METODICHE DI MACHINE LEARNING E STRUCTURE BASED DRUG DESIGN PER L'INDIVIDUAZIONE DI NUOVI LIGANDI DI K-RAS

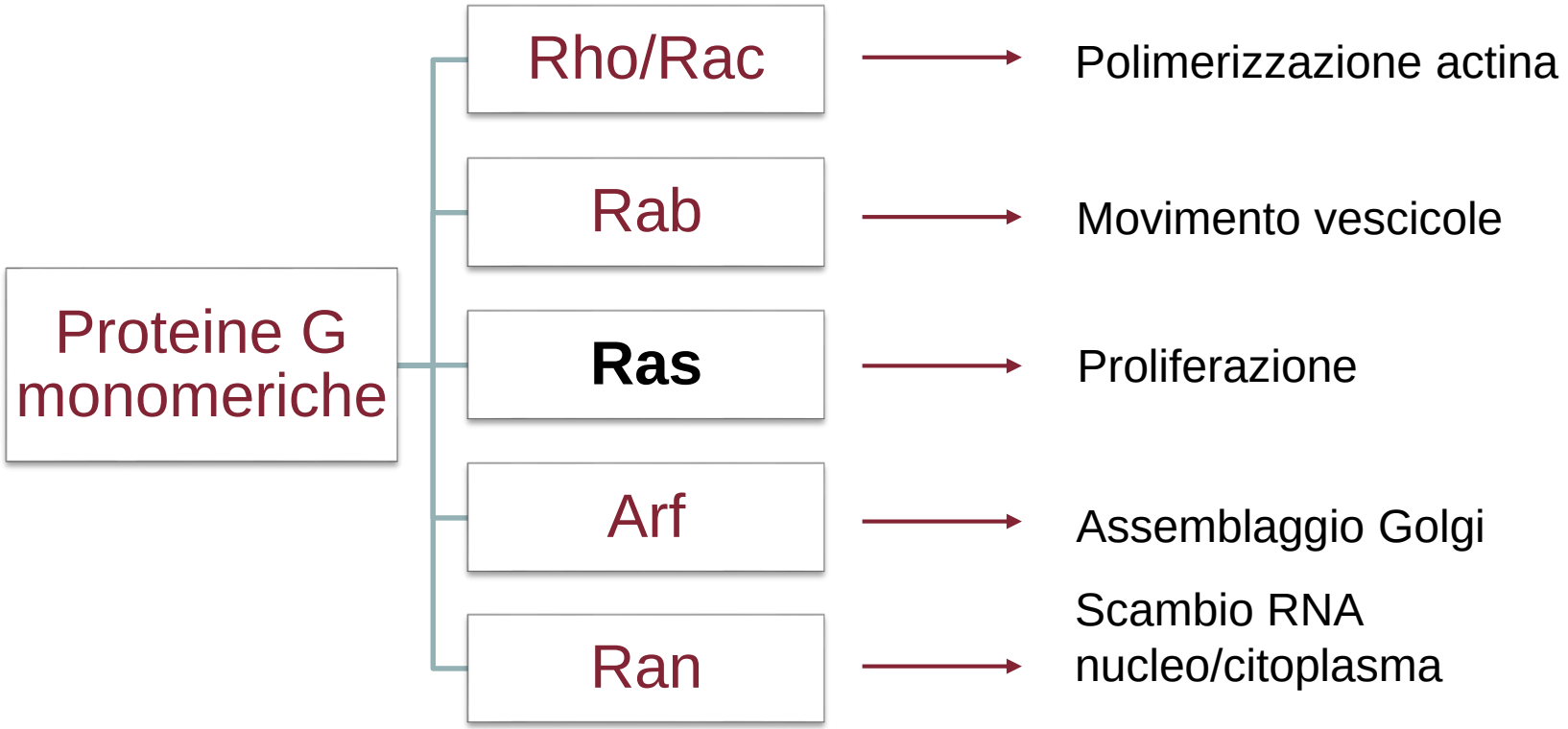
SAPIENZA
UNIVERSITÀ DI ROMA

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

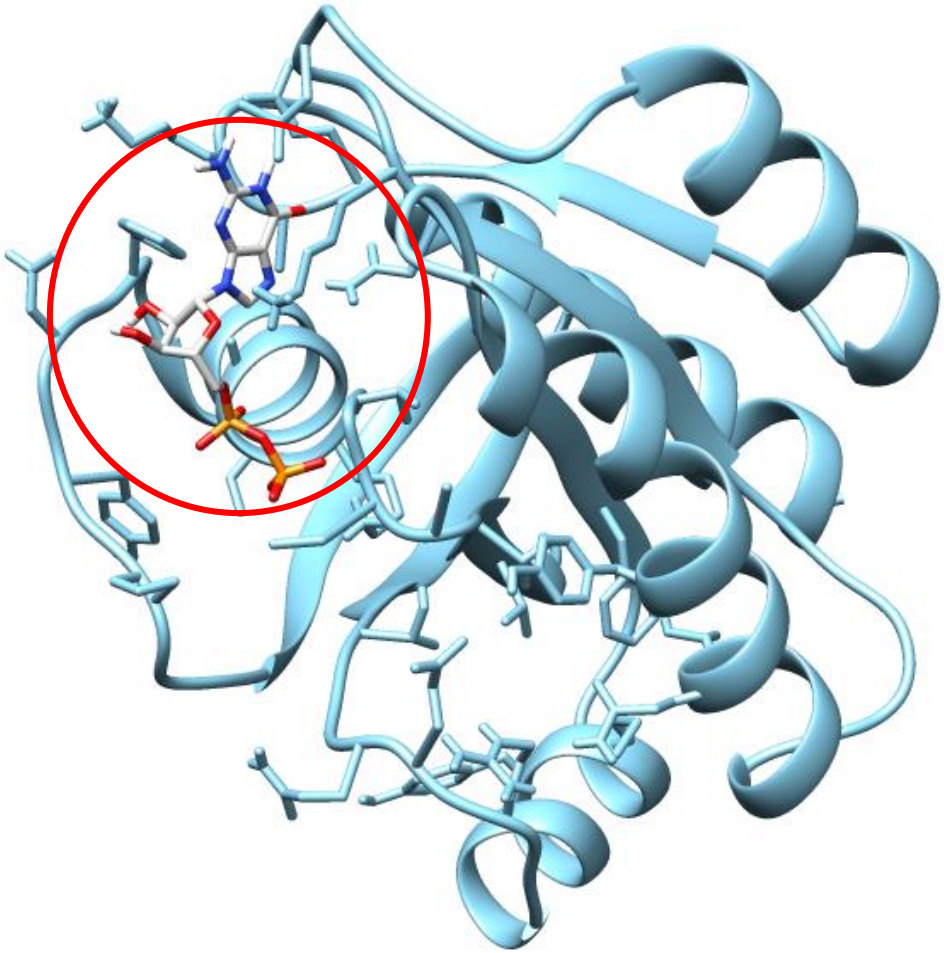
Laureando: Leonardo Contreas
Matricola: 1556506

