Studio di modelli farmacoforici attraverso l'uso del codice PyPharm, confrontati con modelli ottenuti mediante l'uso del software commerciale LigandScout. Applicazione a serie di inibitori della proteina chinasi TYK2

SAPIENZA
UNIVERSITÀ DI ROMA

Facoltà di Farmacia e Medicina
Corso di Laurea in Biotecnologie Farmaceutiche
Tesi Sperimentale in Chimica Farmaceutica
A.A. 2021/2022

CANDIDATO:
Francesco Ciano (Matr. 1527501)

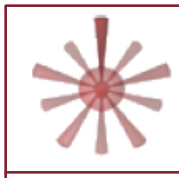
RELATORE:
Prof. Rino Ragno

FARMACOFORO

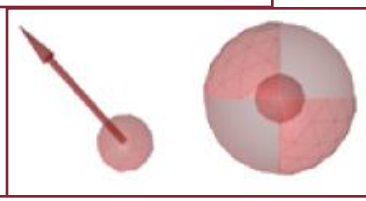
Figura geometrica virtuale ai cui vertici sono posizionate diverse caratteristiche farmacoforiche, legate mediante regole geometriche.

PRINCIPALI CARATTERISTICHE FARMACOFORICHE

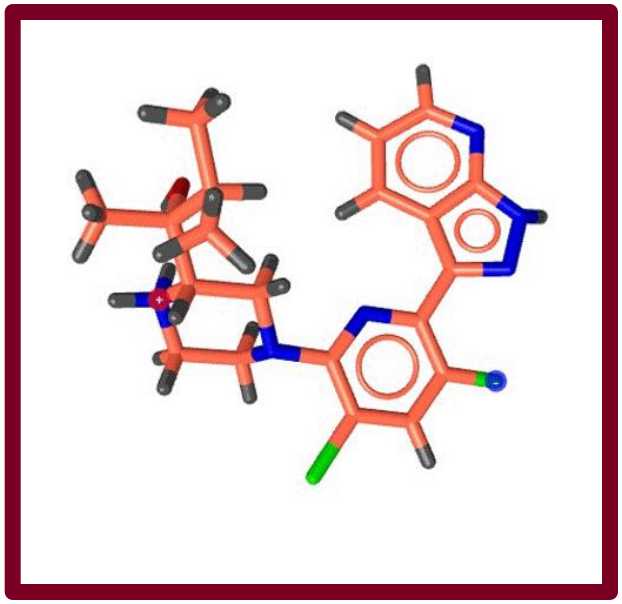
Centroide Idrofobico



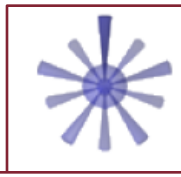
Area Ionizzabile Negativa



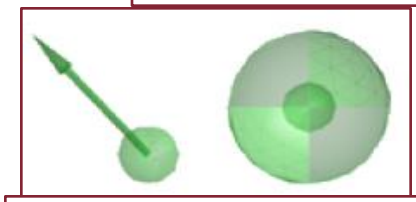
Accettore Legame–H



Anello Aromatico



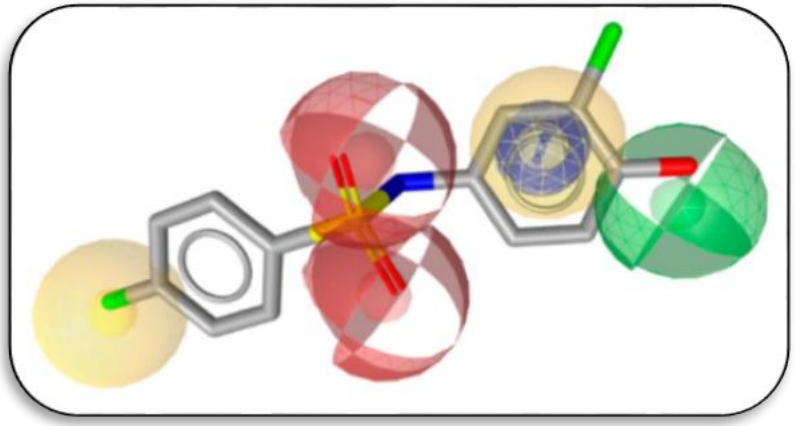
Area Ionizzabile Positiva



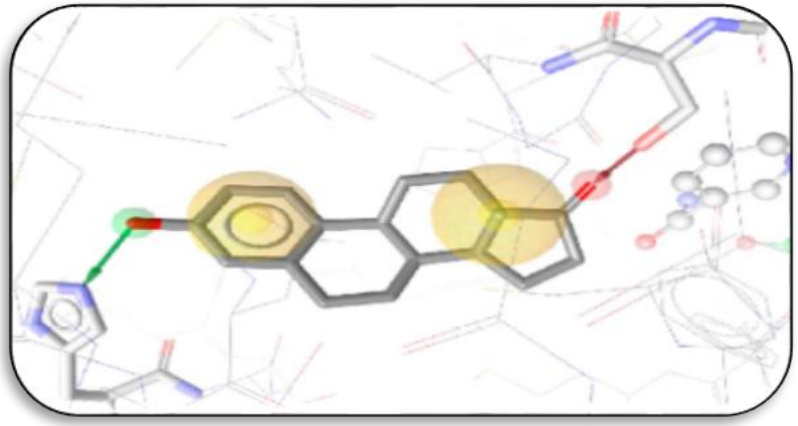
Donatore Legame–H

Ciano Francesco

LIGAND-BASED PHARMACOPHORE MODELING

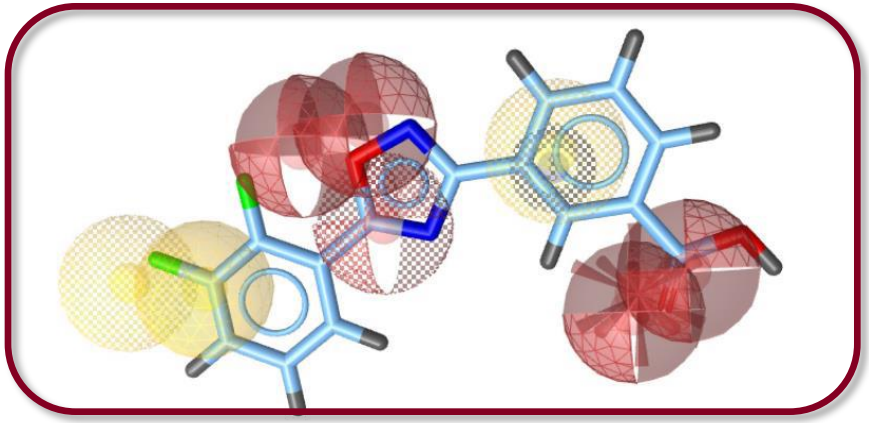


STRUCTURE-BASED PHARMACOPHORE MODELING

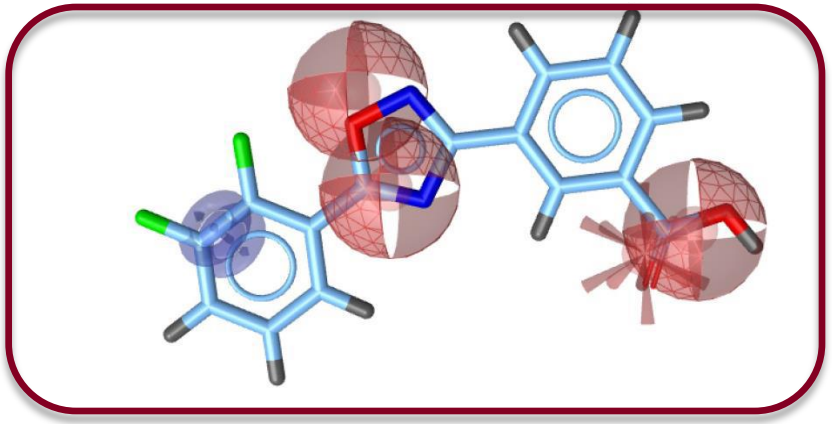


Oltre alla divisione in structure based e ligand based, dal punto di vista della condivisione dei punti farmacoforici è possibile distinguere:

MERGED PHARMACOPHORE

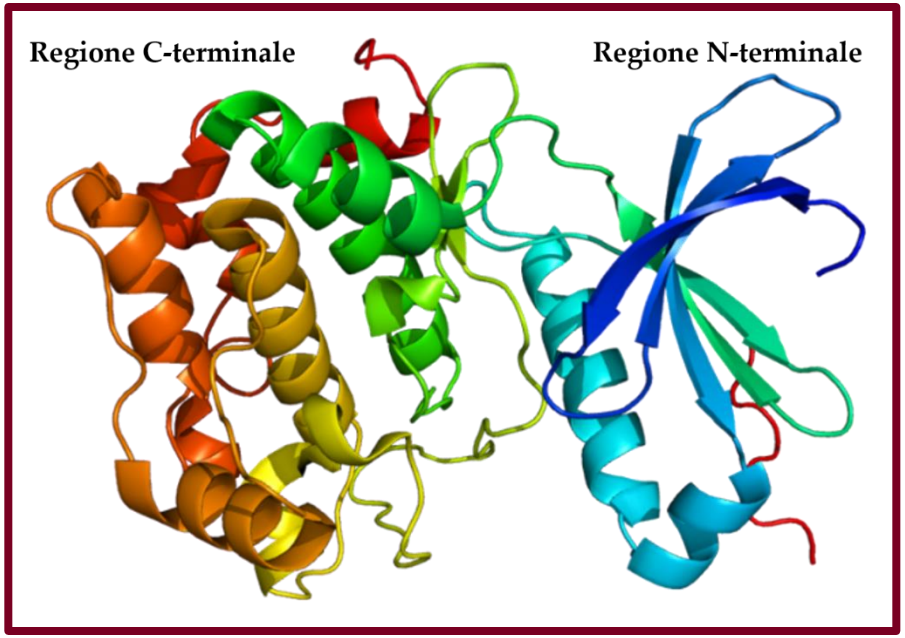


SHARED PHARMACOPHORE



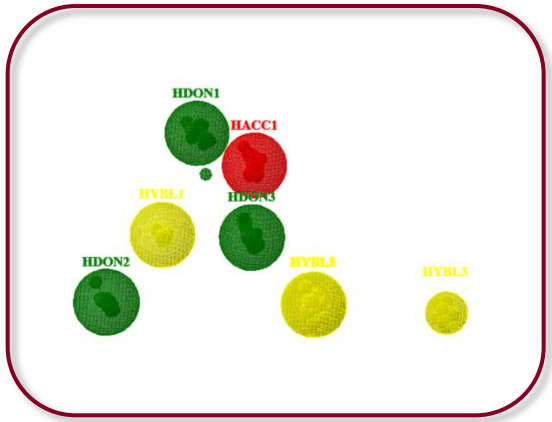
1. Creazione di un dataset di molecole
2. Analisi conformazionale
3. Allineamento molecolare
4. Generazione farmacoforo

Le JAK attivate vanno a fosforilare specifici residui di tirosina presenti sul recettore associato, che agiscono da sito di reclutamento delle proteine STAT.

