Sviluppo e Validazione di un protocollo per la generazione di modelli 3-D QSAR predittivi mediante il portale 3d-qsar.com

SAPIENZA
UNIVERSITÀ DI ROMA

Facoltà di Farmacia e Medicina
Corso di laurea in Chimica e Tecnologia Farmaceutiche
Tesi Sperimentale in Chimica Farmaceutica
a.a. 2023-2024

Laureando: Mattia Frezzini

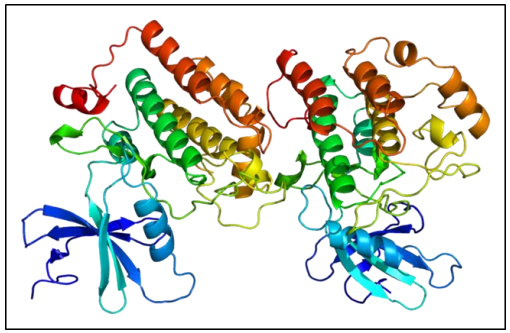
Matricola: 1846088

Relatore: Prof. Rino Ragno

DRUG DESIGN

DISPONIBILITÀ STRUTTURA TRIDIMENSIONALE

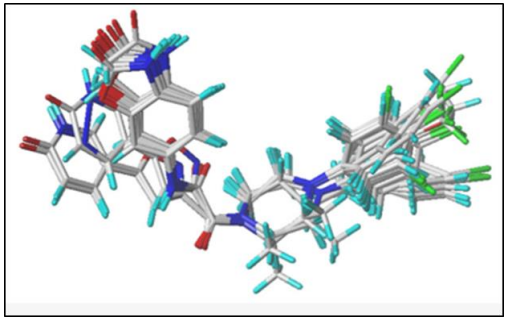
SI



STRUCTURE BASED

- DOCKING MOLECOLARE
- DINAMICA MOLECOLARE

NO



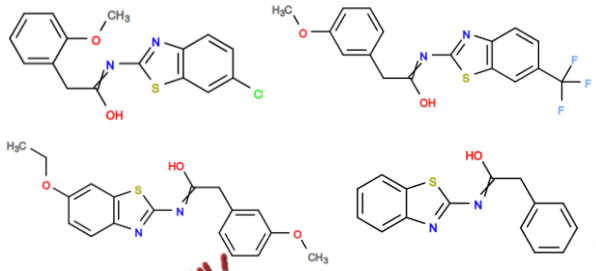
LIGAND BASED

- 3-D QSAR
- METODI FARMACOFORICI

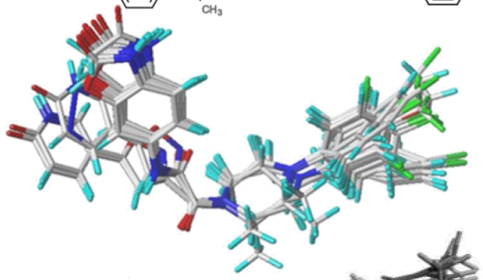
GENERAZIONE DEI MODELLI STATISTICO-MATEMATICI

3-D Quantitative Structure-Activity Relationships

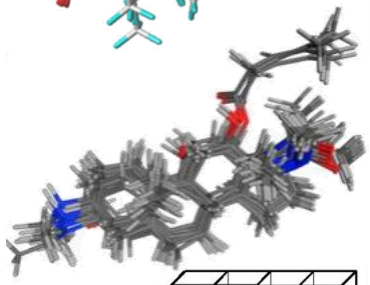
Scelta del Data Set → **Py-MolEdit**



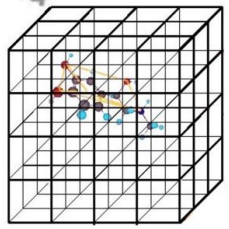
Analisi conformazionale → **Py-ConfSearch**