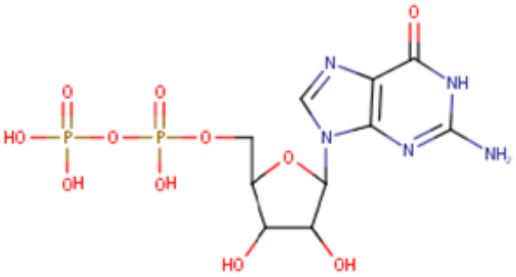
Relatore: prof. Rino Ragno

La superfamiglia



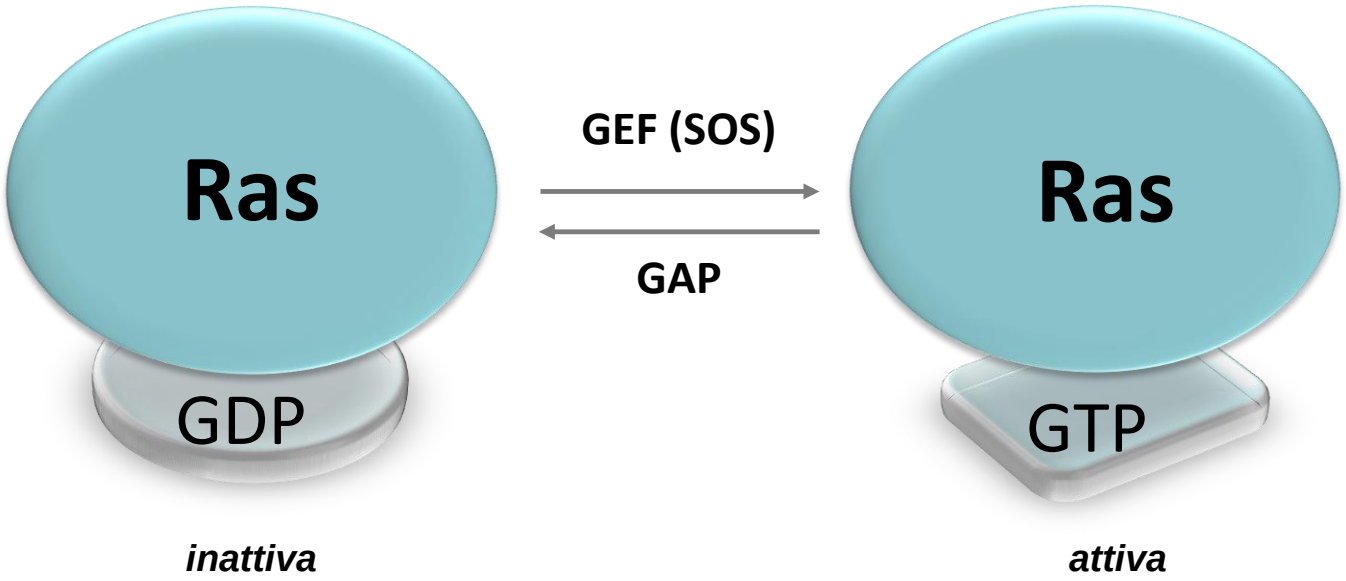
La proteina – KRAS4B





Le proteine Ras

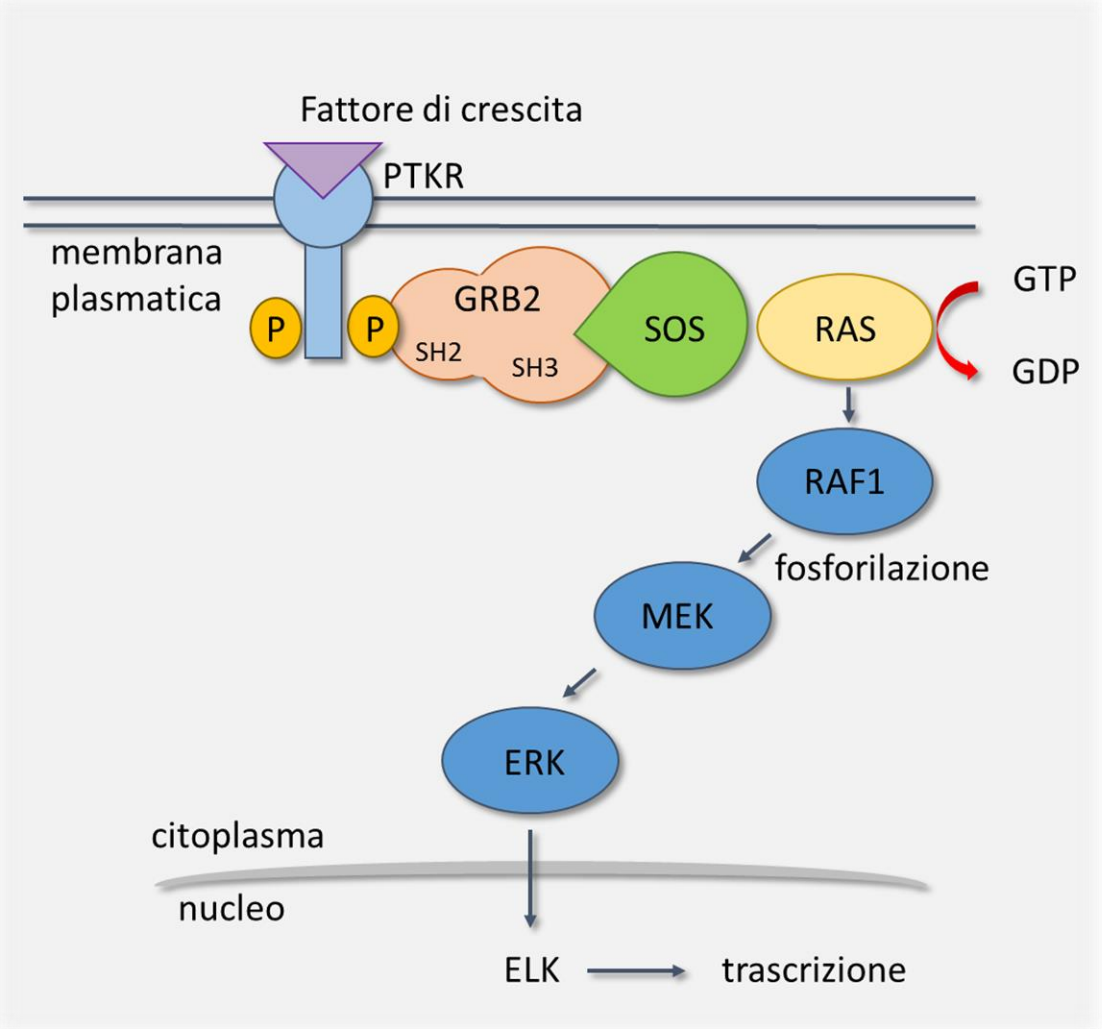
Interruttori molecolari
"on/off"

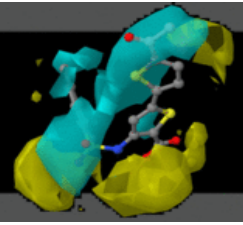


Lo stato funzionale delle proteine Ras è regolato da:

- **GEFs** → guanine nucleotide exchange factors
- **GAPs** → GTPases-activating proteins

Il pathway


Fattore di crescita
PTKR
membrana
plasmatica
P
P
GRB2
SH2
SH3
SOS
RAS
GTP
GDP
RAF1
fosforilazione
MEK
ERK
citoplasma
nucleo
ELK
trascrizione



Come agire?

- Proteina considerata «undruggable»
- Strategie:

✓ *Blocco delle modifiche post-traduzionali*

✓ *Inibire il legame KRAS/GDP*

Covalenti

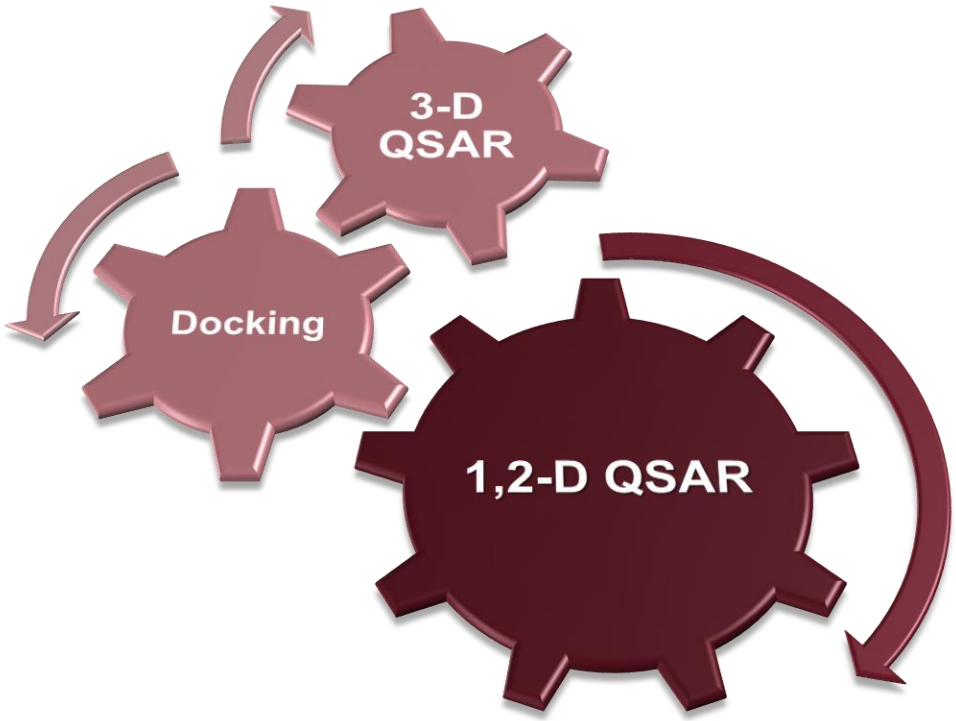
~~✓ Competitivi~~

✓ Allosterici

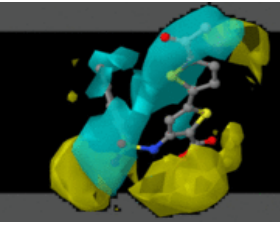
Reversibili

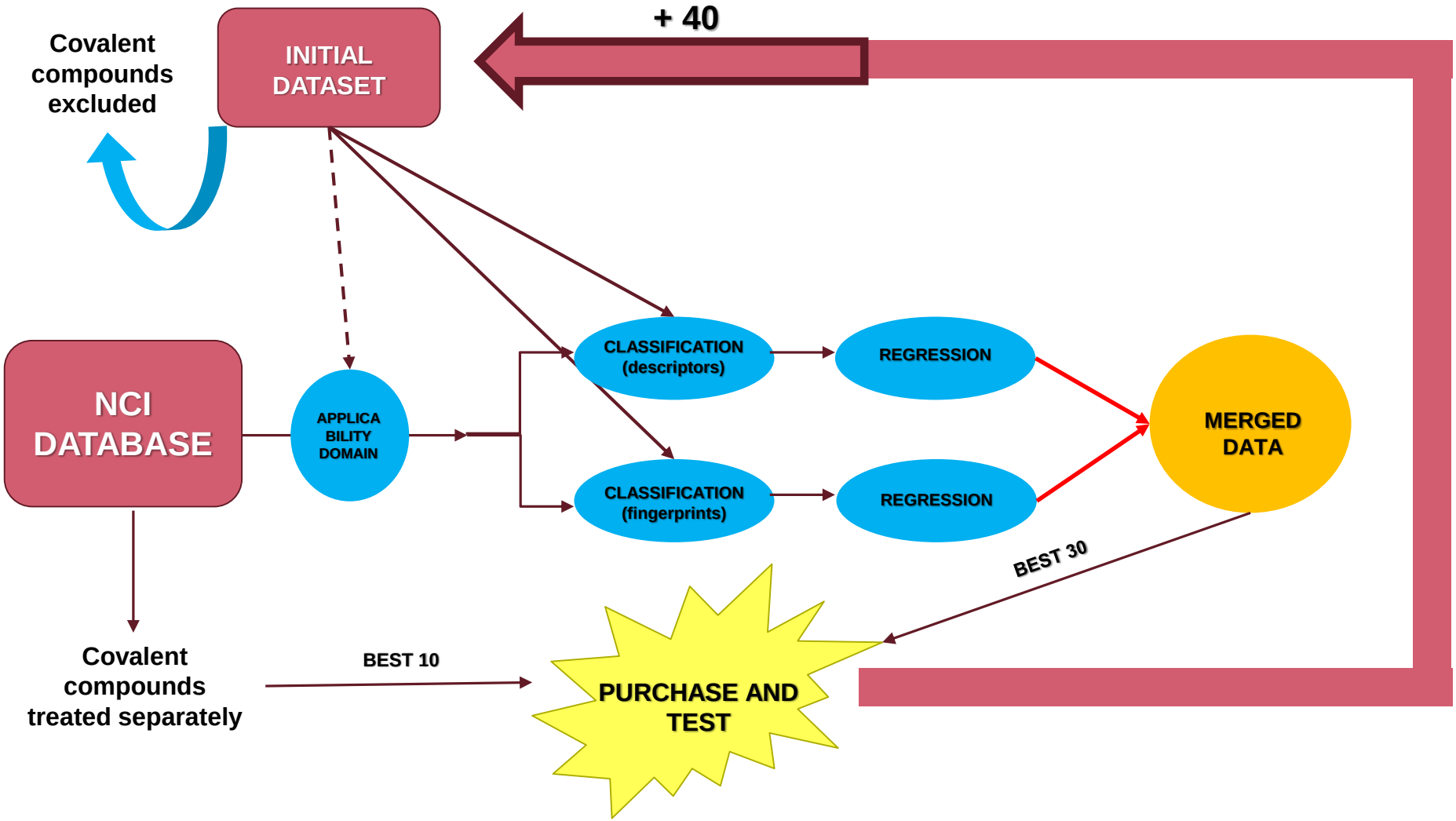
✓ Allosterici

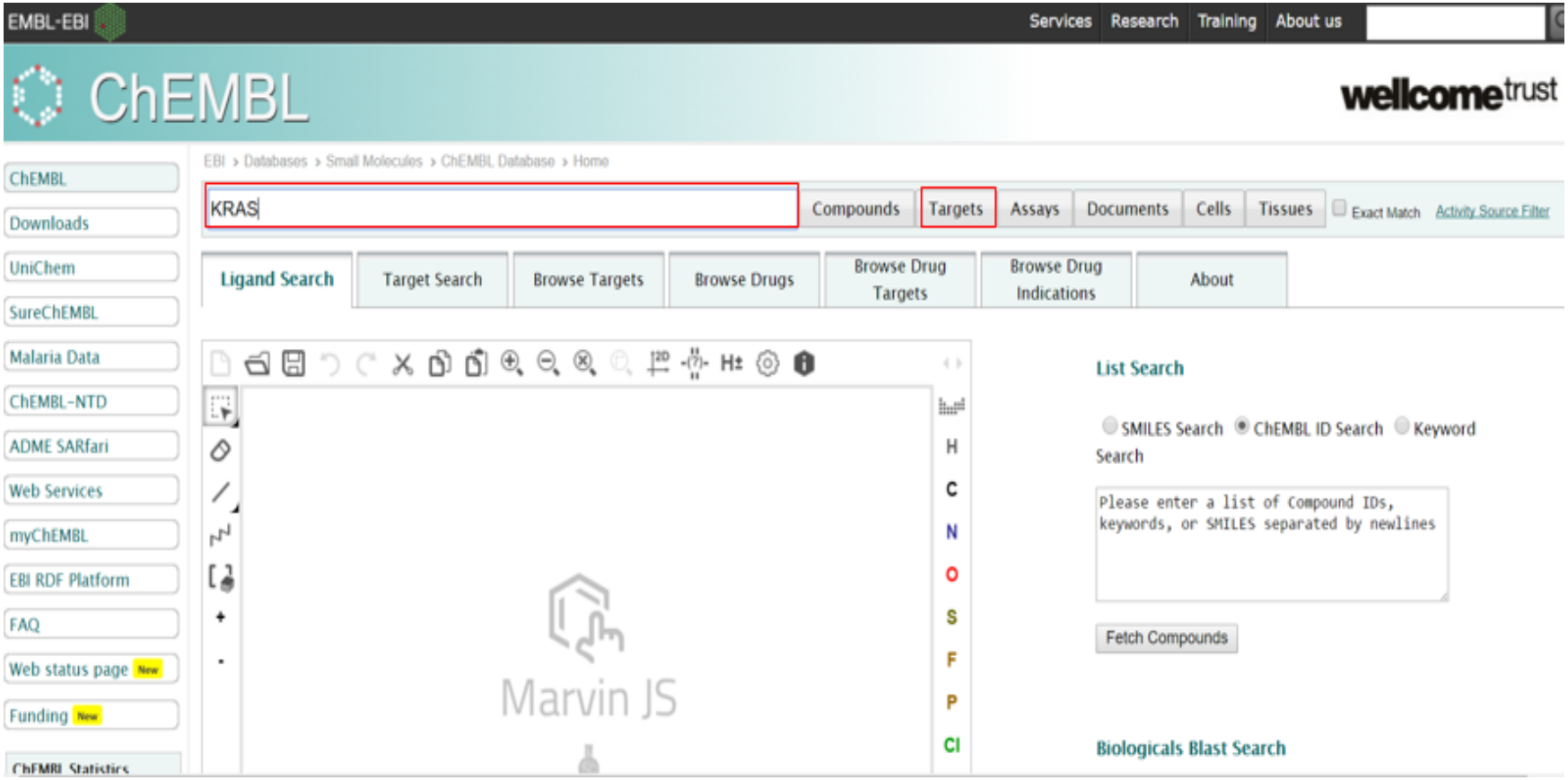
QSAR e Docking

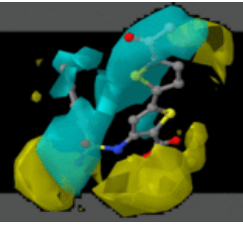


Strategia 1,2-D QSAR by www.RCMD.it






EMBL-EBI
Services
Research
Training
About us
ChEMBL
wellcome trust
ChEMBL
Downloads
UniChem
SureChEMBL
Malaria Data
ChEMBL-NTD
ADME SARfari
Web Services
myChEMBL
EBI RDF Platform
FAQ
Web status page
Funding
EBI › Databases › Small Molecules › ChEMBL Database › Home
KRAS
Compounds
Targets
Assays
Documents
Cells
Tissues
Exact Match
Activity Source Filter
Ligand Search
Target Search
Browse Targets
Browse Drugs
Browse Drug Targets
Browse Drug Indications
About
Marvin JS
List Search
SMILES Search
ChEMBL ID Search
Keyword Search
Please enter a list of Compound IDs, keywords, or SMILES separated by newlines
Fetch Compounds
Biologicals Blast Search

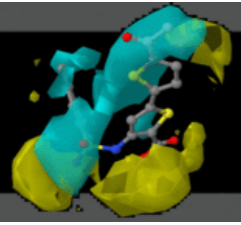


Database iniziale

by www.RCMD.it

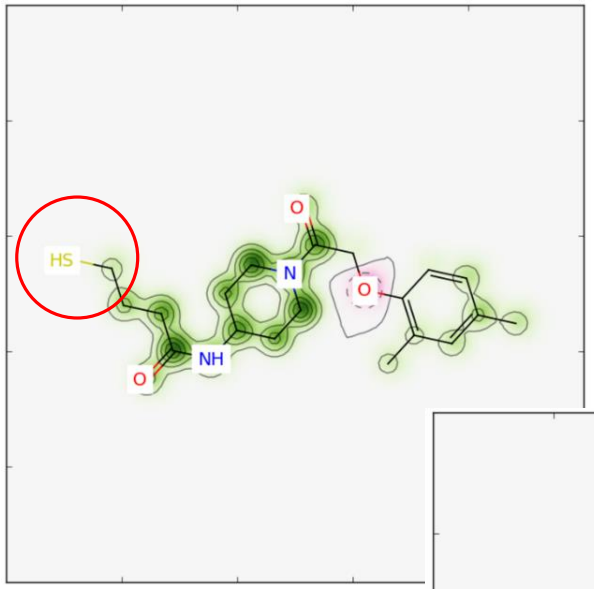
59 molecole (38 reversibili)