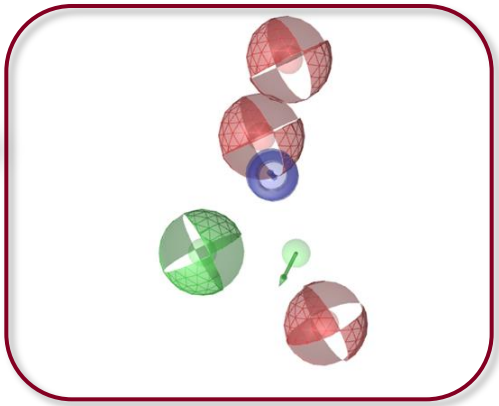


- **FAMIGLIA:** TIROSIN-PROTEIN CHINASICA NON RECETTORIALE
- **SOTTOFAMIGLIA:** JAK
- **ISOFORMA:** TYK2
- **PRINCIPALE LOCALIZZAZIONE:** CELLULE EMATOPOIETICHE

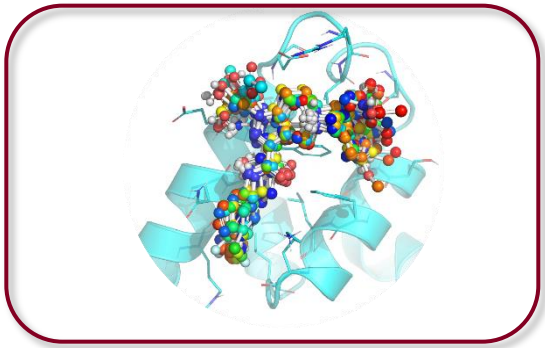
I modelli farmacoforici sono stati generati con due metodiche di lavoro differenti al fine di individuare somiglianze ed eventuali differenze nelle procedure operative. I modelli ottenuti sono stati poi utilizzati in analisi di VS.



PYPHARM



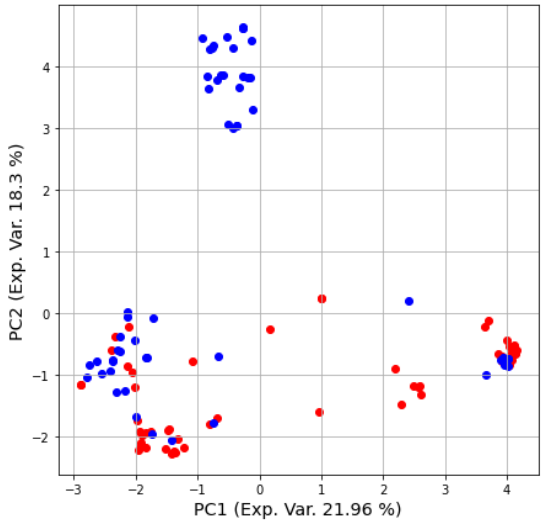
LIGANDSCOUT



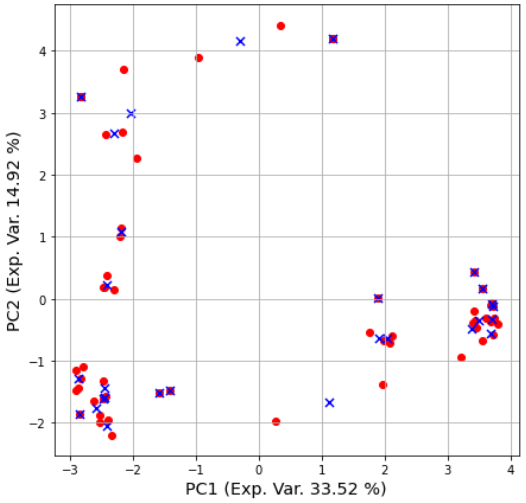
VIRTUAL SCREENING

SUDDIVISIONE MOLECOLE DATASET

ID MOLECOLA	SMILES	IC_{50}(nM)
C_35	C1CN(CC[NH2+]1)c1cccc(n1)-c1n[nH]c2ncccc12	0,75
B_20	CC(C)[C@@](C)(O)[C@@H]1CN(CC[NH2+]1)c1ccc(Cl)c(n1)-c1n[nH]c2ncccc12	1,69
D_26	CC(C)[C@@](C)(O)[C@@H]1CN(CC[NH2+]1)c1nc(-c2n[nH]c3ncccc23)c(F)cc1O	325