Allineamento delle molecole → **Py-Align**

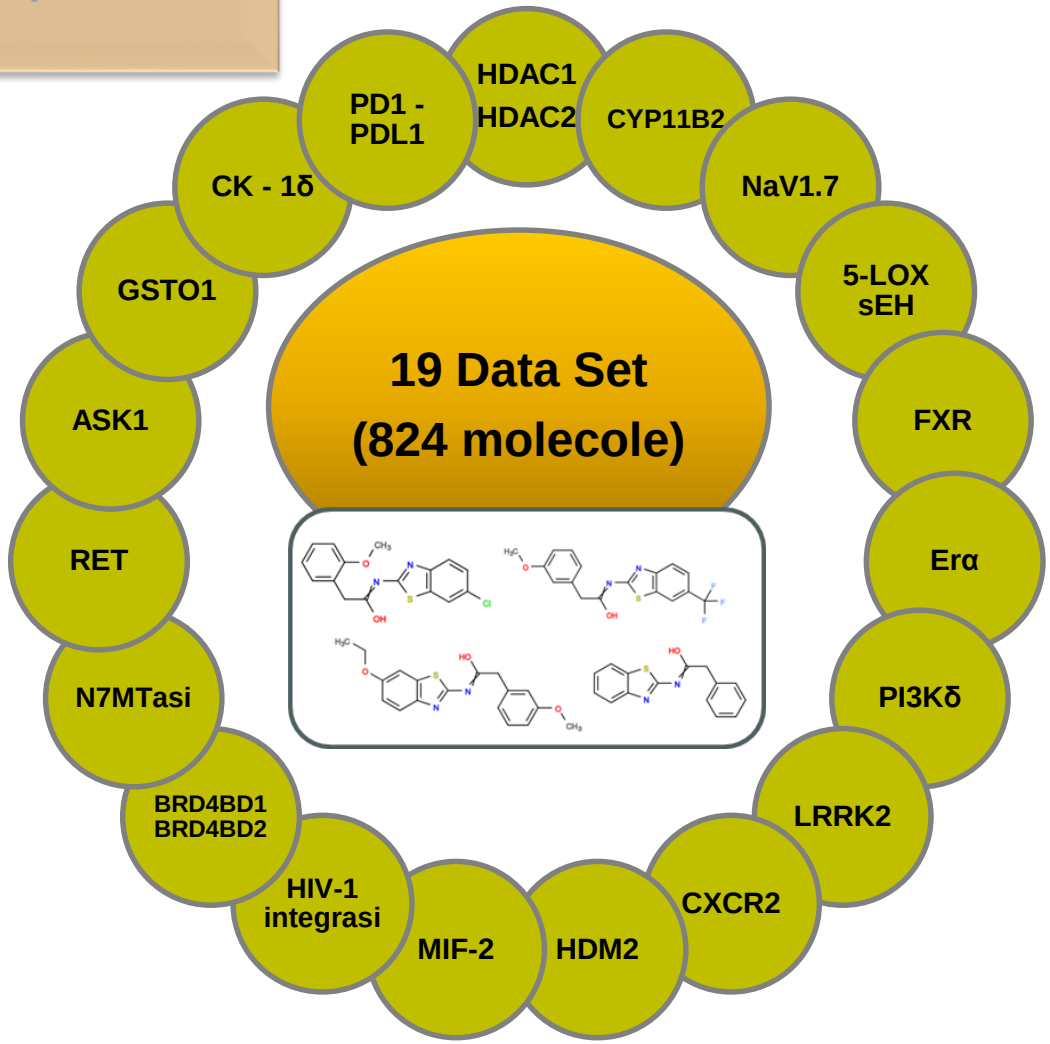


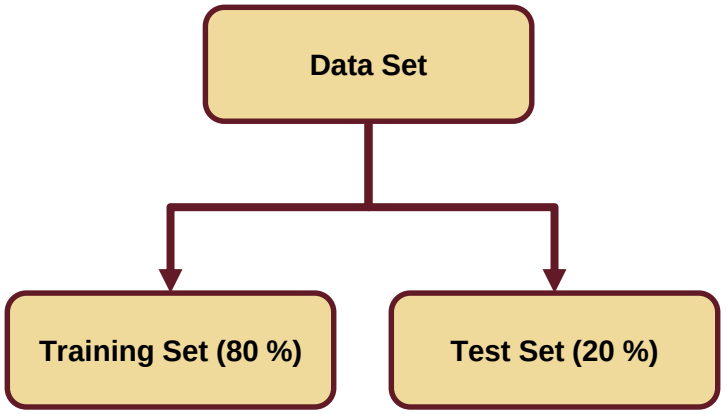
Calcolo dei campi di interazione molecolare (MIF) → **Py-CoMFA**



Scelta del Data Set

Py-MolEdit





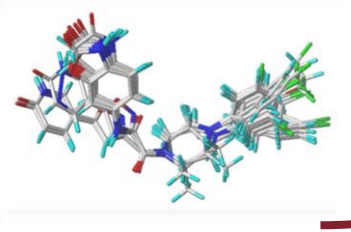
- Serie congeneriche
- Stesso meccanismo d'azione
- Stesse unità di misura
- Ampi intervalli di attività

Analisi conformazionale

Py-ConfSearch

Analisi conformazionale

- **Algoritmi Genetici**
- **Algoritmo della Distanza Geometrica**
  - **RDKit**
    - MMFF94
    - MMFF94s
    - UFF
  - **OpenBabel**
    - WEIGHTED
    - RANDOM
      - MMFF94
      - MMFF94s
      - UFF
      - GAF
      - GCHEMICAL
  - **Balloon**
    - EEM
    - SFKEEM
    - MMFF94
- **Ricerca Sistematica**
- **Ricerca Casuale**
- **Metodo di Monte Carlo** — Soon!
- **Dinamica Molecolare**
- **Simulated Annealing** — Soon!