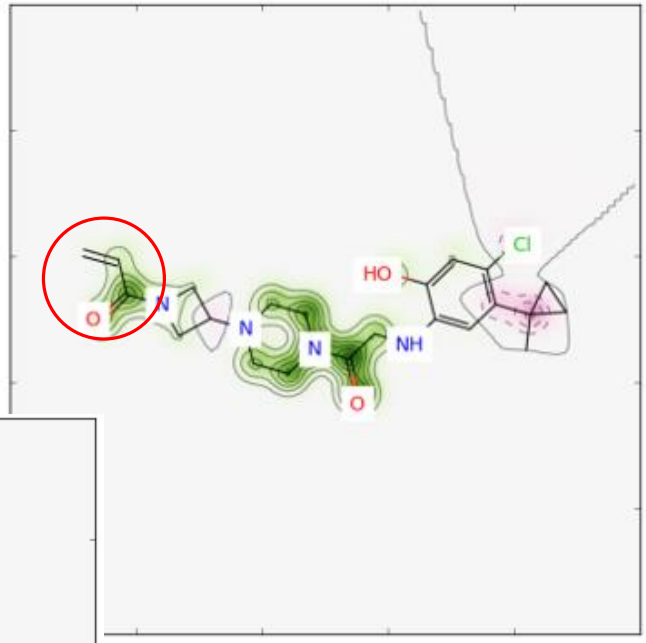
	CANONICAL_SMILES	STANDARD_TYPE	STANDARD_UNITS	pVALUE
33	OCC1OC(C(O)C(O)C1O)N(CCOc1ccc2ccccc2c1)S(=O)(=...	IC50	microM	6.15
22	O=C1CN2CCN1Cc3cc(Oc4cc(Cn5cncc5C2)ccc4C#N)ccc3...	EC50	microM	5.97
1	Nc1ccc(cc1)S(=O)(=O)N1C(=O)CCCC1	Kd	microM	5.52
0	Nc1ccc(cc1)S(=O)(=O)NC(=C)C1CC1	Kd	microM	5.22
21	O=C1CN2CCN1c3cccc4ccc(Oc5cc(Cn6cncc6C2)ccc5C#N...	EC50	microM	5.00
30	Cc1ccc(NC(=S)NNc2c(cc(cc2[N+]([O-])=O)C(F)(F)F...	IC50	microM	5.00
29	OC(=O)CC1=C(/C(=C/c2cocc2)/c2c1cc(F)cc2)C	IC50	microM	4.52
20	ONC1=CC=C(C=C1)C(=O)NC[C@@H]1CC2O[C@H](COCC3=C...	IC50	microM	4.45
32	ONc1ccc(cc1)S(=O)(=O)NCCOc1ccc2ccccc2c1	IC50	microM	4.25
18	ONC1=CC=C(C=C1)S(=O)(=O)NC[C@H]1CC2O[C@H](COCC...	IC50	microM	4.24
23	ONC1=CC=C(C=C1)C(=O)NC[C@H]1CC2O[C@H](COCC3=CC...	IC50	microM	4.21
17	ONC1=CC=C(C=C1)S(=O)(=O)NC[C@@H]1CC2O[C@H](COC...	IC50	microM	4.12
25	OC(=O)CC1=C(/C(=C/c2cscc2)/c2c1cc(Cl)cc2)C	IC50	microM	4.00
31	[O-][N+](=O)c1cc(cc(c1NNC(=S)Nc1ccc(F)cc1)[N+]...	IC50	microM	4.00
26	OC(=O)CC1=C(/C(=C/c2cocc2)/c2c1cc(cc2)C)C	IC50	microM	3.80
24	OC(=O)CC1=C(/C(=C/c2cocc2)/c2c1cc(Cl)cc2)C	IC50	microM	3.74
15	CCC(C)C(N)C(=O)NC1=CC2N=C(CC3=CNC4=CC=CC=C34)N...	Kd	microM	3.72
19	OC(=O)CC1=C(/C(=C/c2ccc(SC)cc2)/c2c1cc(F)cc2)C	IC50	microM	3.68
27	CC1=C(CC(O)=O)c2ccc(Br)cc2\C1=C/C1=CSC=C1	IC50	microM	3.66
14	CC(C)C(N)C(=O)NC1=CC2N=C(CC3=CNC4=CC=CC=C34)NC...	Kd	microM	3.57

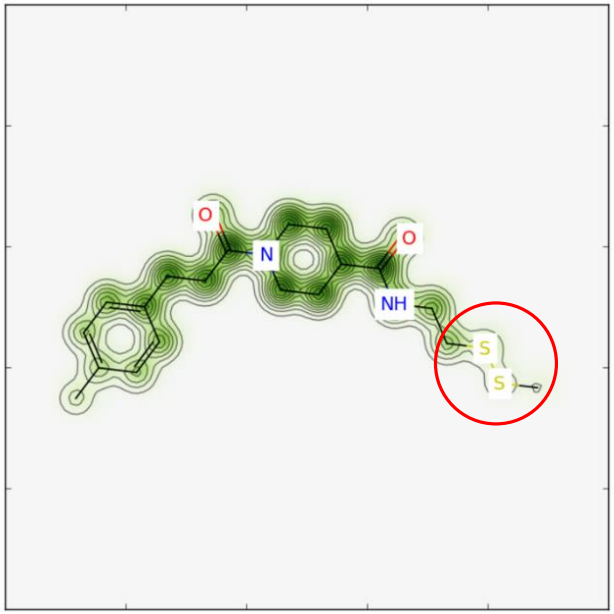


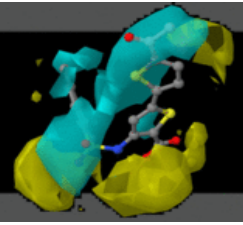
Molecole covalenti

by www.RCMD.it





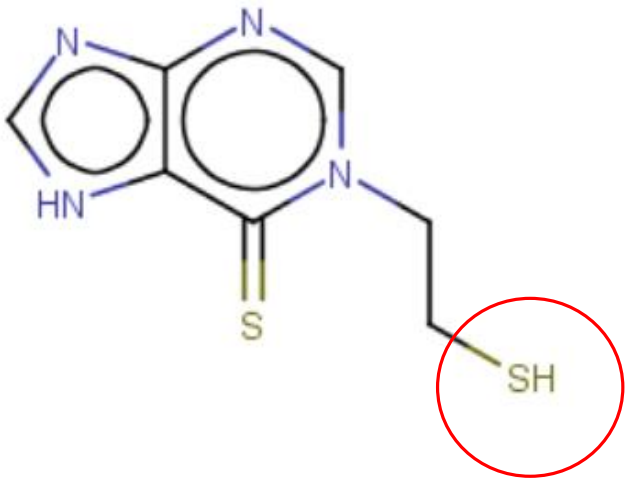




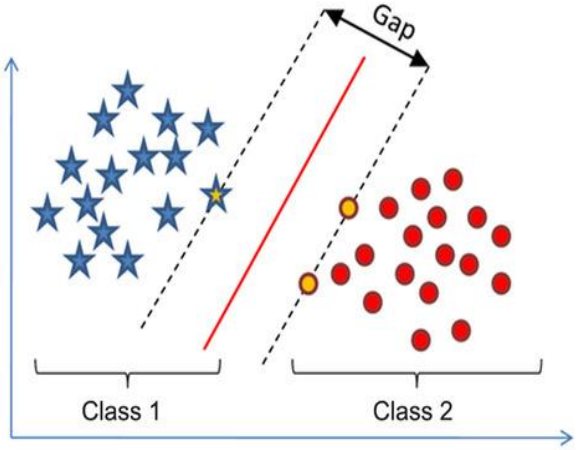
Molecole covalenti

by www.RCMD.it

	CANONICAL_SMILES_NO_SALT	NSC	[fp]median_activity	[fp]Rank	[desc]median_activity	[desc]Rank	mean_rank
0	S=c1c2[nH]cnc2ncn1CCS	404823.0	5.346527	33	6.047139	2	35.0
1	O=C(NCCS)C(=O)NCCS	120714.0	5.418880	13	5.880083	26	39.0
2	C=CC(=O)NC(=N)NC(N)=O	372157.0	5.369232	29	5.899250	14	43.0
3	C=CC(=O)NC(O)C(Cl)(Cl)Cl	22403.0	5.369427	28	5.889128	18	46.0
4	C=CC(=O)OCC(C)(C)COC(=O)C=C	97407.0	5.482500	8	5.865417	42	50.0

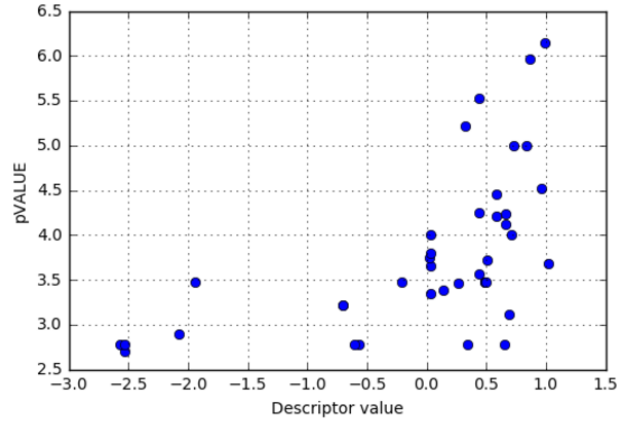


Molecole reversibili

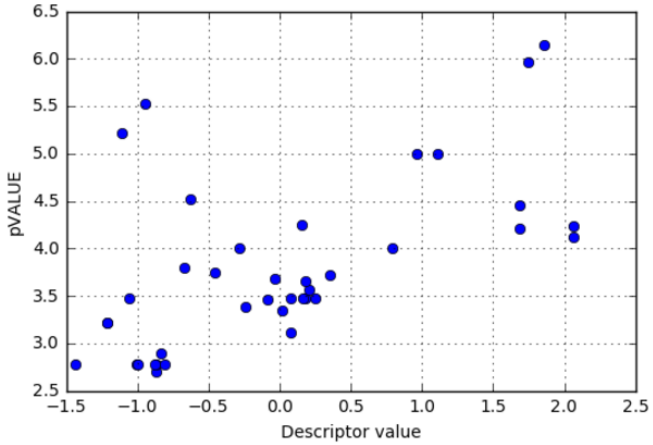


	Algorithm	q^2 K-Fold	q^2 Shufflesplit
Descriptors	Logistic Regression	0.673	0.592
Fingerprints	PassiveAggressive	0.577	0.562

