Studio di modelli farmacoforici mediante l'uso del linguaggio di programmazione *Python*. Applicazione a serie di inibitori della chinasi PKC Theta

FACOLTÀ DI FARMACIA E MEDICINA
CORSO DI LAUREA IN BIOTECNOLOGIE FARMACEUTICHE

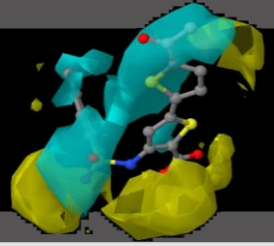
SAPIENZA
UNIVERSITÀ DI ROMA

Tesi Sperimentale
in
Chimica Farmaceutica (CHIM08)

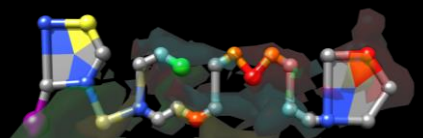
CANDIDATO:
Stefano Masiello (Matr. 1469704)

RELATORE:
Prof. Rino Ragno

ANNO ACCADEMICO
2019/2020



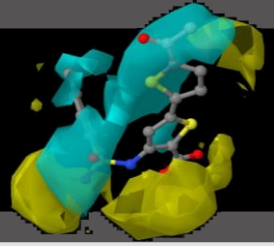
SCOPO DEL LAVORO



Generazione di modelli farmacoforici mediante RDKit e LigandScout

Confronto dei risultati ottenuti mediante i due metodi

Generazione di modelli 3-D QSAR attraverso la piattaforma www.3d-qsar.com



FARMACOFORO

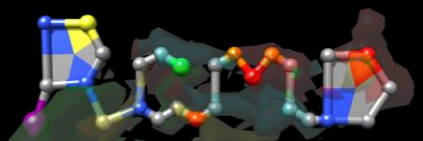
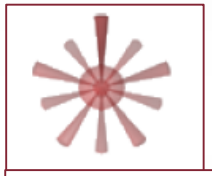


Figura geometrica virtuale ai cui vertici sono posizionate diverse caratteristiche farmacoforiche, legate mediante regole geometriche.

PRINCIPALI CARATTERISTICHE FARMACOFORICHE

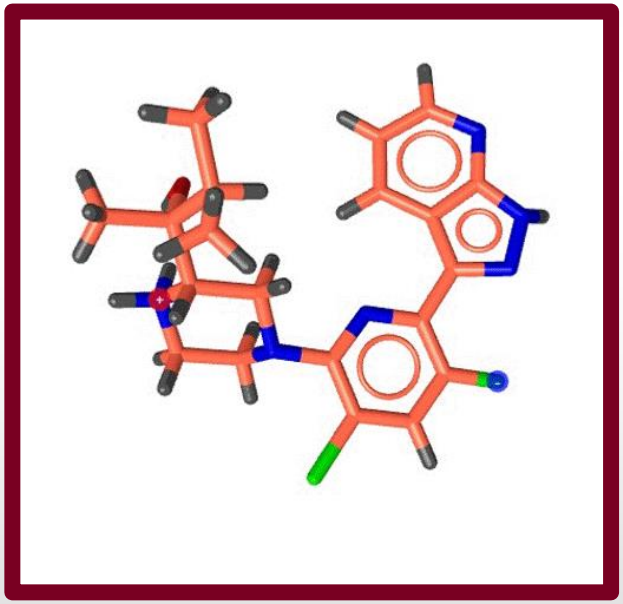
Centroide Idrofobico

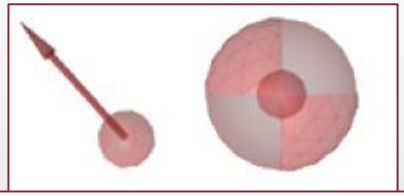
Anello Aromatico



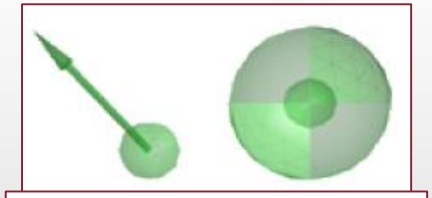
Area Ionizzabile Negativa

Area Ionizzabile Positiva

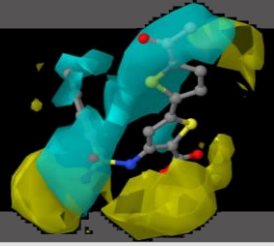




Accettore Legame–H



Donatore Legame–H



FARMACOFORO

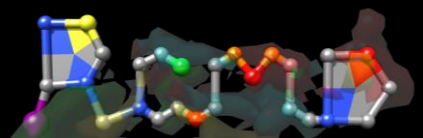
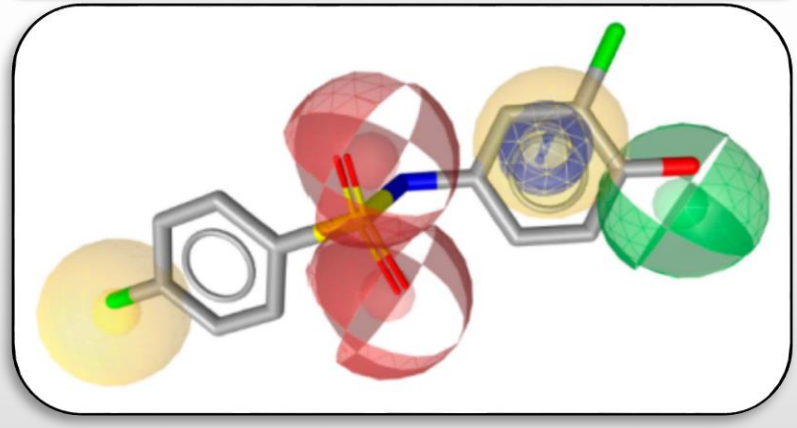


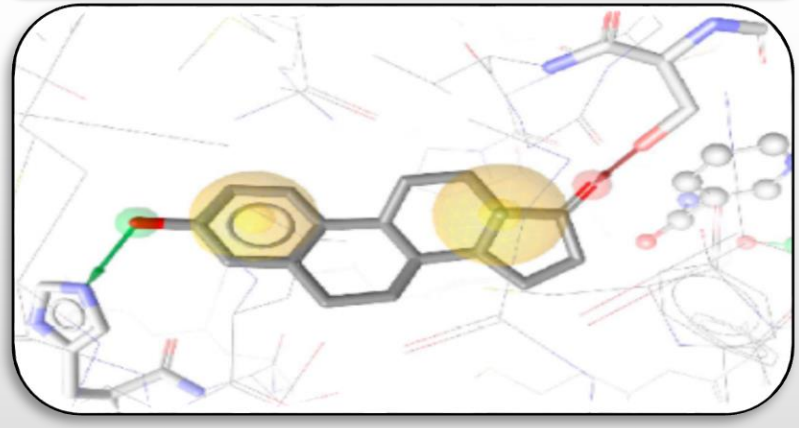
Figura geometrica virtuale ai cui vertici sono posizionate diverse caratteristiche farmacoforiche, legate mediante regole geometriche.