16 analisi conformazionali

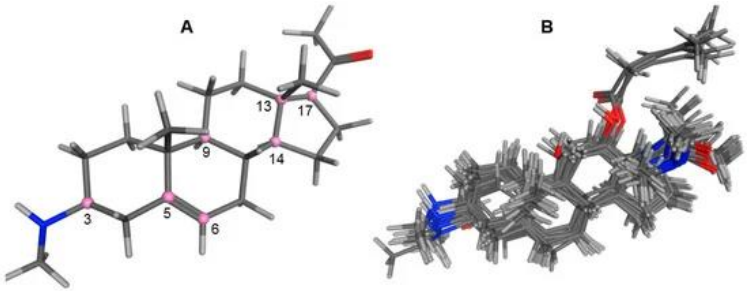
Allineamento delle molecole

Py-Align

Allineamento

RDKit
- BEST SCORE
- LOWEST RMSD
- TANIMOTO DISTANCE
- PROTRUDE DISTANCE

SHAEP
- ONLY SHAEP
- SIMILARITY

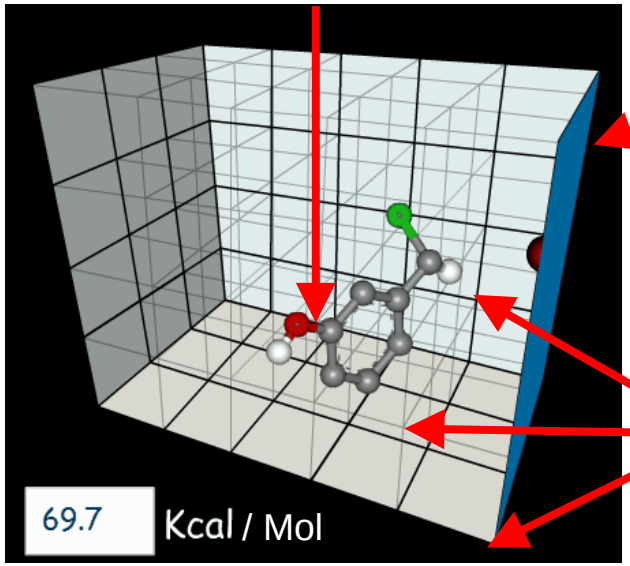


17 allineamenti ciascuno:
- Least active (Biological Activity)
- Most active (Biological Activity)
- Most flexible (Rotatable Bonds)
- Most rigid (Rotatable Bonds)
- Heaviest (MW)
- Longest (Conformation)
- Biggest (Molecular Volume)
- Least polar (TPSA)
- Most polar (TPSA)
- Highest MR (Molecular Refractivity)
- Lowest MR (Molecular Refractivity)
- Highest HD (Hydrogen Bond Donators)
- Lowest HD (Hydrogen Bond Donators)
- Highest HA (Hydrogen Bond Acceptors)
- Lowest HA (Hydrogen Bond Acceptors)
- Highest LogP (Partition Coefficient)
- Lowest LogP (Partition Coefficient)

Py-CoMFA

Molecular Interaction Fields

MOLECOLE ALLINEATE



GRIGLIA (GRID)

NODI CON SONDA (PROBES)

PARAMETRI DI DEFAULT
Probe: C.3
Probe Charge: 1.0
Dielectric Constant: 8
Grid Spacing: 2 Å
Cut off Value: 30 kcal/mol
Minimum Sigma: 2.0

Partial Least Square

$$r^2 = 1 - \frac{\sum(Y_i - Y_{i\,calc})^2}{\sum(Y_i - Y_{im})^2}$$

$$q^2 = 1 - \frac{\sum(Y_i - Y_{i\,pred})^2}{\sum(Y_i - Y_{im})^2}$$

$$FS = \frac{r^2 + q^2_{CV} + q^2_{ext} + [1 - |r^2 - q^2_{CV}|] + [1 - |r^2 - q^2_{ext}|] + [1 - |q^2_{CV} - q^2_{ext}|]}{6}$$

Validazione interna

MODELLI IN PREDIZIONE (scelta allineamento)

PARAMETRI DI DEFAULT
Method: ALL
Probe: C.3
Probe Charge: 1.0
Dielectric Constant: 8
Grid Spacing: 2 Å
Cut off Value: 30 kcal/mol
Minimum Sigma: 2

COMBINED CV Score	COMBINED Pred Score	COMBINED Full Score	MixedFields CV Score	MixedFields Pred Score	MixedFields Full Score	AllFields CV Score	AllFields Pred Score	AllFields Full Score
0.591	0.836	0.658	0.640	0.811	0.703	0.638	0.791	0.710
0.668	0.725	0.741	0.672	0.744	0.747	0.666	0.738	0.742
0.690	0.803	0.745	0.686	0.805	0.742	0.658	0.762	0.723
0.667	0.797	0.746	0.657	0.783	0.732	0.588	0.744	0.680

VARIABLE PRETREATMENT OPTIMIZATION (VPO1)

PARAMETRI
Method: ALL
Probe: diversi probe
Probe Charge: 1.0
Dielectric Constant: (1; 81; 1)
Grid Exstension: (3; 11; 1)
Cut off: True e False
Minimum SDEP increment: 0.05
Random Combinations: 200

COMBINED CV Score	COMBINED Pred Score	COMBINED Full Score	MixedFields CV Score	MixedFields Pred Score	MixedFields Full Score	AllFields CV Score	AllFields Pred Score	AllFields Full Score
0.591	0.836	0.658	0.640	0.811	0.703	0.638	0.791	0.710
0.668	0.725	0.741	0.672	0.744	0.747	0.666	0.738	0.742
0.690	0.803	0.745	0.686	0.805	0.742	0.658	0.762	0.723
0.667	0.797	0.746	0.657	0.783	0.732	0.588	0.744	0.680