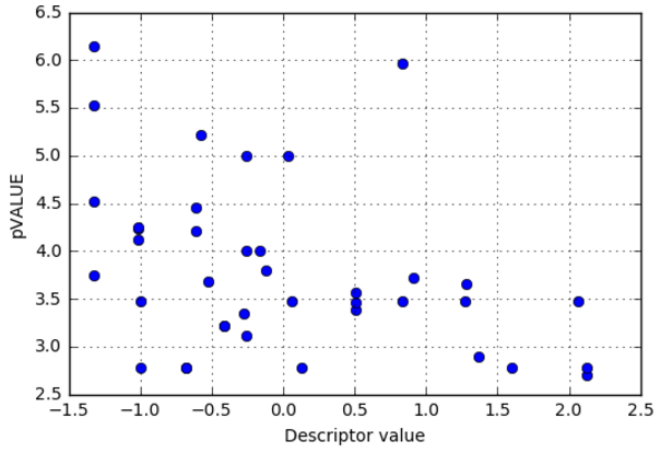
Importanza descrittori



('MaxAbsEStateIndex', <function MaxAbsEStateIndex at 0x00000140A90A1AE8>)

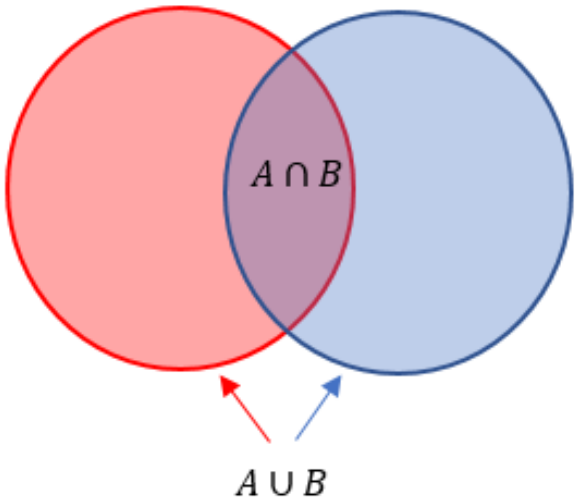


('MolWt', <function <lambda> at 0x00000140A90A1268>)

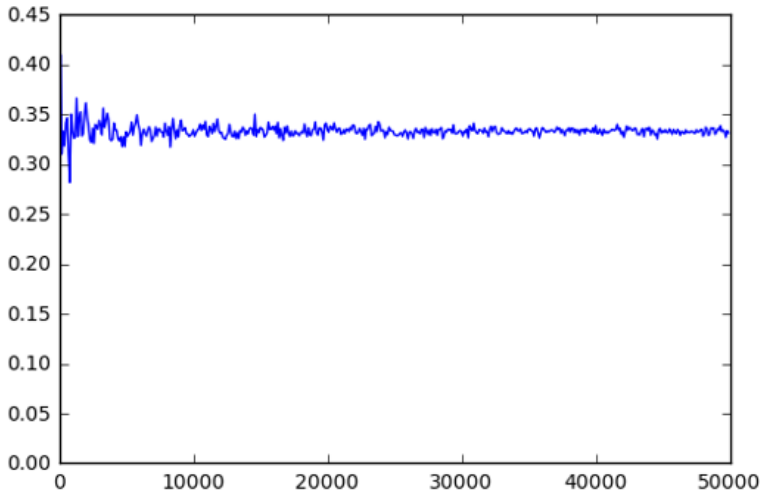


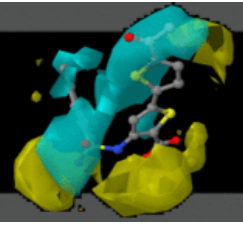
('EState_VSA8', <function _InstallDescriptors.<locals>.<lambda> at 0x00000140A91CFE18>)

$$\frac{A \cap B}{A \cup B}$$



$$\frac{A \cdot B}{|A^2| + |B^2| - A \cdot B}$$



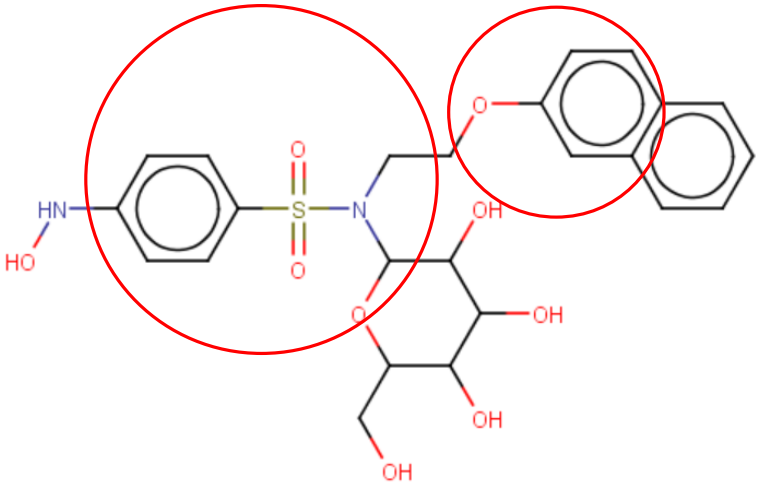


Risultati dello screening

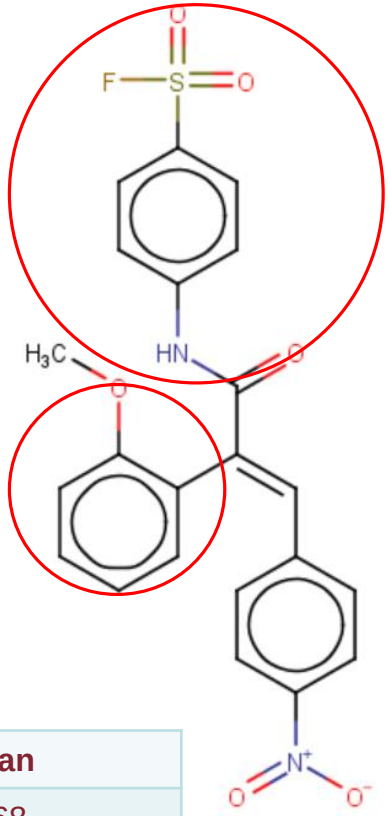
by www.RCMD.it

CANONICAL_SMILES_NO_SALT	Morgan	Atom_pair	Topological_torsion	RDKF
Nc1ccc(S(=O)(=O)N2CCCCC2)cc1	0.571429	0.750000	0.526316	0.727955
Nc1ccc(S(=O)(=O)N2CCCc3ccccc32)cc1	0.560976	0.668831	0.510638	0.362538
CCC(=N)NS(=O)(=O)c1ccc(N)cc1	0.552632	0.442029	0.566667	0.633987
N=C(NS(=O)(=O)c1ccc(N)cc1)NS(=O)(=O)c1ccc(N)cc1	0.542857	0.268085	0.447368	0.464078
NC(=O)NS(=O)(=O)c1ccc(N)cc1	0.542857	0.467742	0.607143	0.494692
Cc1ccc([N+](=O)[O-])cc1NC(=O)NNc1c([N+](=O)[O-...	0.540984	0.580000	0.750000	0.673469
Cc1ccc([N+](=O)[O-])cc1NC(=O)NNc1c([N+](=O)[O-...	0.540984	0.580000	0.750000	0.673469
Cc1ccc(S(=O)(=O)N2CCCCC2=O)cc1[N+](=O)[O-]	0.533333	0.521739	0.555556	0.666239
N=C(N)NS(=O)(=O)c1ccc(N)cc1	0.527778	0.453125	0.607143	0.500000
NC(=S)NS(=O)(=O)c1ccc(N)cc1	0.527778	0.476190	0.607143	0.493644
Nc1ccc(S(=O)(=O)NC2CCC(NS(=O)(=O)c3ccc(N)cc3)C...	0.527778	0.277372	0.319149	0.641350
Nc1ccc(S(=O)(=O)NC2CCC(NS(=O)(=O)c3ccc(N)cc3)C...	0.527778	0.277372	0.319149	0.641350
CC(=O)NS(=O)(=O)c1ccc(N)cc1	0.527778	0.456693	0.607143	0.590604
CC(=N)NS(=O)(=O)c1ccc(N)cc1	0.513514	0.449612	0.607143	0.605856
CSC(=N)NS(=O)(=O)c1ccc(N)cc1	0.512821	0.437956	0.566667	0.474206
Nc1ccc(S(=O)(=O)NC2CCC(O)CC2)cc1	0.500000	0.432927	0.384615	0.618661
Nc1ccc(S(=O)(=O)NO)cc1	0.500000	0.417391	0.482759	0.444700
Cc1cc([N+](=O)[O-])c(Nc2ccc(C)c(Cl)c2)c([N+](=...	0.490909	0.378453	0.693878	0.523636
Cc1cc([N+](=O)[O-])c(Nc2ccc(C)c(Cl)c2)c([N+](=...	0.490909	0.378453	0.693878	0.523636
COc1ccc2c(c1)CCCN2S(=O)(=O)c1ccc(N)cc1	0.489362	0.520408	0.480000	0.349908