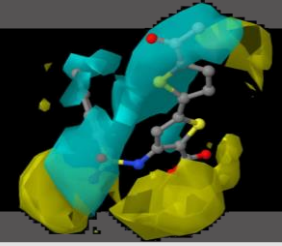
APPROCCI DI MODELING FARMACOFORICO

LIGAND-BASED PHARMACOPHORE MODELING

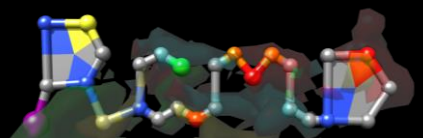


STRUCTURE-BASED PHARMACOPHORE MODELING





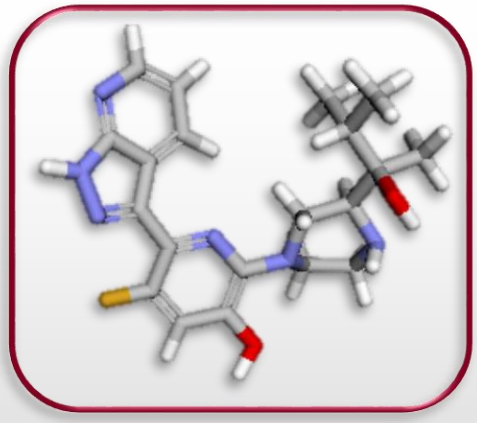
PROCEDIMENTO



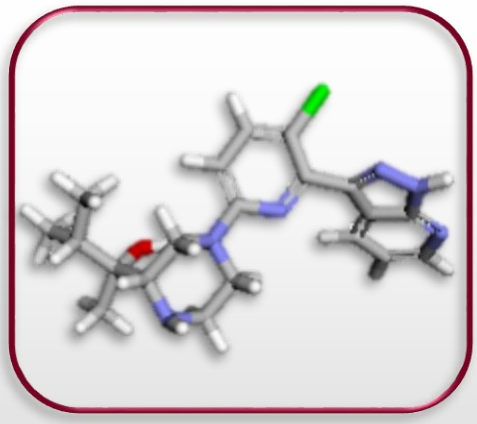
1. • Creazione di un dataset di molecole
2. • Analisi conformazionale
3. • Allineamento molecolare
4. • Generazione farmacoforo

1. CREAZIONE DATASET

ID MOLECOLA	FORMULA SMILES
A_01	C1CN(CC[NH2+]1)c1cccc(n1)-c1n[nH]c2ncccc12
A_18	CC(C)[C@@](C)(O)[C@@H]1CN(CC[NH2+]1)c1ccc(Cl)c(n1)-c1n[nH]c2ncccc12
A_20	CC(C)[C@@](C)(O)[C@@H]1CN(CC[NH2+]1)c1nc(-c2n[nH]c3ncccc23)c(F)cc1O



A_01

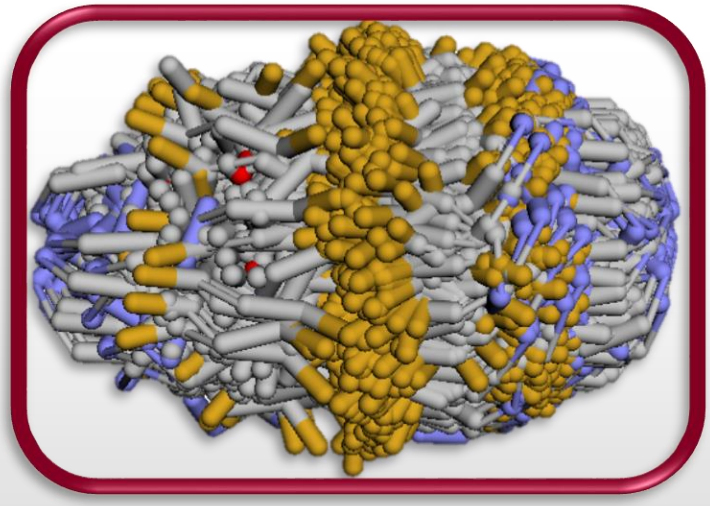


A_18

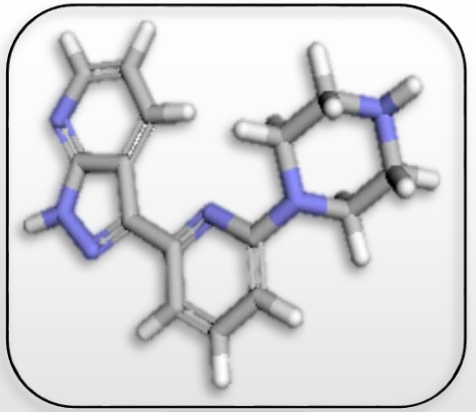
A_20

2. ANALISI CONFORMAZIONALE

ID MOLECOLA	FORMULA SMILES
A_01	C1CN(CC[NH2+]1)c1cccc(n1)-c1n[nH]c2ncccc12
A_18	CC(C)[C@@](C)(O)[C@@H]1CN(CC[NH2+]1)c1ccc(Cl)c(n1)-c1n[nH]c2ncccc12
A_20	**CC(C)[C@@](C)(O)[C@@H]1CN(CC[NH2+]1)c1nc(-c2n[nH]c3ncccc23)c(F)cc1O**