VARIABLE PRETREATMENT OPTIMIZATION (VPO2)

PARAMETRI
Method: specific method
Probe: diversi probe
Probe Charge: 1.0
Dielectric Constant: (1; 81; 1)
Grid Exstension: (3; 11; 1)
Cut off: True e False
Minimum SDEP increment: 0.05
Random Combinations: 200

COMBINED CV Score	COMBINED Pred Score	COMBINED Full Score	MixedFields CV Score	MixedFields Pred Score	MixedFields Full Score	AllFields CV Score	AllFields Pred Score	AllFields Full Score
0.591	0.836	0.658	0.640	0.811	0.703	0.638	0.791	0.710
0.668	0.725	0.741	0.672	0.744	0.747	0.666	0.738	0.742
0.690	0.803	0.745	0.686	0.805	0.742	0.658	0.762	0.723
0.667	0.797	0.746	0.657	0.783	0.732	0.588	0.744	0.680

Y-Scrambling

Y-scrambling (YS) → **Individuazione modelli migliori** → Nessun modello affetto da correlazione fortuita

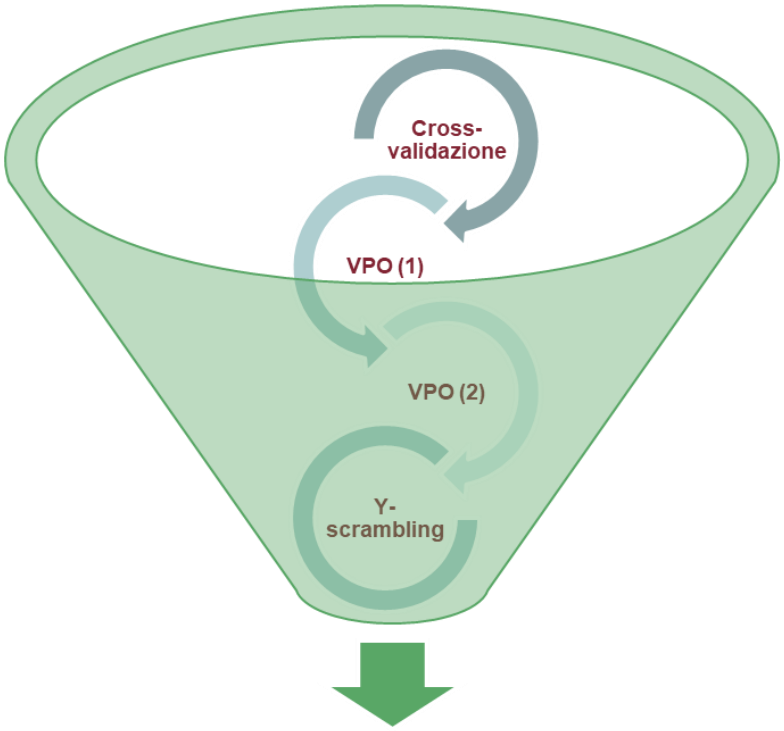
	DatasetName	DatasetNumber	Model ID	VPO	Method	nTrMols	nTsMols	r2 STE	q2cv STE	q2ext STE	r2 ELE	q2cv ELE	q2ext ELE	r2 SUM	q2cv SUM	q2ext SUM	r2 COMBINED	q2cv COMBINED	q2ext COMBINED	PC STE
0		108931	1136559	True	Autogrid	38	10	0.234	-0.300	0.255	0.419	0.048	0.232	0.775	-0.040	0.422	0.484	-0.008	0.323	1
1		108931	1136562	True	EasyMIF	38	10	0.211	-0.348	0.073	0.503	0.087	0.212	0.529	0.024	0.159	0.542	0.029	0.226	1
2		108931	1136554	True	Autogrid	38	10	0.335	-0.325	0.123	0.445	0.078	0.301	0.564	0.159	0.332	0.536	-0.045	0.338	1

```
Modify the dataset:             https://db.3d-qsar.com/v2/pycomfa/dataset/new?data_set=108931
To launch the YS and save maps: https://db.3d-qsar.com/v2/pycomfa/comfa/new?data_set=108931&model=1136559

        Please in the settings check the followings:
                Minimum SDEP Increment: --> 0.05
                Make Y-Scrambling:      --> True
                Save Maps:              --> True

View the dataset:               https://db.3d-qsar.com/v2/pycomfa/comfa?data_set=108931
View the alignment:             https://db.3d-qsar.com/v2/pycomfa/molecule/view?data_set=108931&

Modify the dataset:             https://db.3d-qsar.com/v2/pycomfa/dataset/new?data_set=108931
To launch the YS and save maps: https://db.3d-qsar.com/v2/pycomfa/comfa/new?data_set=108931&model=1136562

        Please in the settings check the followings:
                Minimum SDEP Increment: --> 0.05
                Make Y-Scrambling:      --> True
                Save Maps:              --> True

View the dataset:               https://db.3d-qsar.com/v2/pycomfa/comfa?data_set=108931
View the alignment:             https://db.3d-qsar.com/v2/pycomfa/molecule/view?data_set=108931&
```