Risultati dello screening



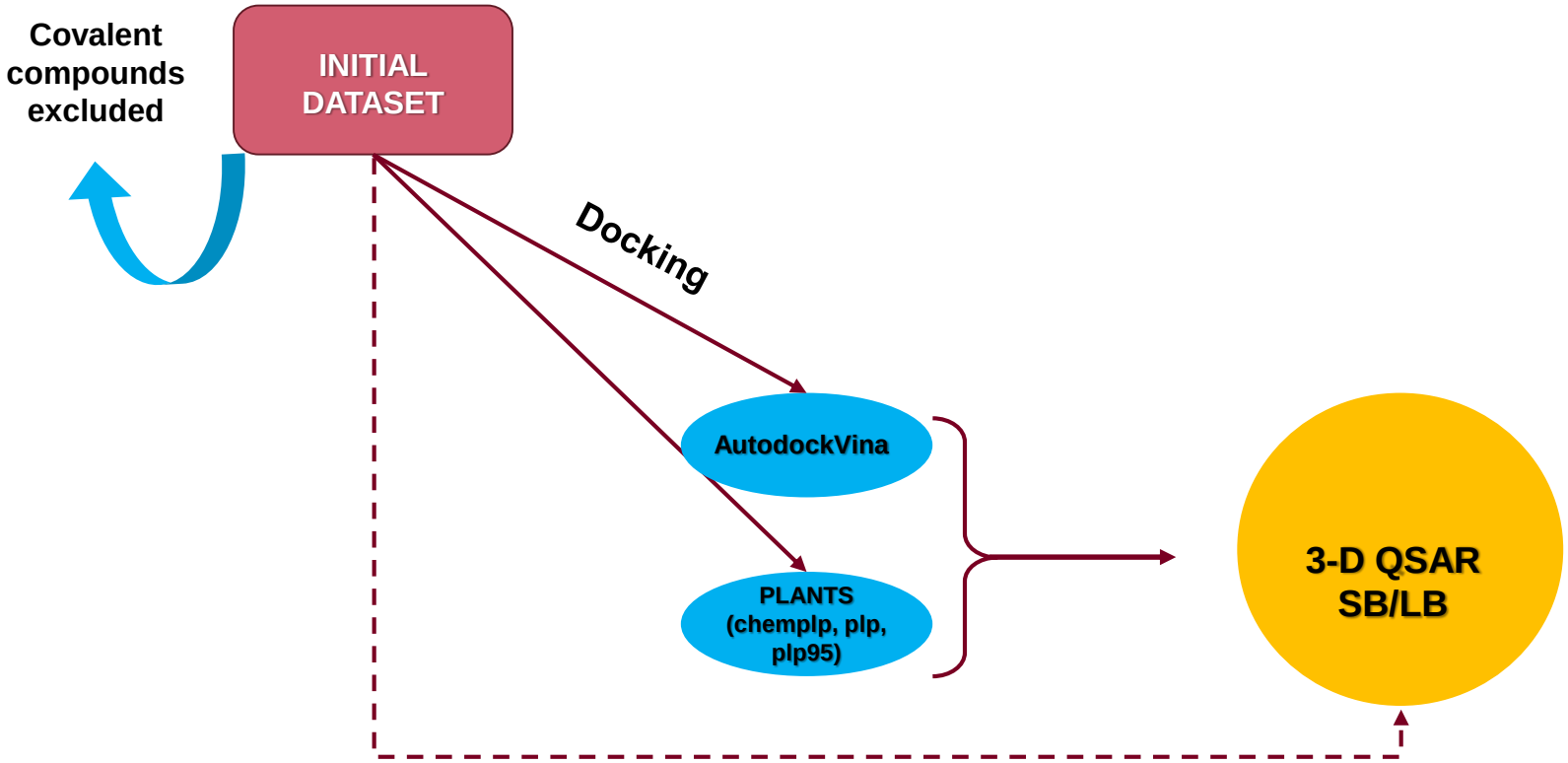
Attività (pVALUE)	6.15

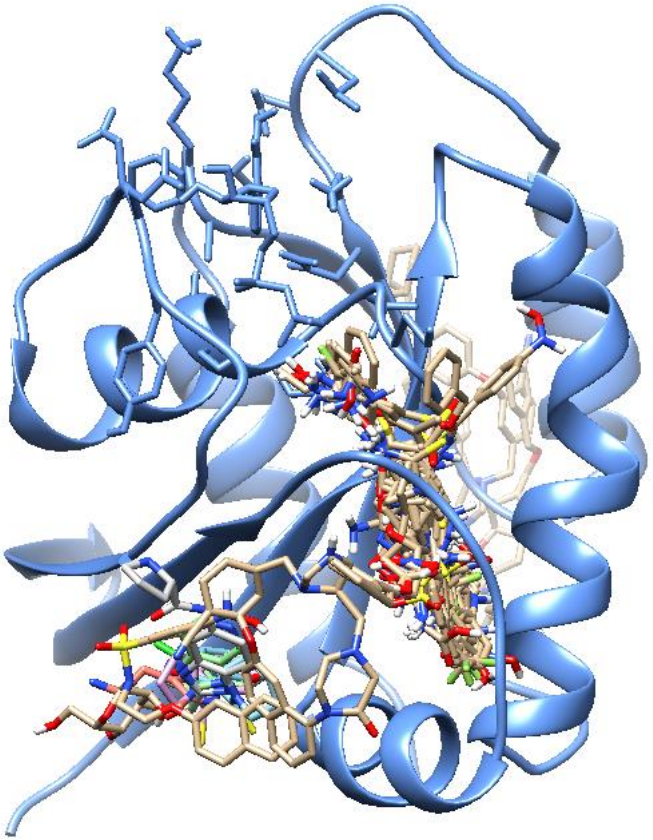


	desc	fp	Mean
Attività (pVALUE)	4.26	5.11	4.68

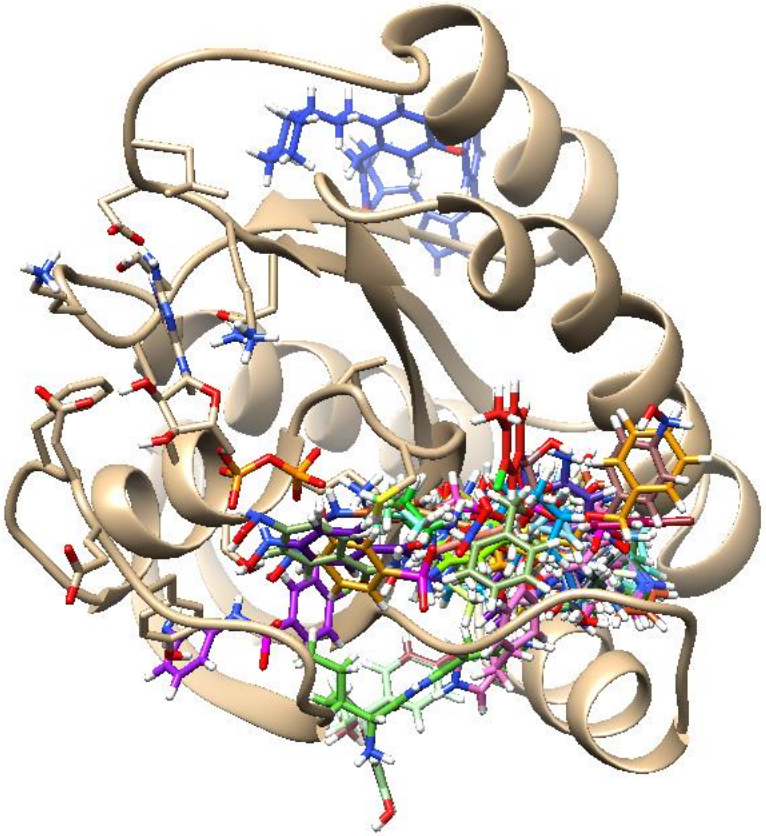
Strategia Docking e 3-D QSAR

by www.RCMD.it

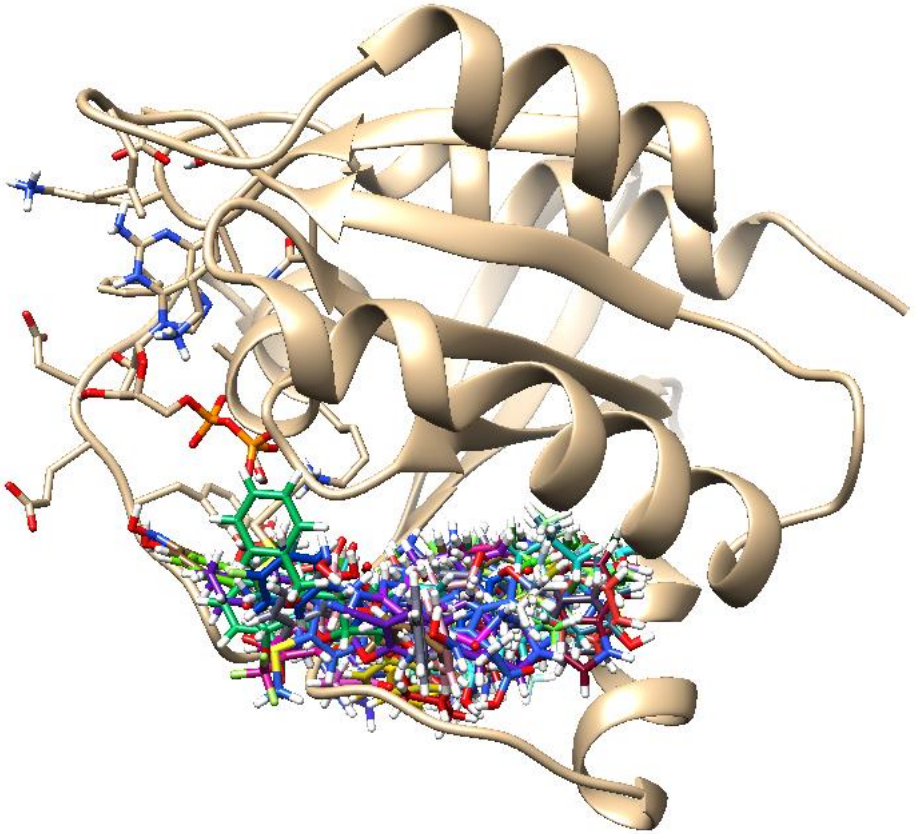




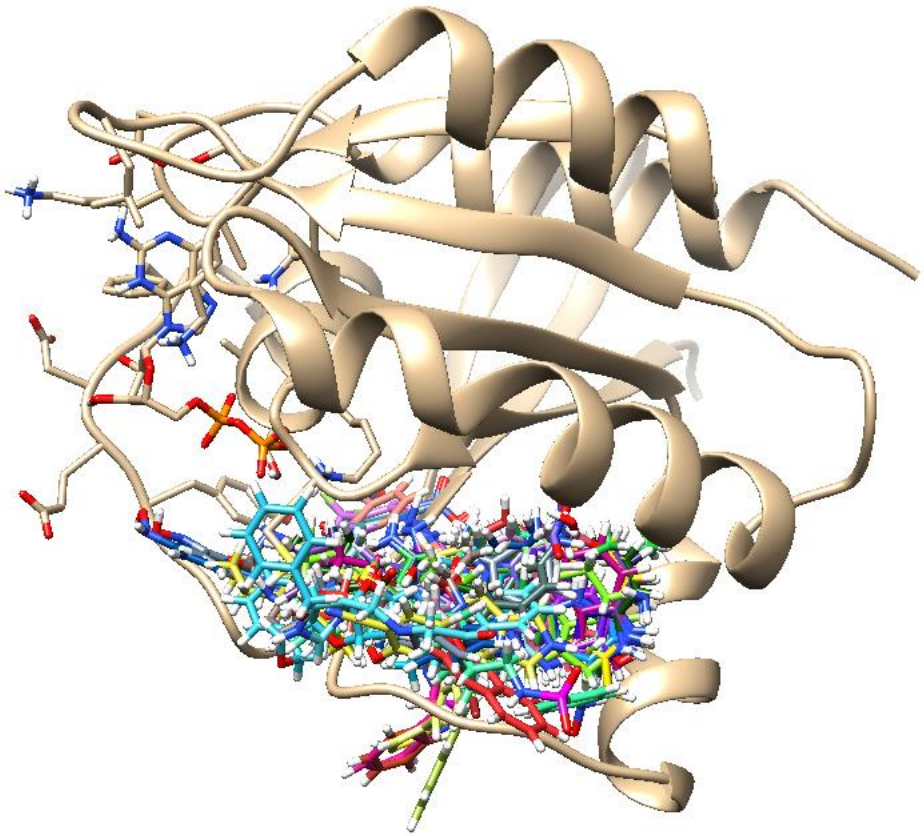
PLANTS - chemplp



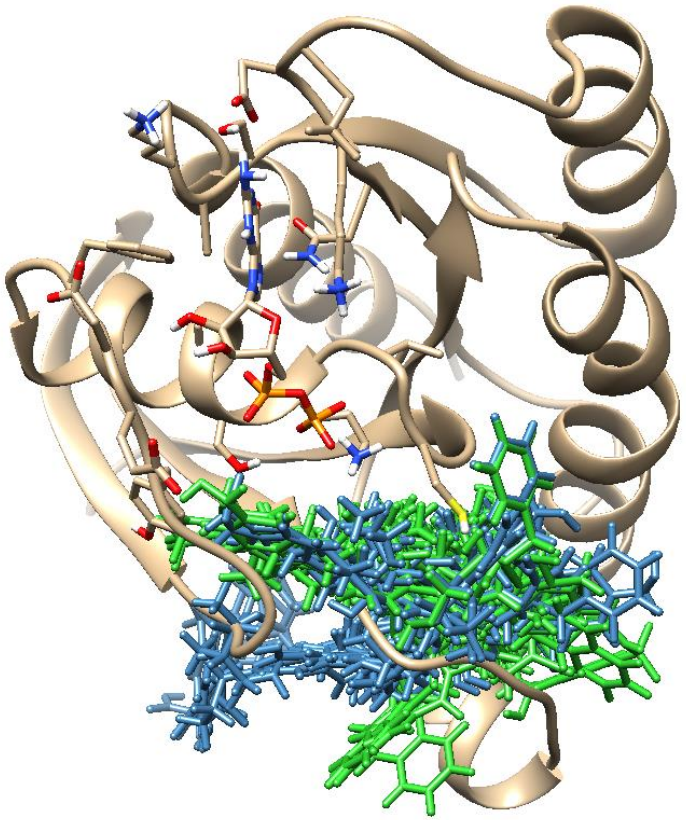
PLANTS - plp



PLANTS – plp95

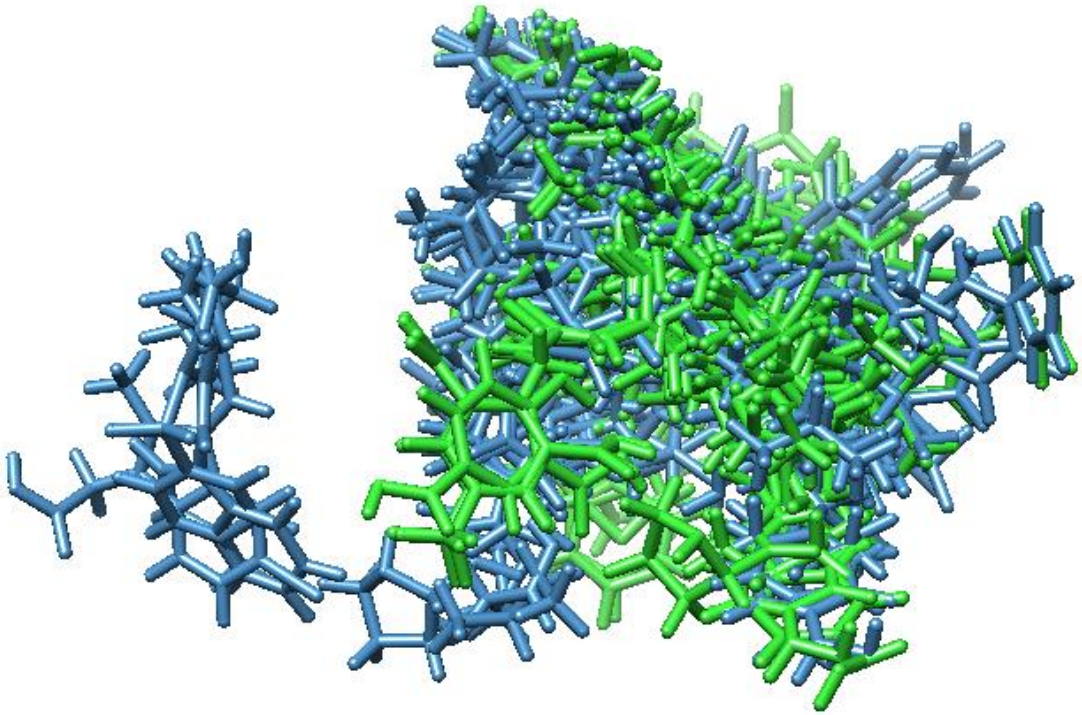


Scelta allineamento migliore



- plp95
- plp

Scelta allineamento migliore



- plp95
- plp

www.3D-QSAR.com

Home Contacts

Menu

Home

Contacts

Login

Leonardo

••••••••••••••••

Remember Me

Login

Not Registered? Sign Up

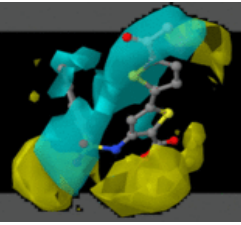
Forgot Password? Click Here

