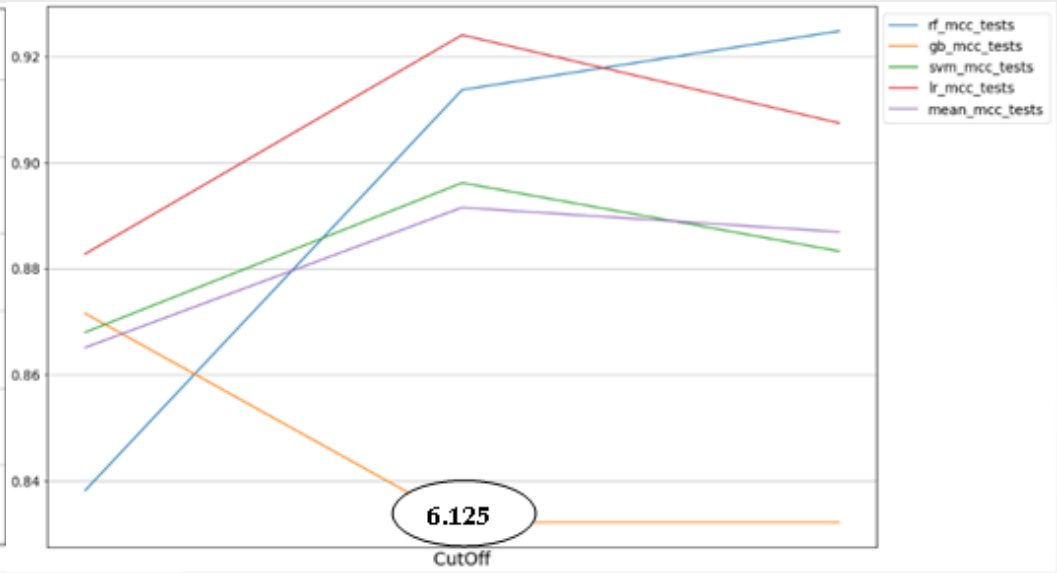


descs

RICERCA CUT-OFF (HDAC8 S.M.)