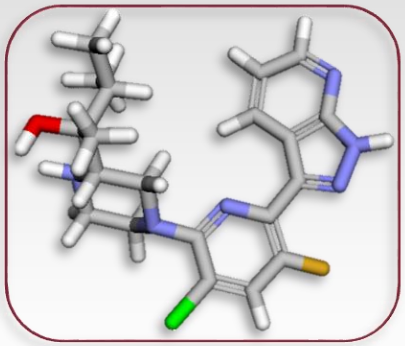


1000 conformazioni

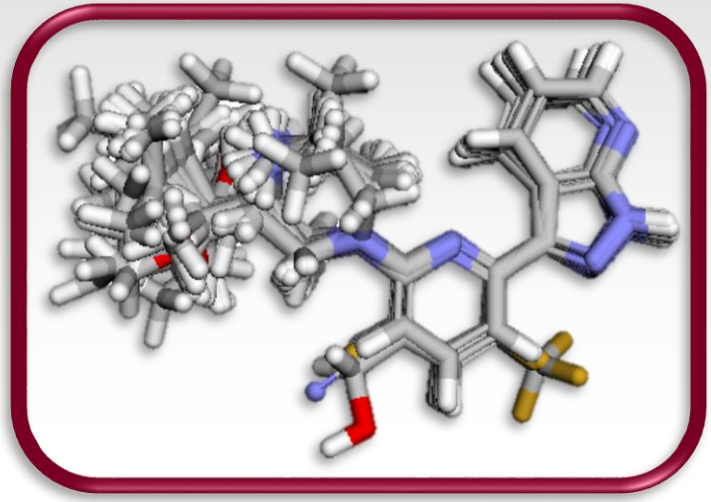


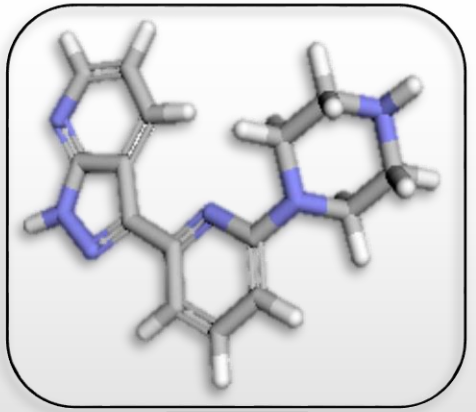
A_20

3. ALLINEAMENTO MOLECOLARE

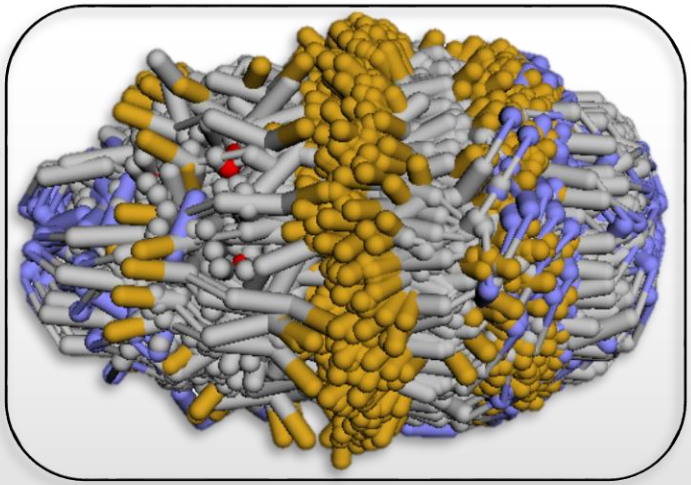


Molecola di riferimento (A_24)

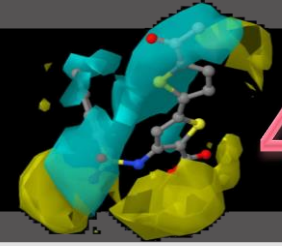




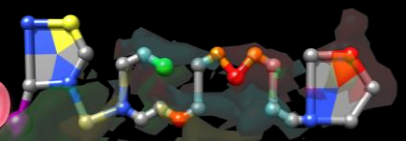
A_20



1000 conformazioni



4. GENERAZIONE FARMACOFORO



Estrazione delle caratteristiche farmacoforiche

INTERAZIONI IDROFOBICHE (HB)

INTERAZIONI AROMATICHE (ARO)

ACCETTORE / DONATORE LEGAME-H (HBA / HBD)

AREA IONIZZABILE POSITIVA / NEGATIVA (POS / NEG)

Clusterizzazione delle caratteristiche farmacoforiche

K-Means

N° CLUSTER PER TIPOLOGIA DI CARATTERISTICA

SPAZIO MOLECOLARE DELLE CARATTERISTICHE

Generazione del modello farmacoforico finale

4. GENERAZIONE FARMACOFORO

Estrazione delle caratteristiche farmacoforiche

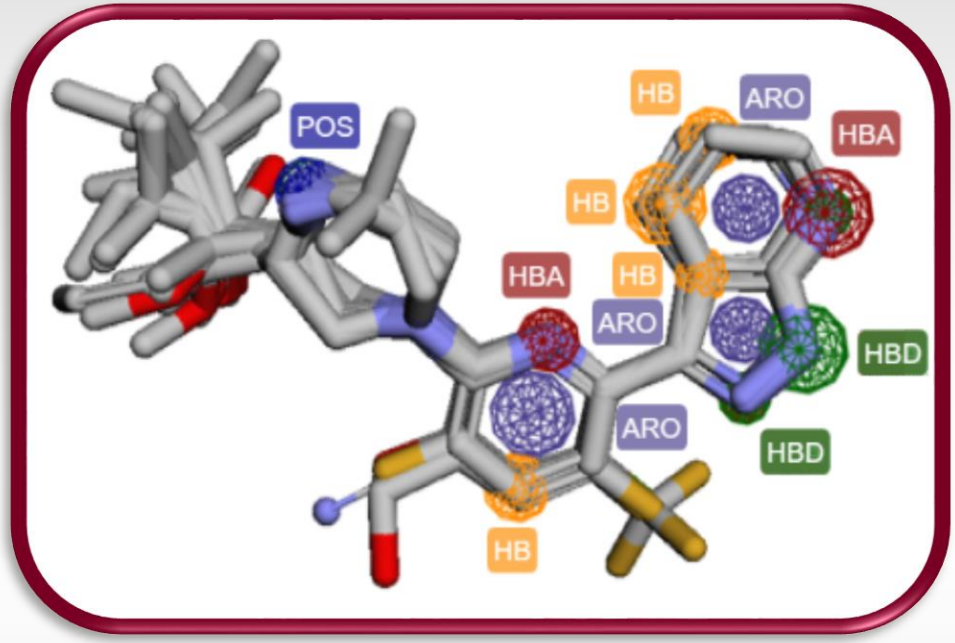
Clusterizzazione delle caratteristiche farmacoforiche