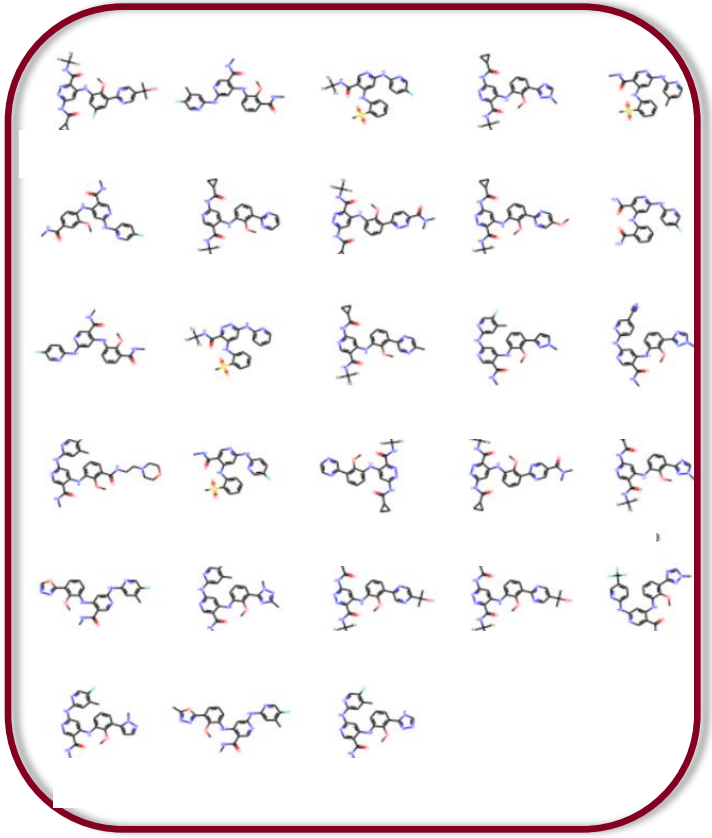


SUDDIVISIONE MOLECOLE IN BASE ALLA LORO POTENZA

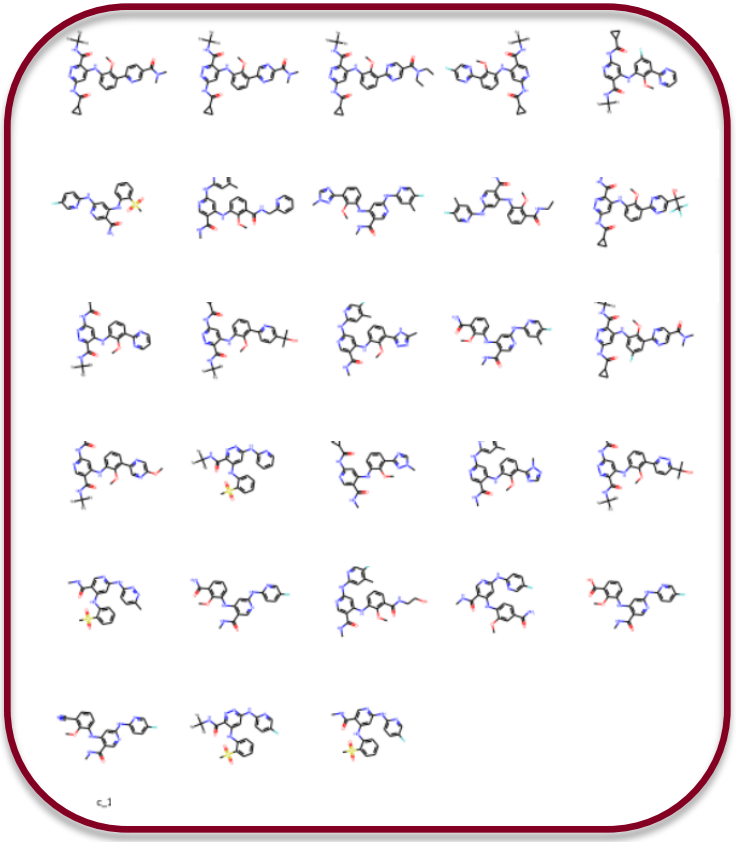


CLUSTER DI MOLECOLE

Le molecole classificate come attive sono state suddivise in due sotto dataset, dove ogni dataset conteneva molecole che tra di loro risultavano strutturalmente simili:



MOLECOLE RAPPRESENTATIVE DI OGNI CLUSTER

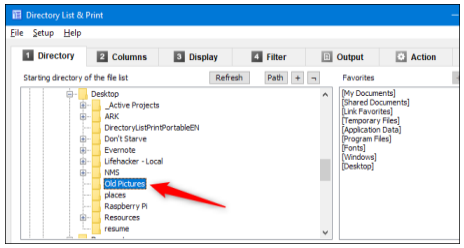


MOLECOLE NON RAPPRESENTATIVE DI OGNI CLUSTER

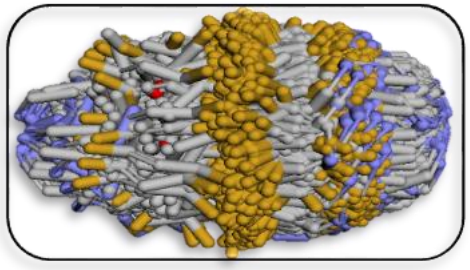
Prima di avviare la costruzione di un modello farmacoforico sarà opera dell'utente settare alcuni paramenti di lavoro

```
## Arguments, set them up for your system

workdir = "" # Full path better

project = 'pharma2' #Name of project

smiles = 'data/molecole_piu_rappresentative_di_ogni_cluster.smi' #Comma seperated list of smile, numerical ID

conf = '200' #Max number of conformations per molecule

ref_select_method = 'Max_points' #Manual, Max_Points

ref = '' #if ref_select_method is manual, Max_points is usually best unless activity is known
         #also supports other common formats (SDF, Mol2)

align_it = "/home/jepolitsch/.conda/envs/build/bin/align_it" #Path to exectable
```

NOME PROGETTO



FILE DIRECTORY



NUMERO DI CONFORMAZIONI

SCELTA DEL REFERENCE

```
def ref_max_points():
    max_points = 0
    max_mol = None
    mols = open(smiles, 'r+').readlines()
```

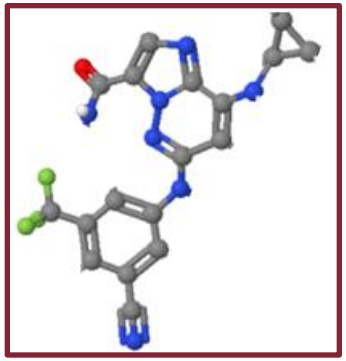
SCELTA DEL REFERENCE IN BASE AL MAX NUMERO DI PUNTI FARMACOFORICI

```
for i in mols:
    mol = i.split(',')
    smile, id = mol[0:2]
    mol = Chem.MolFromSmiles(smile)
    mol = Chem.AddHs(mol)
    AllChem.EmbedMolecule(mol, randomSeed=42)
    pharm = P()
    pharm.load_from_mol(mol)
    features = pharm.get_features_count()
    tot_feat = (sum(list(features.values())))
```

CICLO FOR APPLICATO A TUTTE LE MOLECOLE

```
ref = select_ref()
ref3D = gen3D(ref,str(project_path+"_ref.sdf"), '1', 1000)
training3D = gen3D(smiles, str((project_path + "_gen3D.sdf")), '10', 1000)

Molecule d_22 selected as reference with 23 pharmacophore points
```

IDENTIFICAZIONE REFERENCE



MOLECOLE d_22

Per ogni molecola viene scelta la conformazione migliore, a questa vengono associati tutta una serie di parametri che l'utente può visualizzare e confrontare con la molecola reference