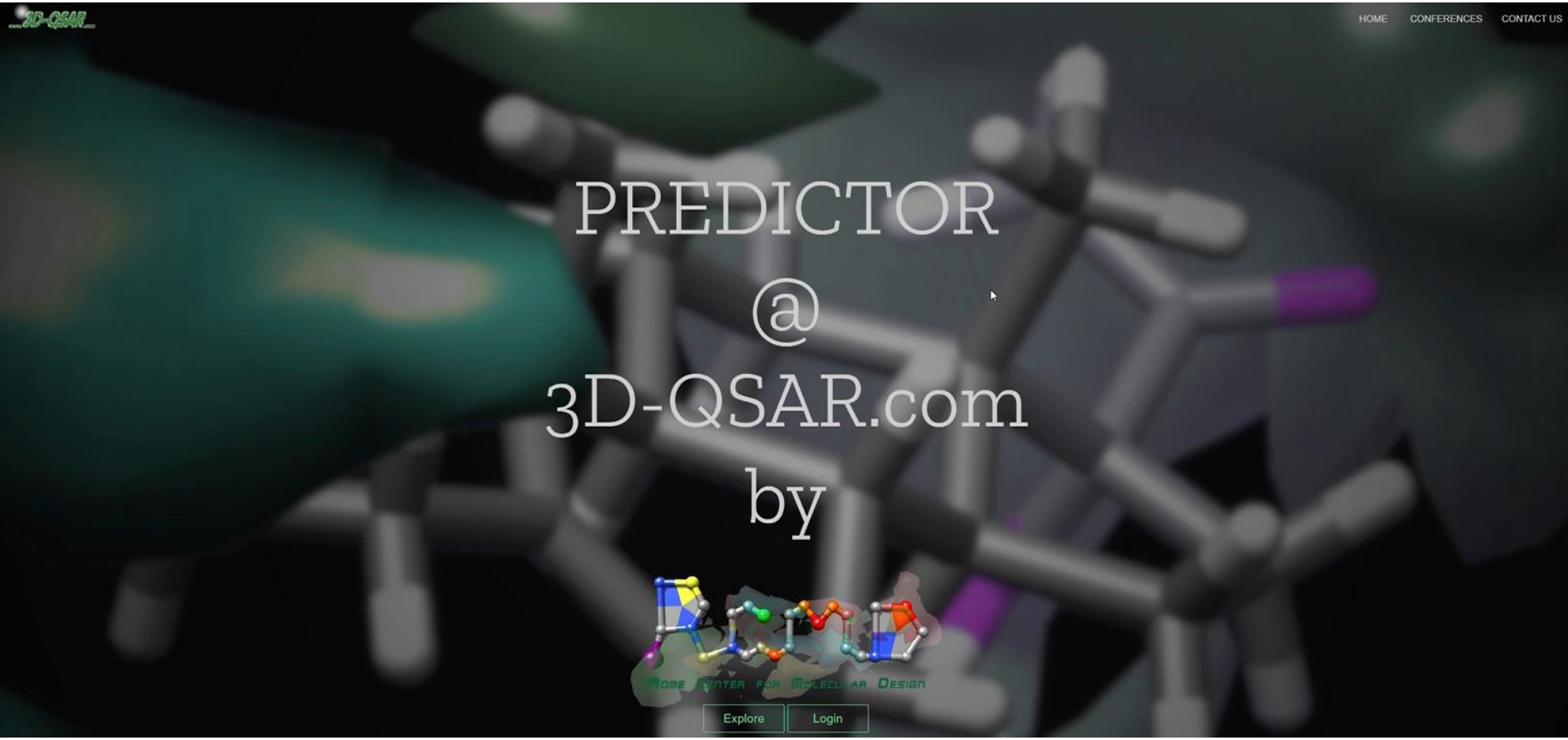
Modelli

Target	# modelli
HDAC1	4
HDAC2	4
CYP11B2	4
NaV1.7	4
5LOX	1
sEH	1
FXR	3
Erα	3
PI3Kδ	5
LRRK2	3
CXCR2	4
HDM2	4
MIF2_R110	3
T124/T125	3
BRD4BD1	3
BRD4BD2	4
N7_Mtasi	3
RET	3
ASK1	4
GSTO1	4
CK1δ	2
PD1_PDL1	2