Welcome to www.3D-QSAR.com

Hi! Click the App Button you want to work on!

Our Apps

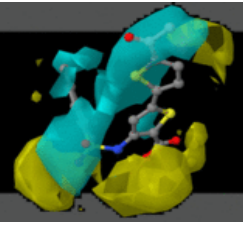
Py-MolEdit | Py-ConfSearch | Py-Align | Py-CoMFA



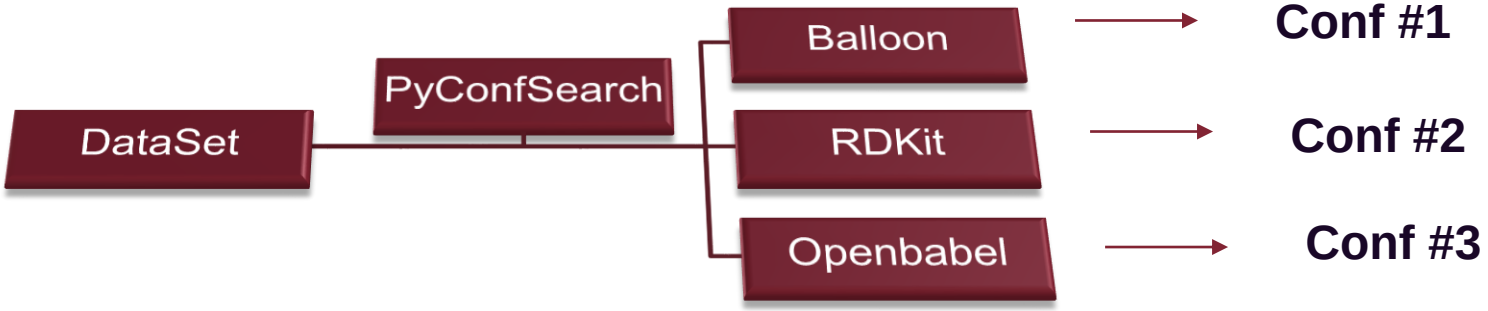
3DQSAR – StructureBased

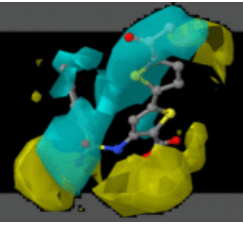
by www.RCMD.it

Probe	Q^2 ELE (PC)	Q^2 STE (PC)	Q^2 BOTH (PC)	CV
C.2	**0.037(1)**	**0.131(8)**	**0.06(2)**	**LOO**
C.2	**0.084(8)**	**0.231(8)**	**0.106(4)**	**LSO**
C.3	0.037(1)	0.08(8)	0.06(2)	LOO
C.3	0.082(7)	0.149(5)	0.071(2)	LSO
C.cat	0.037(1)	0.12(8)	0.06(2)	LOO
C.cat	0.034(1)	0.136(7)	0.062(5)	LSO
H	0.033(1)	0.12(8)	0.056(2)	LOO
H	0.042(4)	0.256(8)	0.054(2)	LSO
N3	0.036(1)	0.117(8)	0.059(2)	LOO
N3	0.053(1)	0.154(8)	0.1(4)	LSO
O.3	0.036(1)	0.093(8)	0.06(2)	LOO
O.3	0.088(1)	0.138(8)	0.091(4)	LSO