Generazione del modello farmacoforico finale

File .pdb

```
FINAL MODEL
ATOM      0 HBD  HBD     1      -1.985   1.510  -0.733
ATOM      1 HBD  HBD     1       5.788  -1.065   1.572
ATOM      2 HBD  HBD     1       6.621  -1.265   0.535
ATOM      3 HBD  HBD     1       6.610  -1.487  -1.861
ATOM      4 HBA  HBA     1       2.318  -0.155   1.139
ATOM      5 HBA  HBA     1       5.788  -1.065   1.572
ATOM      6 HBA  HBA     1       6.610  -1.487  -1.861
ATOM      7  HB   HB     1       2.128  -0.766   3.837
ATOM      8  HB   HB     1       4.371  -1.286  -2.773
ATOM      9  HB   HB     1       3.809  -1.096  -1.510
ATOM     10  HB   HB     1       4.670  -1.100  -0.410
ATOM     11 ARO  ARO     1       5.533  -1.137   0.407
ATOM     12 ARO  ARO     1       2.223  -0.458   2.482
ATOM     13 ARO  ARO     1       5.201  -1.289  -1.685
ATOM     14 POS  POS     1      -1.985   1.510  -0.733
ENDMDL
END
```

4. GENERAZIONE FARMACOFORO



- *INTERAZIONI IDROFOBICHE - HB (GIALLO)*
- *INTERAZIONI AROMATICHE - ARO (VIOLETTO)*
- *ACCETTORE LEGAME-H - HBA (ROSSO)*
- *DONATORE LEGAME-H - HBD (VERDE)*
- *AREA IONIZZABILE POSITIVA - POS (BLU)*

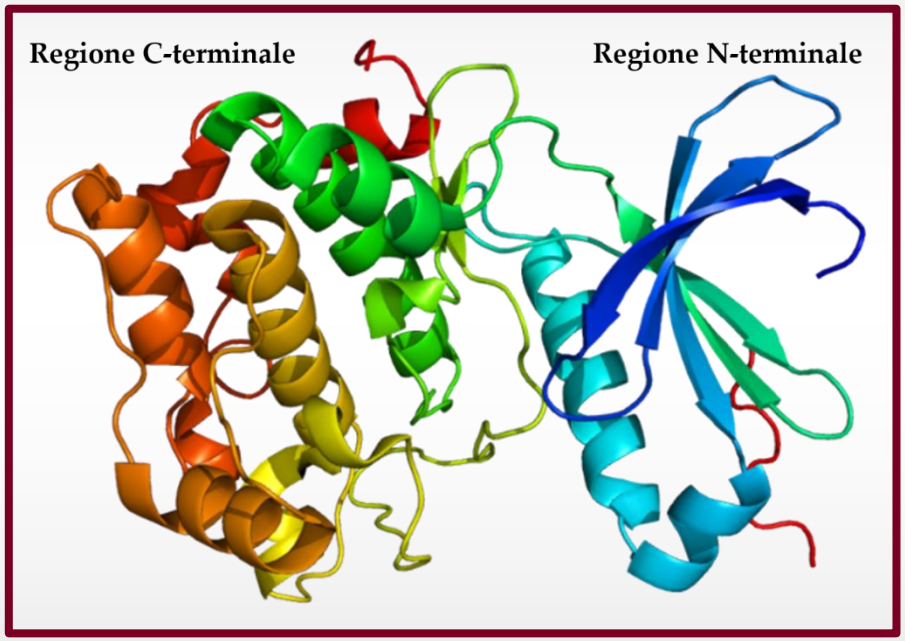
Generazione del modello farmacoforico finale

File .pdb

```
FINAL MODEL
ATOM      0 HBD  HBD     1      -1.985    1.510   -0.733
ATOM      1 HBD  HBD     1       5.788   -1.065    1.572
ATOM      2 HBD  HBD     1       6.621   -1.265    0.535
ATOM      3 HBD  HBD     1       6.610   -1.487   -1.861
ATOM      4 HBA  HBA     1       2.318   -0.155    1.139
ATOM      5 HBA  HBA     1       5.788   -1.065    1.572
ATOM      6 HBA  HBA     1       6.610   -1.487   -1.861
ATOM      7  HB   HB     1       2.128   -0.766    3.837
ATOM      8  HB   HB     1       4.371   -1.286   -2.773
ATOM      9  HB   HB     1       3.809   -1.096   -1.510
ATOM     10  HB   HB     1       4.670   -1.100   -0.410
ATOM     11 ARO  ARO     1       5.533   -1.137    0.407
ATOM     12 ARO  ARO     1       2.223   -0.458    2.482
ATOM     13 ARO  ARO     1       5.201   -1.289   -1.685
ATOM     14 POS  POS     1      -1.985    1.510   -0.733
ENDMDL
END
```

APPLICAZIONE: PKC-θ

Catalizza le reazioni di fosforilazione mediando il trasferimento del fosfato gamma di una molecola di ATP sui gruppi idrossilici di serina, treonina o tirosina.

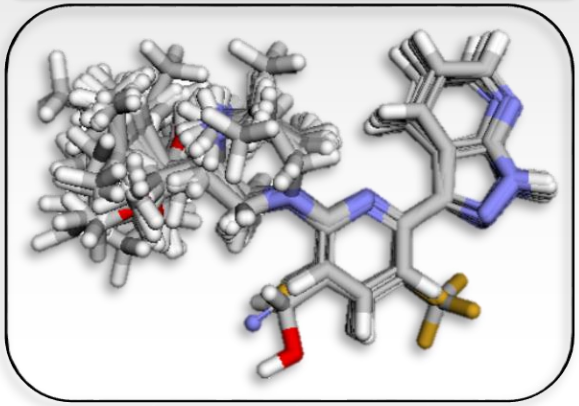


- **FAMIGLIA:** PROTEIN CHINASI C
- **SOTTOFAMIGLIA:** PKC "NUOVE"
- **ISOFORMA:** THETA θ
- **PRINCIPALE LOCALIZZAZIONE:** CELLULE EMATOPOIETICHE