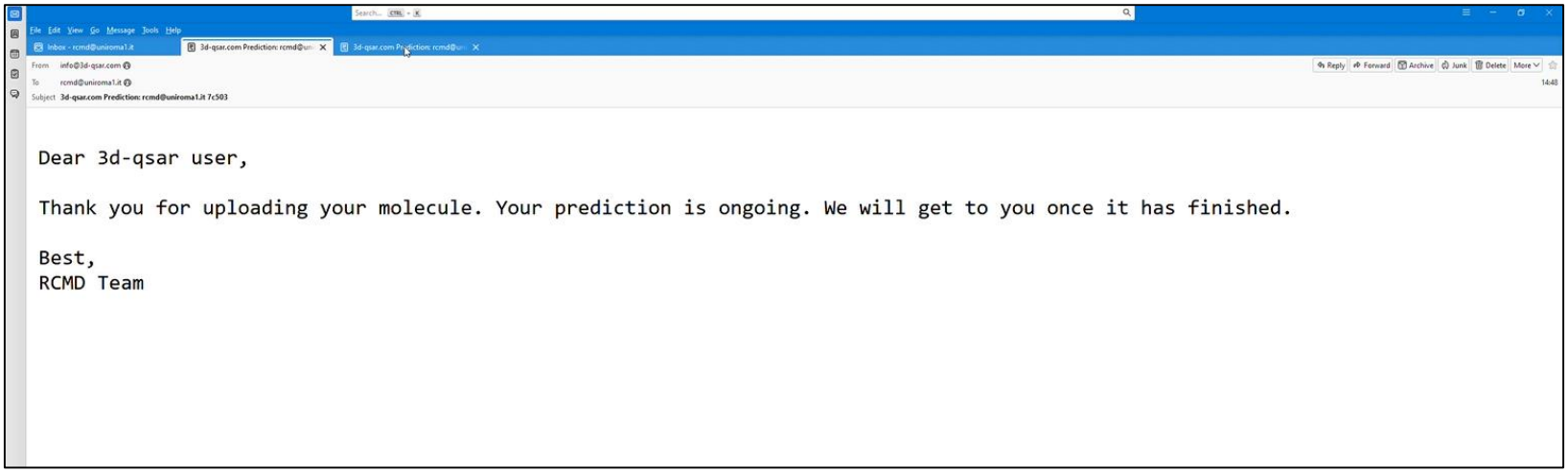
64·592 MODELLI



79 MODELLI


3D-QSAR
HOME
CONFERENCES
CONTACT US
PREDICTOR
@
3D-QSAR.com
by
Explore
Login

db.3dqsar.com



Dear 3d-qsar user,

Thank you for uploading your molecule. Your prediction is ongoing. We will get to you once it has finished.

Best,
RCMD Team

Dear 3d-qsar user,

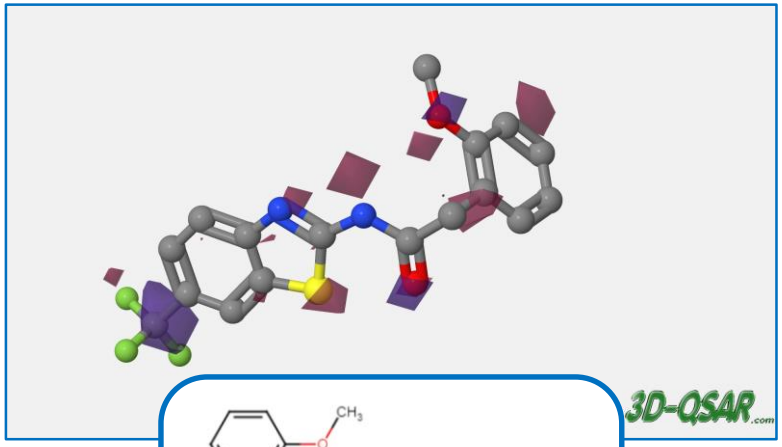
Your prediction has finished.

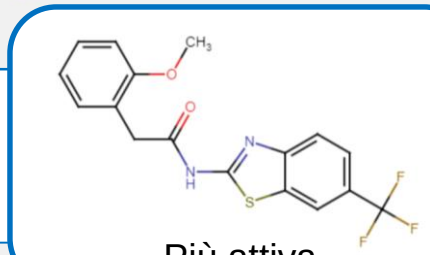
You can see the results at the following link: https://db.3d-qsar.com/details?model=CHEMBL1075104-1&email=rcmd@uniroma1.it&code=7c503

Otherwise can use the following credentials in the model page:
- email: rcmd@uniroma1.it
- code: 7c503
- link: https://db.3d-qsar.com/details?model=CHEMBL1075104-1