	Ref_ID	Vol_Ref	Probe_ID	Max_Vol_Probe	Max_Overlap	Exclu_Spheres	Adj_Max_Vol	Num_Shared_Feats	TANIMOTO	TVERSKY_Ref	TVERSKY_Probe
0	d_22.0	228.958	d_18.6	228.958	209.584	0	209.584	11	0.8440	0.9154	0.9154
1	d_22.0	228.958	d_22.0	228.958	228.958	0	228.958	11	1.0000	1.0000	1.0000
2	d_22.0	228.958	d_31.2	213.208	205.699	0	205.699	10	0.8699	0.8984	0.9648
3	d_22.0	228.958	d_12.0	213.208	211.034	0	211.034	10	0.9130	0.9217	0.9898
4	d_22.0	228.958	d_33.6	213.208	208.496	0	208.496	10	0.8923	0.9106	0.9779
5	d_22.0	228.958	d_23.3	228.958	190.240	0	190.240	10	0.7107	0.8309	0.8309
6	d_22.0	228.958	d_19.9	228.958	199.747	0	199.747	10	0.7737	0.8724	0.8724
7	d_22.0	228.958	c_43.3	213.208	175.905	0	175.905	9	0.6606	0.7683	0.8250
8	d_22.0	228.958	d_29.2	255.850	153.213	0	153.213	9	0.4620	0.6692	0.5988
9	d_22.0	228.958	d_26.1	228.958	171.069	0	171.069	9	0.5964	0.7472	0.7472
10	d_22.0	228.958	d_24.1	228.958	184.453	0	184.453	9	0.6745	0.8056	0.8056
11	d_22.0	228.958	d_30.8	213.208	176.911	0	176.911	9	0.6670	0.7727	0.8298
12	d_22.0	228.958	c_33.6	224.350	173.512	0	173.512	9	0.6201	0.7578	0.7734
13	d_22.0	228.958	c_30.5	224.350	168.067	0	168.067	9	0.5892	0.7341	0.7491
14	d_22.0	228.958	c_38.0	224.350	161.175	0	161.175	8	0.5517	0.7040	0.7184

VALORI DI RIFERIMENTO REFERENCE

VALORI DI CONFRONTO DELLE MOLECOLE

CALCOLO DEL NUMERO DI PUNTI FARMACOFORICI UTILIZZANDO LA DISTANZA EUCLIDEA

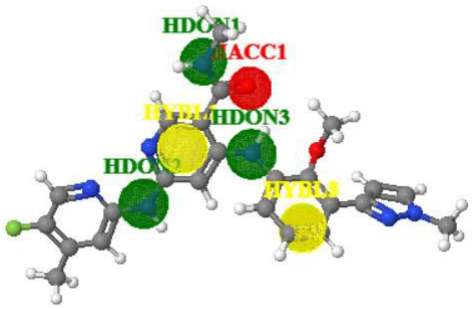
Per ogni molecola viene calcolato, tramite la distanza euclidea, quanti punti farmacoforici condivide con il reference (treshold customizable).
Da questo tipo di dataframe è facilmente ottenibile un farmacoforo shared che l'utente può selezionare sia richiedendo unicamente i punti condivisi oltre un certo treshold sia richiedendo un farmacoforo con un numero di punti definito.

	HYBL1	HYBL2	HYBL3	HDON1	HDON2	HDON3	HYBH1	HACC1	HACC2	HYBL4	HYBL5
c_23.9	1	1	0	1	1	1	0	1	0	0	1
d_12.0	1	1	1	1	1	1	0	1	1	1	1
b_30.6	1	1	0	1	1	1	0	1	0	0	1
c_18.8	1	1	0	1	1	1	0	1	0	0	1
d_31.2	1	1	1	1	1	1	0	1	1	1	1
c_40.8	1	1	1	1	1	1	0	1	0	0	1
b_29.8	1	1	0	1	1	1	0	1	0	0	1
c_21.9	1	1	0	1	1	1	0	1	0	0	1
d_19.9	1	1	1	1	1	1	0	1	1	0	1
c_43.3	1	1	1	1	1	1	0	1	1	0	1
d_23.3	1	1	1	1	1	1	0	1	1	1	1