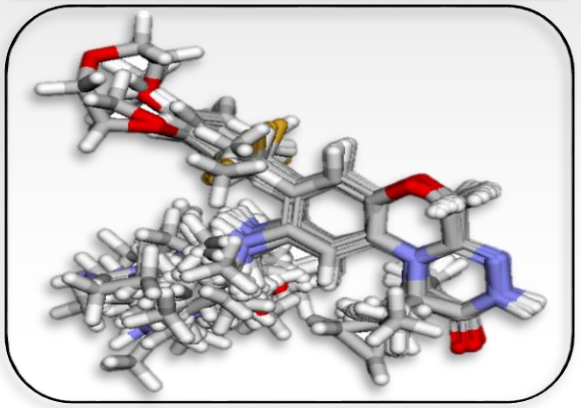
CODICE PDB: 1xjd

ALLINEAMENTI RDKit

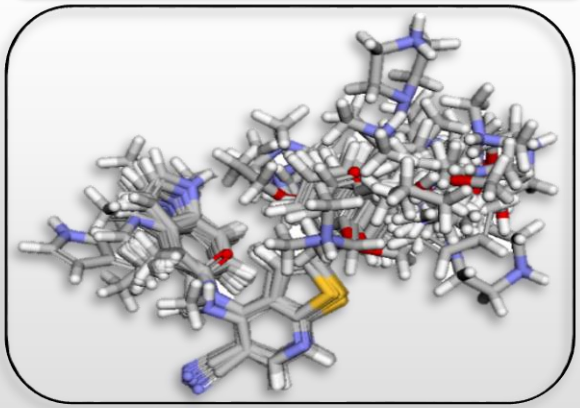
DATASET A



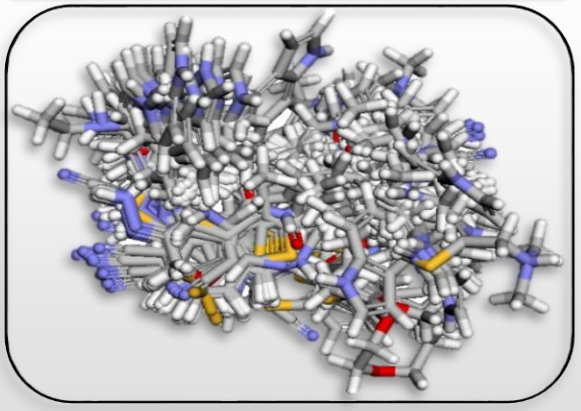
DATASET B



DATASET C

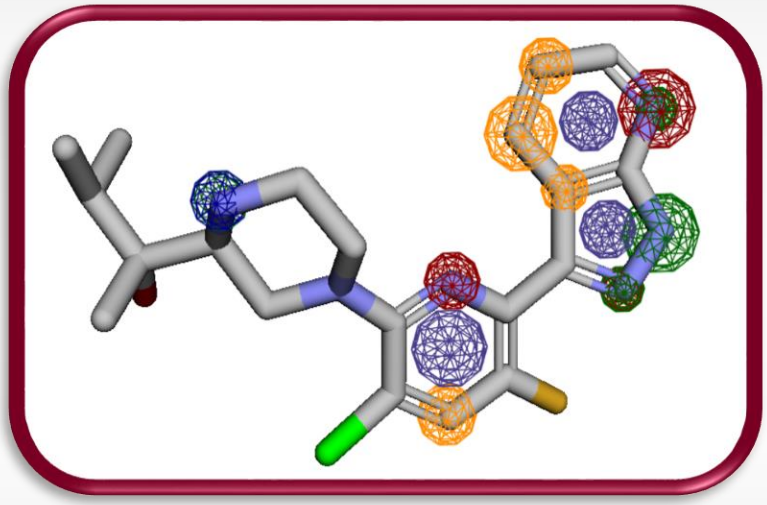


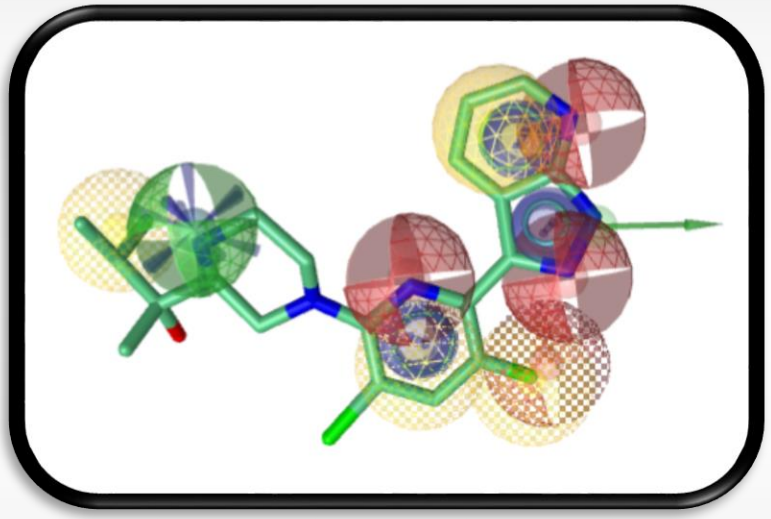
DATASET UNICO



RDKit *vs* LIGANDSCOUT

RDKit	DATASET A	LigandScout (LS)
15 CARATTERISTICHE		14 CARATTERISTICHE



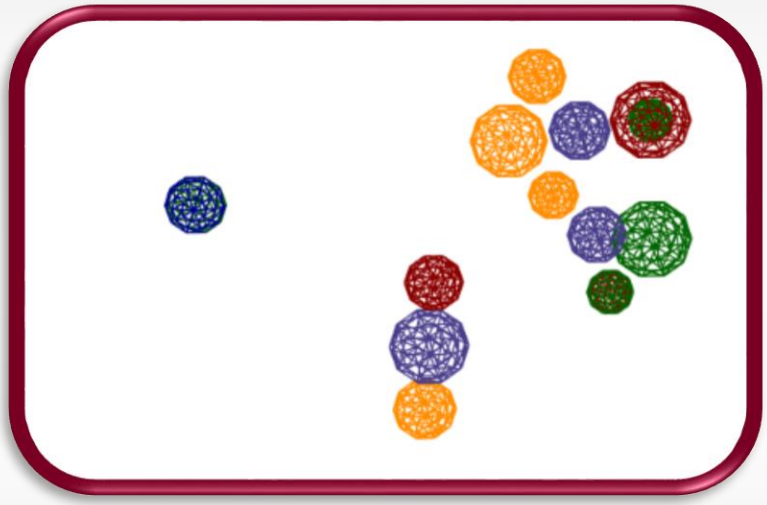


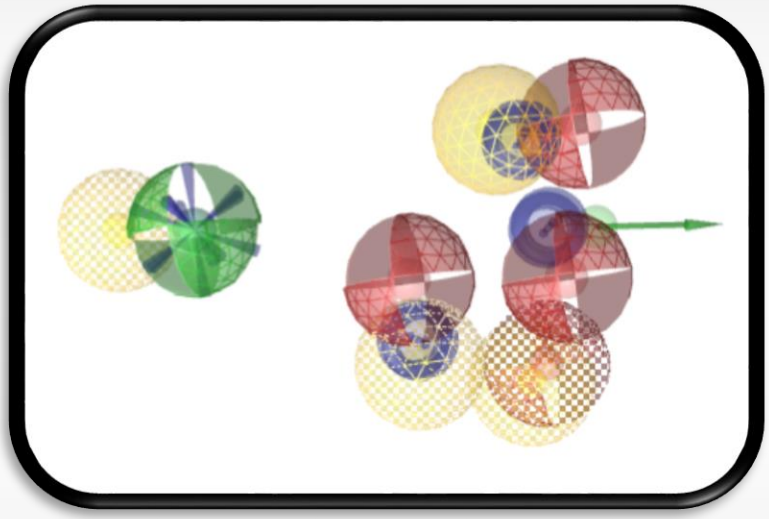
4	INTERAZIONI IDROFOBICHE (GIALLO)	4
3	INTERAZIONI AROMATICHE (VIOLETTO)	3
1	AREA IONIZZABILE POSITIVA (BLU)	1
3	ACCETTORE LEGAME-H (ROSSO)	4
4	DONATORE LEGAME-H (VERDE)	2