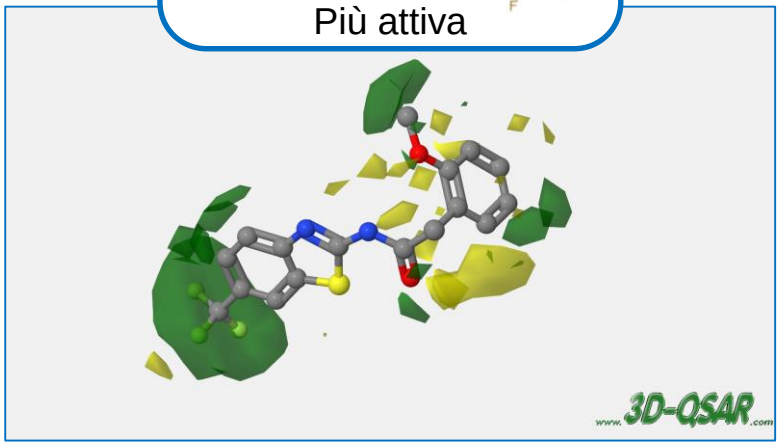
Best,
RCMD Team

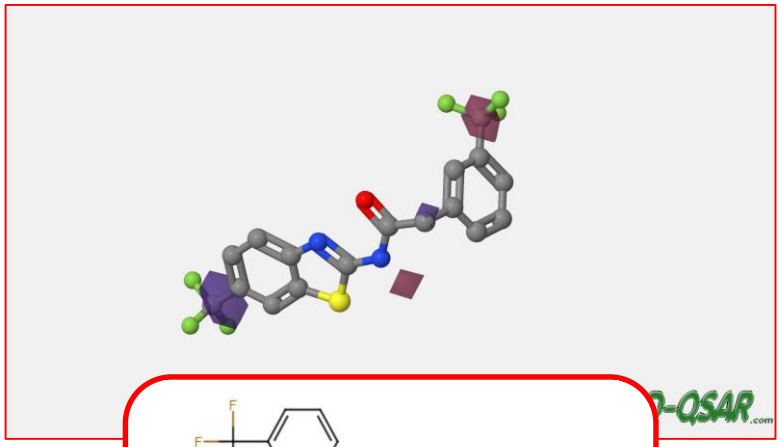
CHEMBL-4523610-1 – Inibitori di CK1δ

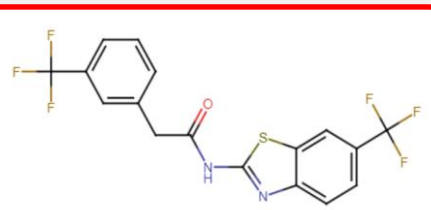




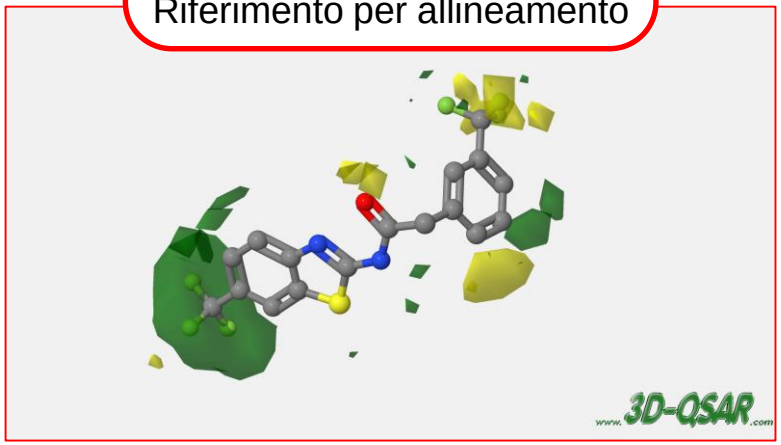
Più attiva







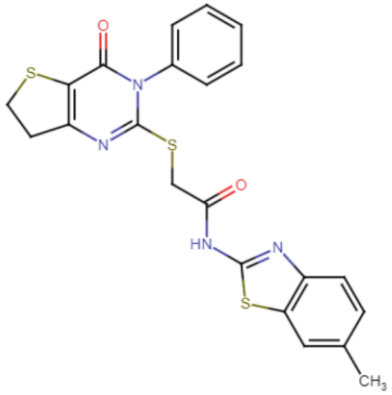
Riferimento per allineamento



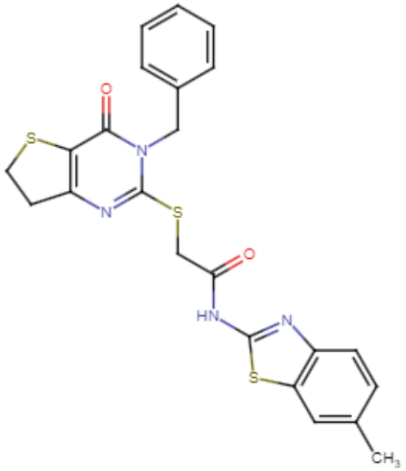
CHEMBLID	PC	r^2	SDEC	q^2_{cv}	$SDEP_{cv}$	q^2_{ext}	r^2_{YS}	$SDEC_{YS}$	$SDEP_{ext}$	q^2_{YS}	$SDEP_{YS}$	Probe
CHEMBL 4523610-1	3	0.86	0.228	0.567	0.505	0.828	0.313	0.758	0.378	0.021	0.759	C.3H3