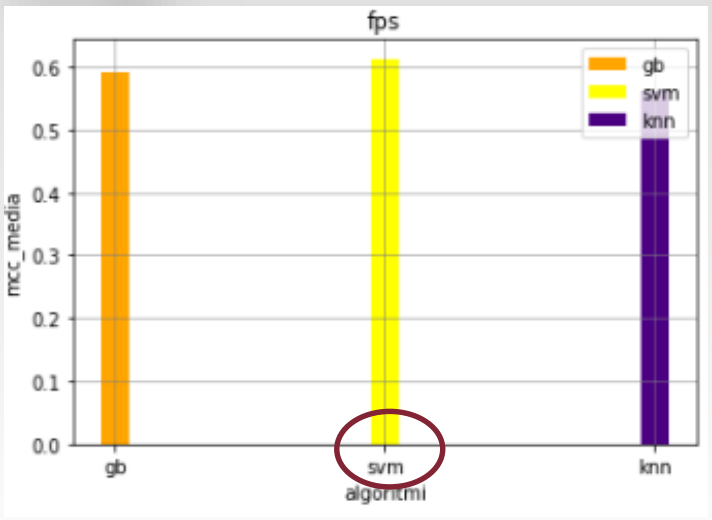


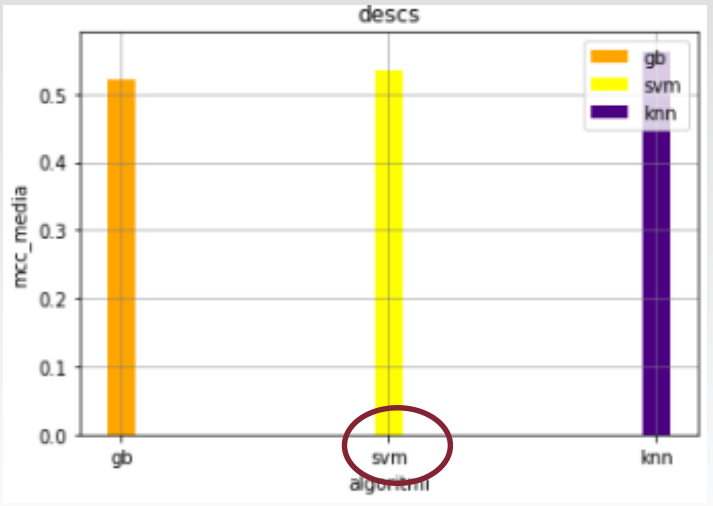
fps



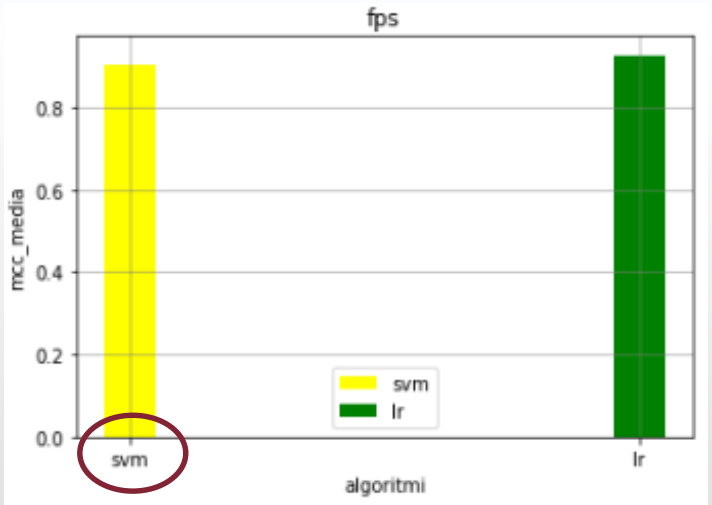
descs

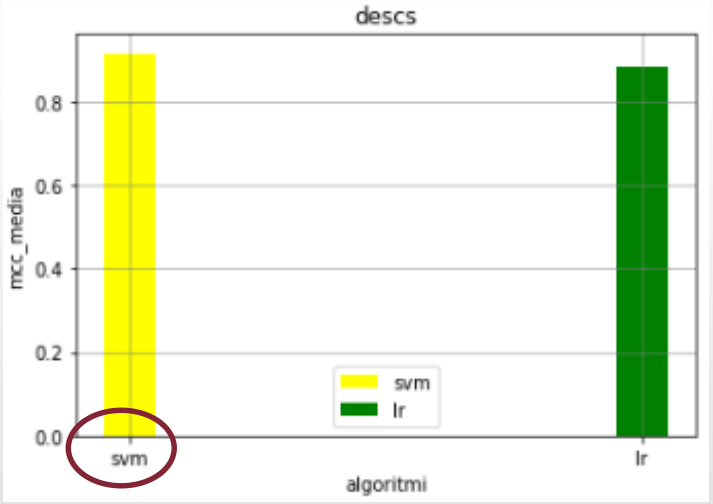
RISULTATI RICERCA MIGLIORE ALGORITMO



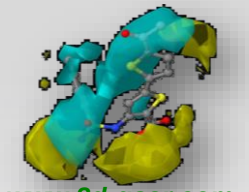


HDAC8 umano





HDAC8 S.M.



www.3d-qsar.com

MODELLO FINALE

Rome Center for Molecular Design
www.rcmd.it

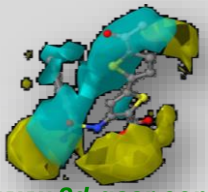
iper-parametri

*h*HDAC8

SVM	*fps*	mcc = 0.64	{'class_weight': {0: 1.0, 1: 1.1}, 'C': 41}
	descs	mcc = 0.37	

*sm*HDAC8

SVM	*fps*	mcc = 0.92	{'class_weight': {0: 1.1, 1: 1.0}, 'C': 11}
	descs	mcc = 0.83	



www.3d-qsar.com

CONCLUSIONI

Rome Center for Molecular Design
www.rcmd.it

- **L'approccio utilizzato ha permesso di costruire due modelli QSAR per *h*HDAC8 e *sm*HDAC8.**
- **Le combinazioni ottimali di iper-parametri hanno condotto a modelli molto robusti e altamente predittivi nei confronti di test set esterni.**
- **L'applicazione a due isoenzimi prevede anche la progettazione di molecole selettivamente attive nei confronti di *sm*HDAC8 selezionando in modo opportuno quelle che sono predette non attive nei confronti di *h*HDAC8.**
- **Questo protocollo permetterà di effettuare un Virtual Screening per selezionare nuove molecole potenzialmente attive nei confronti delle HDAC considerate riducendo così i tempi della progettazione iniziale di nuove molecole.**