RDKit *vs* LIGANDSCOUT

DATASET A

RDKit	LigandScout (LS)
15 CARATTERISTICHE	14 CARATTERISTICHE

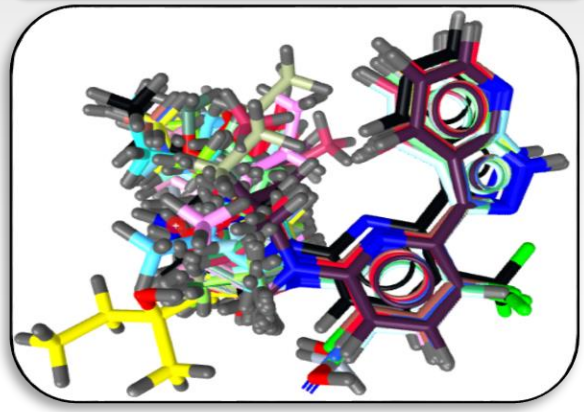




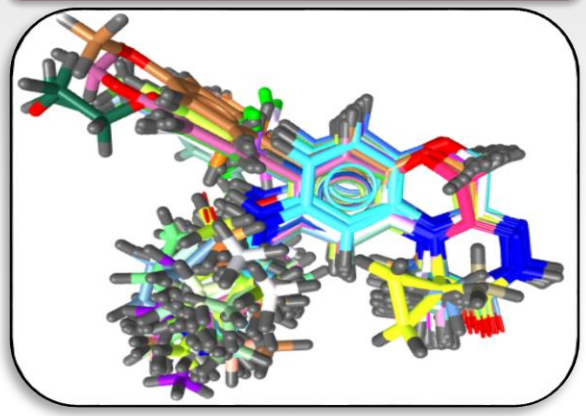
4	INTERAZIONI IDROFOBICHE (GIALLO)	4
3	INTERAZIONI AROMATICHE (VIOLETTO)	3
1	AREA IONIZZABILE POSITIVA (BLU)	1
3	ACCETTORE LEGAME-H (ROSSO)	4
4	DONATORE LEGAME-H (VERDE)	2