CHEMBL-4523610-1 - Inibitori di CK1δ

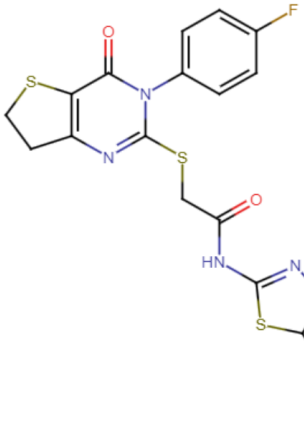
Nome Dataset	Smiles	IC_{50} (μM)	Attività testata pAct	Attività predetta pAct	Modello usato per predizione	% errore
CK1δ	O=C1N(C2=CC=CC=C2)C(SCC(NC3=NC(C=CC(C)=C4)=C4S3)=O)=NC5=C1SCC5	0.32	6.495	6.018	CHEMBL4523610-1	7.344
CK1δ	O=C1N(CC2=CC=CC=C2)C(SCC(NC3=NC(C=CC(C)=C4)=C4S3)=O)=NC5=C1SCC5	0.42	6.376	5.974	CHEMBL4523610-1	6.305
CK1δ	O=C1N(C2=CC=C(F)C=C2)C(SCC(NC3=NC(C=CC(C)=C4)=C4S3)=O)=NC5=C1SCC5	0.55	6.26	6.343	CHEMBL4523610-1	1.326
CK1δ	O=C1N(C2=CC=CC=C2OC)C(SCC(NC3=NC(C=CC(C)=C4)=C4S3)=O)=NC5=C1SCC5	1.2	5.92	5.996	CHEMBL4523610-1	1.284



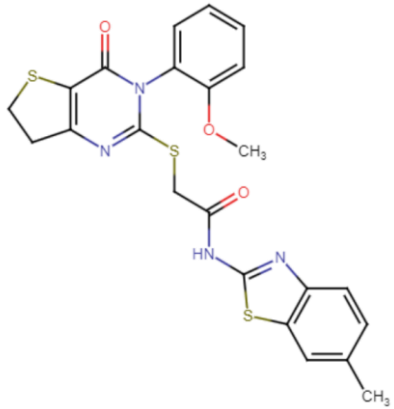
Avg Similarity	Max Similarity	Min Similarity
0.351	0.471	0.248



Avg Similarity	Max Similarity	Min Similarity
0.356	0.458	0.264



Avg Similarity	Max Similarity	Min Similarity
0.329	0.421	0.230



Avg Similarity	Max Similarity	Min Similarity
0.362	0.549	0.244

Risultati

1

Proporzionalità diretta tra massima analogia strutturale e attività predetta

2

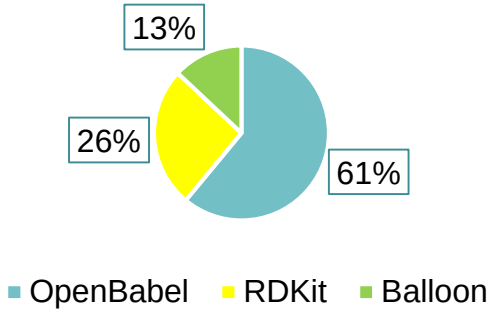
Tutti i modelli generati sullo stesso target hanno previsto una risposta identica circa il campo sterico e quello elettrostatico

3

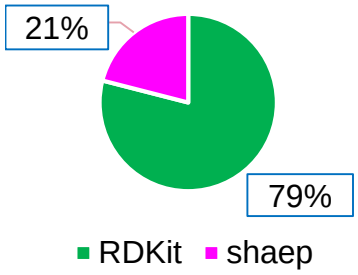
Probe utilizzato come maggior frequenza: C.3H3

4

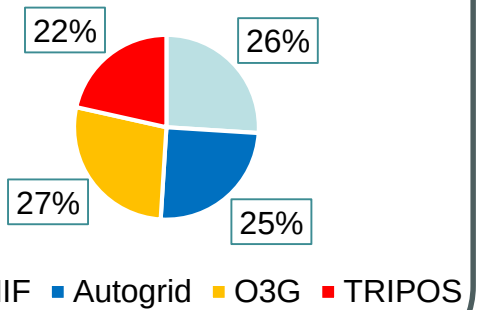
Metodo di analisi conformazionale



Metodo di allineamento



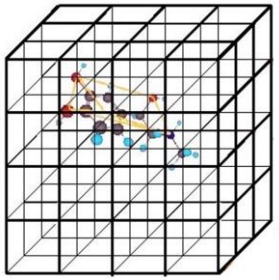
Metodo CoMFA

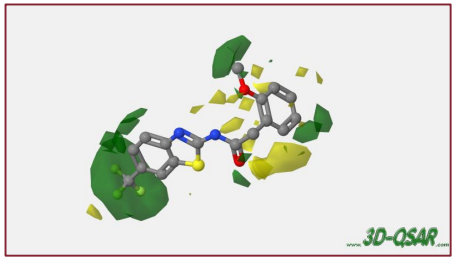


Conclusioni

- PRIMO DATA BASE AL MONDO DI MODELLI 3-D QSAR
- PUBBLICATI 145 MODELLI PER 57 TARGET DIFFERENTI

Possibili applicazioni e sviluppi futuri





- **Facile utilizzo**
- **Aumento di modelli nel data base**
- **Includere nuove strutture nel data set dei modelli per aumentare la predittività**
- **Sviluppare i modelli non predittivi con nuovi metodi di allineamento (già implementati)**
- **Altri database: Structure-Based e QSAR con machine-learning**

GRAZIE PER L'ATTENZIONE!

UN RINGRAZIAMENTO PARTICOLARE PER IL SUPPORTO

PROF. RINO RAGNO
DOTT.SSA ELEONORA PROIA
DOTT.SSA ROBERTA ASTOLFI
DOTT.SSA LIDIA GIULIANI
DOTT. FILIPPO SAPIENZA