3DQSAR – LigandBased

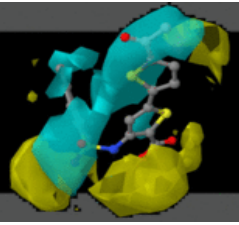




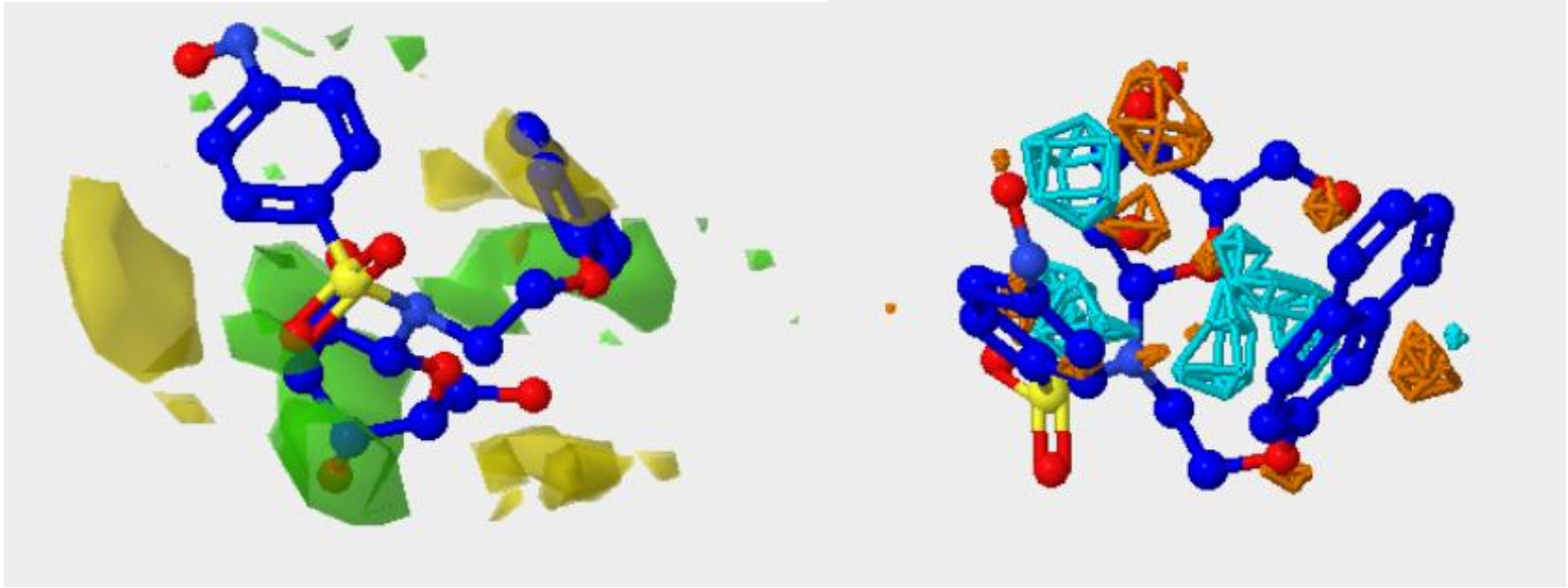
1603_kras_least_active

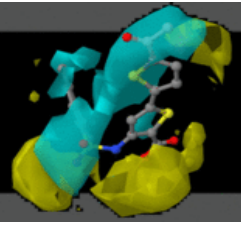
by www.RCMD.it

Probe	Q2 ELE (PC)	Q2 STE (PC)	Q2 BOTH (PC)	CV
C.3	**0.548(8)**	**0.032(2)**	**0.524(7)**	**LOO**
C.cat	0.548(8)	0.032(2)	0.524(7)	LOO
N.3	0.551(8)	0.032(2)	0.521(7)	LOO
O.3	0.550(8)	0.032(2)	0.521(7)	LOO
H	0.538(6)	0.032(2)	0.513(7)	LOO
C.2	0.548(8)	0.032(2)	0.524(7)	LOO
C.3	**0.553(8)**	**0.018(2)**	**0.507(6)**	**LSO**
C.cat	0.535(6)	0.034(2)	0.474(6)	LSO
N.3	0.518(7)	0.027(2)	0.5(7)	LSO
O.3	0.512(6)	0.03(2)	0.496(6)	LSO
H	0.509(7)	0.027(2)	0.491(7)	LSO
C.2	0.512(8)	0.047(2)	0.479(8)	LSO

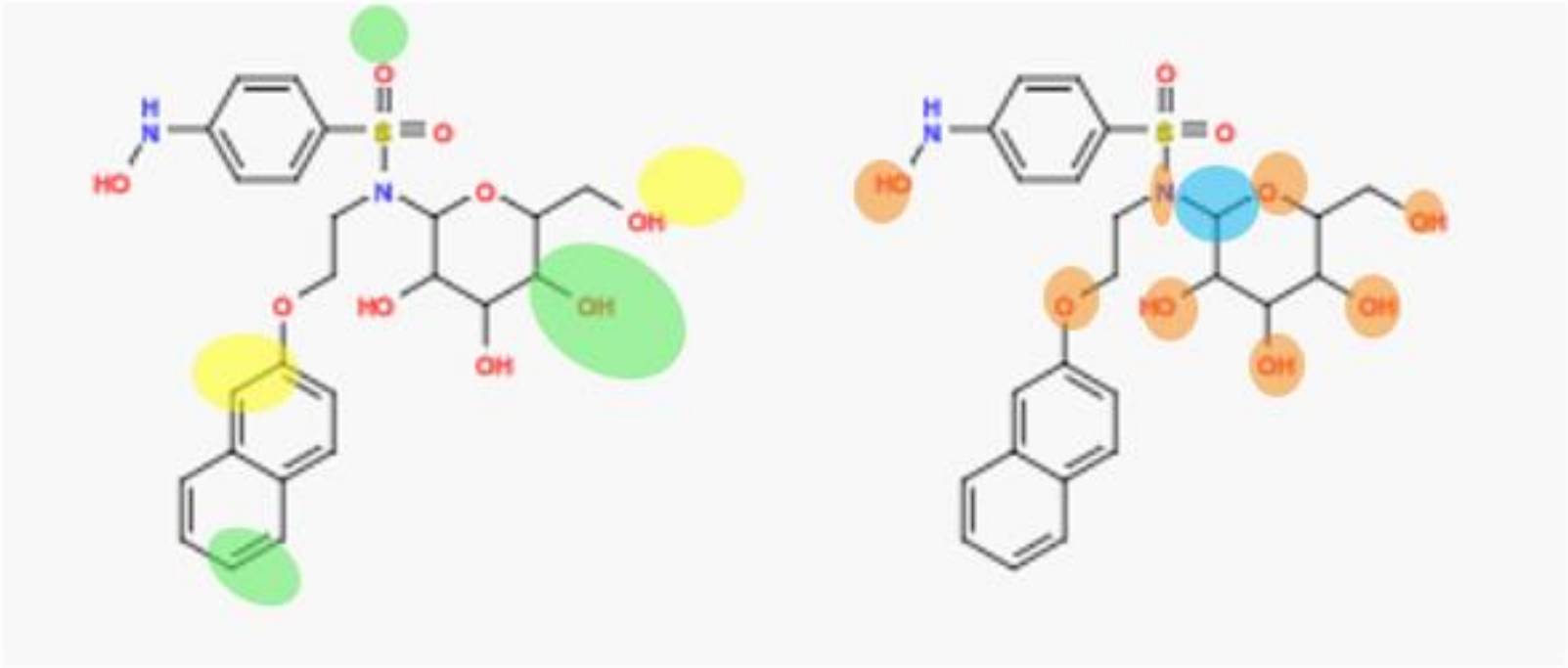


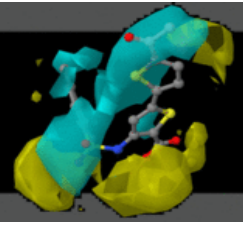
Mappe CoMFA





Mappe CoMFA





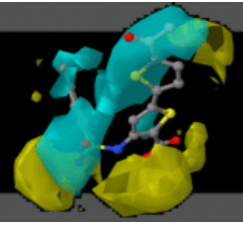
CONCLUSIONI

- Modello 3DQSAR SB: non sufficientemente robusto (q^2 <0.3)
- Modello 3DQSAR LB: molto buono (q^2 >0.5)
- Docking non abbastanza accurato

- 1,2-D QSAR:

✓ Evidenziata correlazione dei singoli descrittori

✓ **Arricchimento del training set con i risultati dello screening**



Grazie per l'attenzione.