ALLINEAMENTI LS

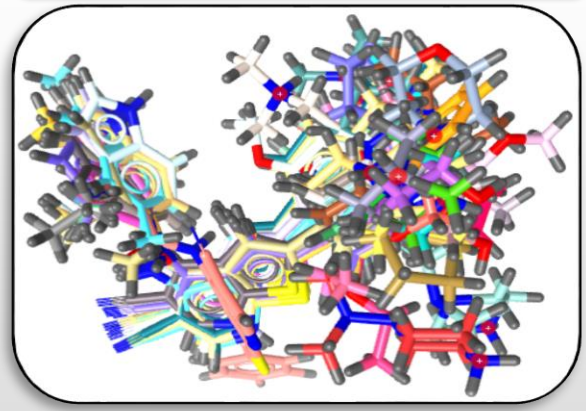
DATASET A



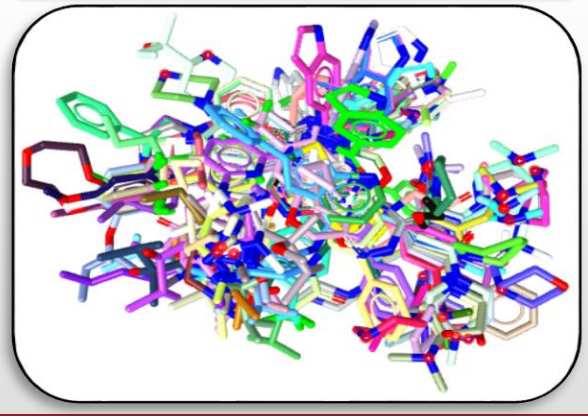
DATASET B



DATASET C

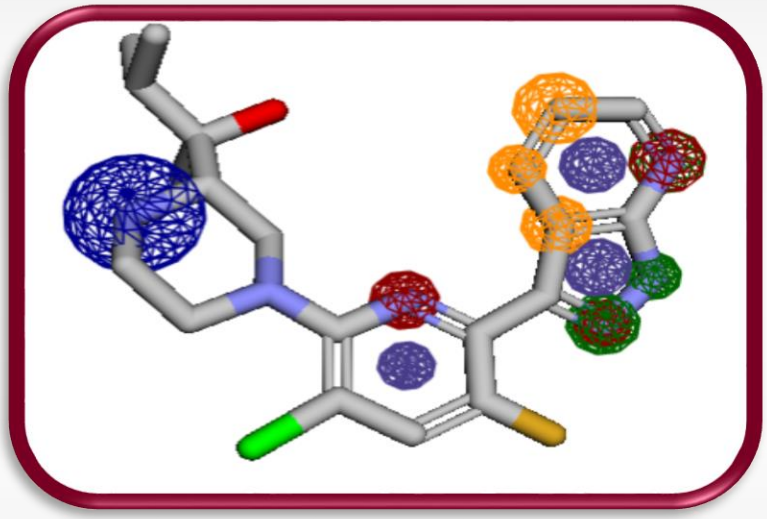


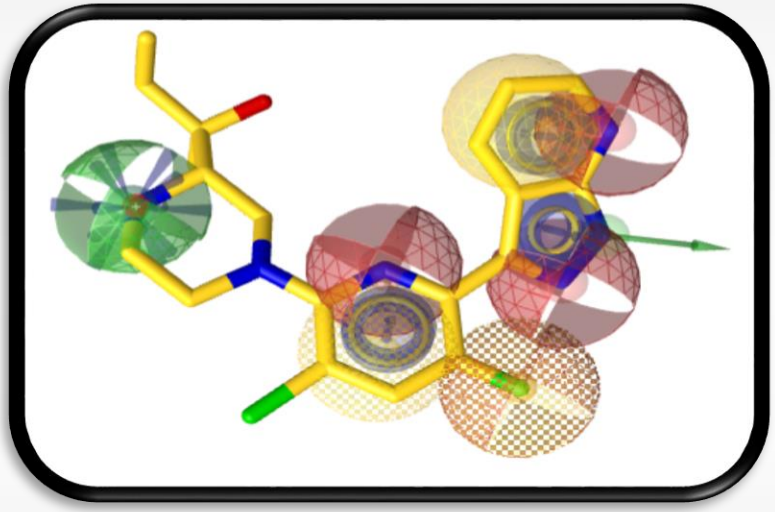
DATASET UNICO



RDKit *vs* LIGANDSCOUT

RDKit	DATASET A	LigandScout (LS)
13 CARATTERISTICHE		13 CARATTERISTICHE

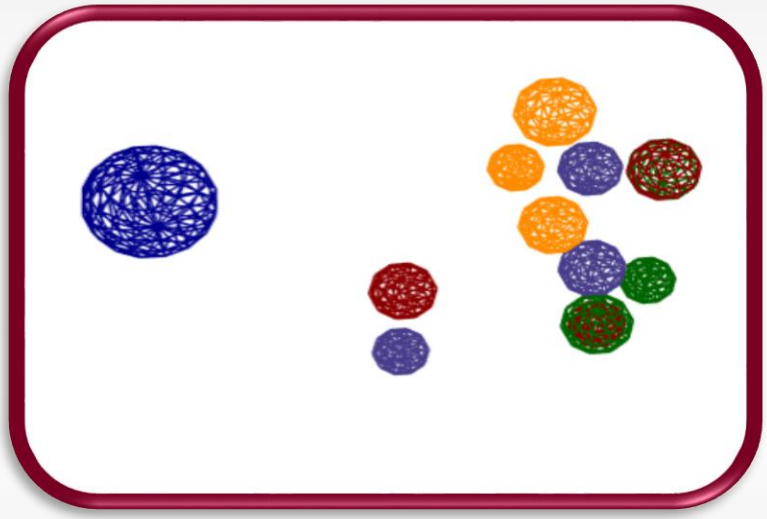


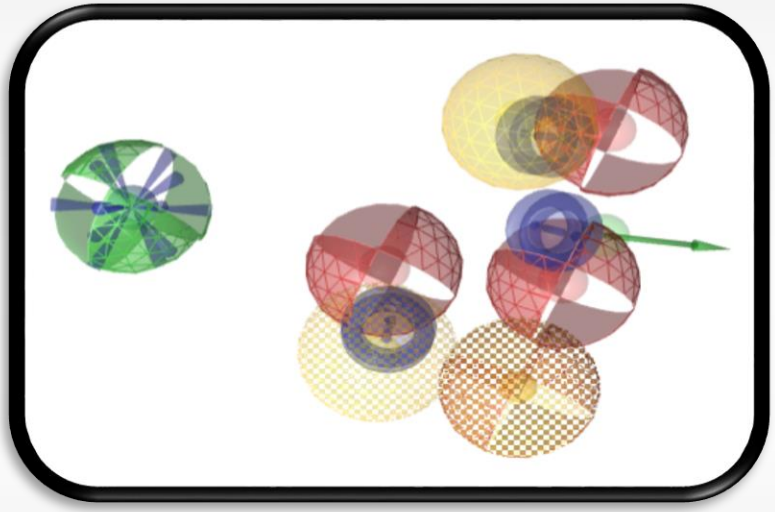


3	INTERAZIONI IDROFOBICHE (GIALLO)	3
3	INTERAZIONI AROMATICHE (VIOLETTO)	3
1	AREA IONIZZABILE POSITIVA (BLU)	1
3	ACCETTORE LEGAME-H (ROSSO)	4
3	DONATORE LEGAME-H (VERDE)	2

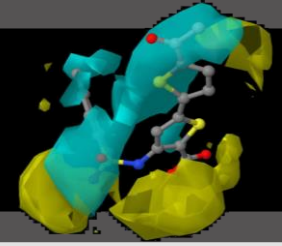
RDKit *vs* LIGANDSCOUT

RDKit	DATASET A	LigandScout (LS)
13 CARATTERISTICHE		13 CARATTERISTICHE

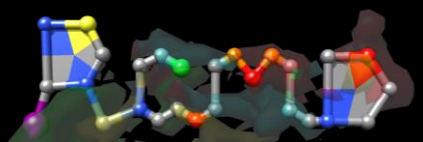




3	INTERAZIONI IDROFOBICHE (GIALLO)	3
3	INTERAZIONI AROMATICHE (VIOLETTO)	3
1	AREA IONIZZABILE POSITIVA (BLU)	1
3	ACCETTORE LEGAME-H (ROSSO)	4
3	DONATORE LEGAME-H (VERDE)	2

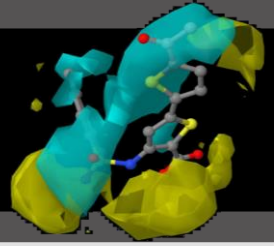


MODELLI 3-D QSAR

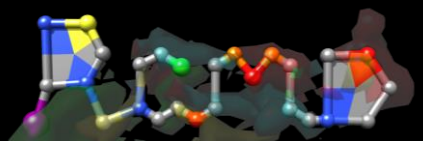


Dataset	ID_Model	Probe Atom	q^2 STE	q^2 ELE	q^2 BOTH	ONPC	Grid Spacing	Grid Extension	CutOff	Min. Sigma	CV Type	r^2	SDEC	SDEP
RDKit_A	98	F	0.567	0.445	**0.642**	3	2.0	10	30.0	1.00	LOO	0.929	0.281	0.630
LS_A	110	S.O	0.415	0.291	**0.525**	8	1.7	5	20.0	1.60	LOO	0.998	0.041	0.725
RDKit_B	101	Ca	0.868	0.632	**0.872**	2	2.5	5	35.0	1.15	LOO	0.909	0.416	0.495
LS_B	117	Al	0.869	0.554	**0.872**	3	1.2	2	20.0	1.95	LOO	0.941	0.336	0.502
RDKit_C	123	K	0.696	0.454	**0.674**	8	1.0	10	20.0	5.00	LSO	0.948	0.213	0.531
LS_C	123	H.P	0.637	0.385	**0.603**	7	1.9	5	20.0	4.00	LSO	0.947	0.214	0.586
RDKit_UNICO	122	H.spc	0.701	0.475	**0.698**	8	1.0	4	35.0	3.00	LSO	0.970	0.223	0.711
LS_UNICO	120	C.3	0.616	0.225	**0.626**	8	2.6	5	30.0	4.00	LOO	0.909	0.390	0.791

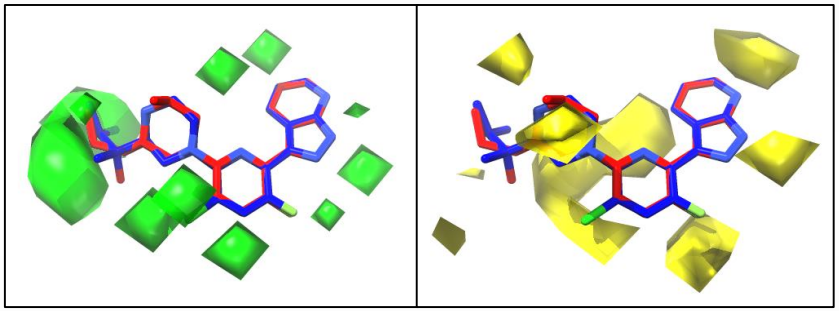
	Dataset A	Dataset B	Dataset C	Dataset UNICO
Incremento percentuale q^2	12 %	0 %	7 %	7 %



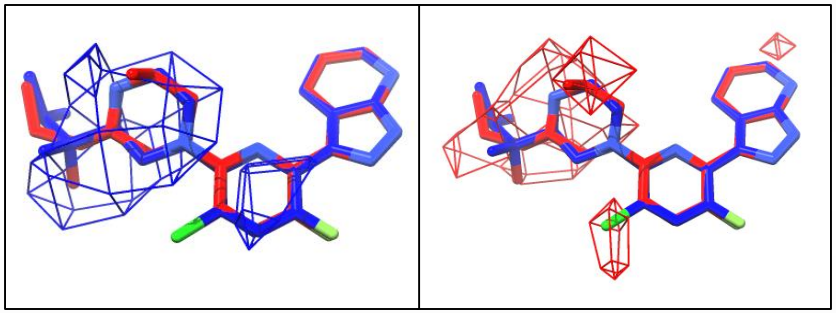
MAPPE CoMFA



Dataset	ID_Model	Probe Atom	q^2 STE	q^2 ELE	q^2 BOTH	ONPC	Grid Spacing	Grid Extension	CutOff	Min. Sigma	CV Type	r^2	SDEC	SDEP
RDKit_A	98	F	0.567	0.445	**0.642**	3	2.0	10	30.0	1.00	LOO	0.929	0.281	0.630
LS_A	110	S.O	0.415	0.291	**0.525**	8	1.7	5	20.0	1.60	LOO	0.998	0.041	0.725

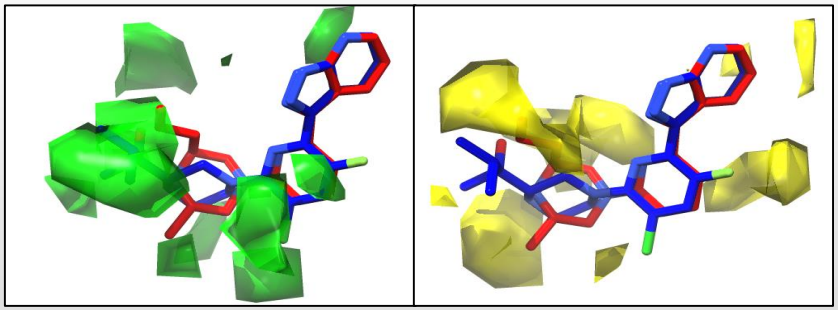


RDKit

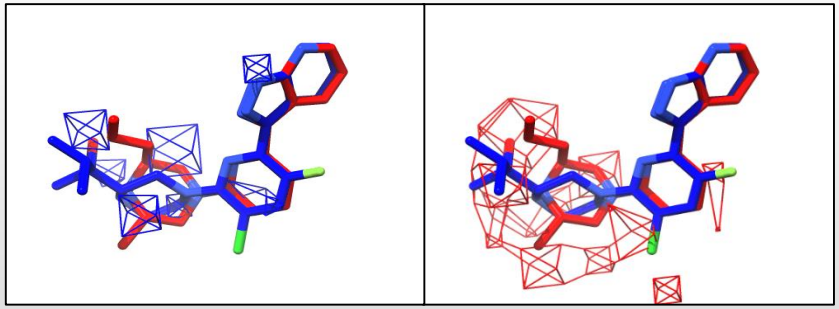


STERICHE (POS / NEG)

ELETTROSTATICHE (POS / NEG)



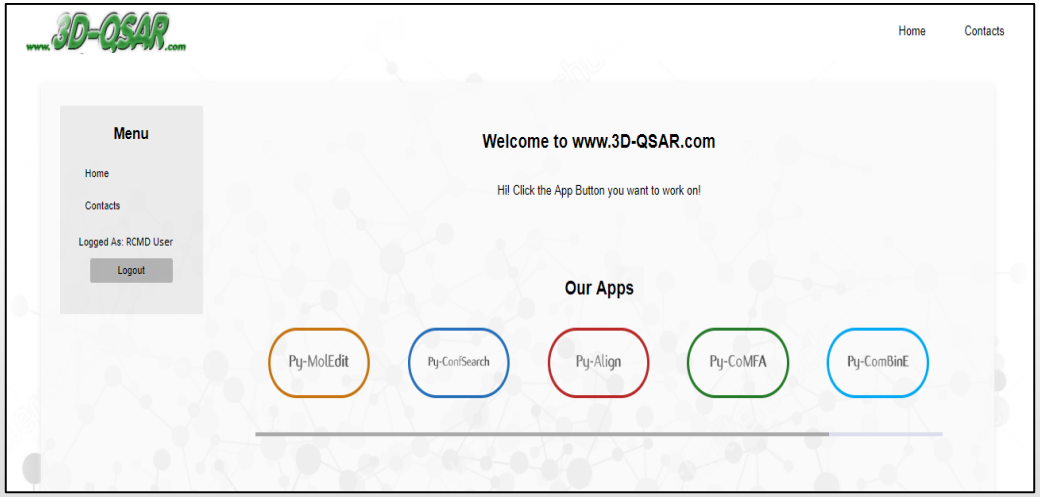
LS

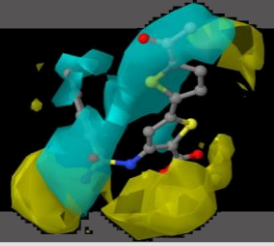


CONCLUSIONI

RDKit:

- Buone potenzialità per approcciarsi a metodiche di modeling farmacoforico *ligand-based*.
- Allineamenti molecolari migliori rispetto al *software* LigandScout.
- Modelli 3-D QSAR più stabili con valori di q^2 maggiori.





Grazie per l'attenzione!