GRIGLIA DI CONDIVISIONE DEI PUNTI FARMACOFORICI

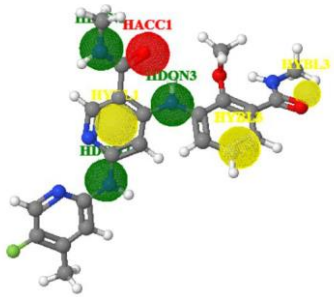
COORDINATE SPAZIALI DEI PUNTI CONDIVISI

	HYBL1	HYBL2	HYBL3	HDON1	HDON2	HDON3	HYBH1	HACC1	HACC2	HYBL4	HYBL5
c_23.9	[3.99764, -0.382187, 1.00448]	[-0.793338, 2.22491, -1.21993]	NaN	[3.34791, -0.645711, 4.63846]	[5.35145, 0.0911031, -1.3849]	[1.19043, -0.0717423, 1.13562]	NaN	[1.49494, -1.37683, 3.49246]	NaN	NaN	[-0.793338, 2.22491, -1.21993]
d_12.0	[3.97623, -0.472332, 1.15176]	[-0.704342, 2.35511, -1.26787]	[-5.0546, 0.637589, -0.87937]	[3.31997, -1.11466, 4.65862]	[5.4951, -0.199035, -1.1445]	[1.17456, -0.120902, 1.02754]	NaN	[1.28795, -0.804885, 3.62939]	[7.65248, -0.582659, -0.340106]	[7.74371, 1.00664, -3.14863]	[-0.704342, 2.35511, -1.26787]
b_30.6	[3.91686, -0.380797, 1.07124]	[-0.928535, 2.03911, -1.32916]	NaN	[3.32202, -1.11537, 4.65096]	[5.21613, -0.490234, -1.36205]	[1.09062, -0.0903116, 1.29172]	NaN	[1.71814, 0.349387, 3.89153]	NaN	NaN	[-0.928535, 2.03911, -1.32916]
c_18.8	[3.98416, -0.308232, 1.00548]	[-0.740701, 2.16294, -1.24673]	NaN	[3.39916, -1.33777, 4.52082]	[5.27403, -0.156555, -1.46365]	[1.15646, -0.196078, 1.10357]	NaN	[1.54202, -0.231291, 3.75777]	NaN	NaN	[-0.740701, 2.16294, -1.24673]



- INTERAZIONI LIPOFILICHE SU ANELLI AROMATICI (GIALLO)
- DONATORI DI LEGAMI IDROGENO DA PARTE DELL'AZOTO (VERDI)
- ACCETTORI DI LEGAMI IDROGENO DA PARTE DELL'OSSIGENO (ROSSO)

Ph4_B

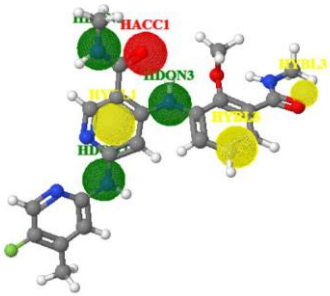


- INTERAZIONI LIPOFILICHE SU ANELLI AROMATICI (GIALLO)
- DONATORI DI LEGAMI IDROGENO DA PARTE DELL'AZOTO (VERDI)
- ACCETTORI DI LEGAMI IDROGENO DA PARTE DELL'OSSIGENO (ROSSO)

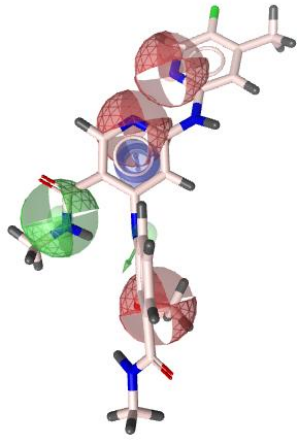
Ph4_A

Ph4_A

PYPHARM



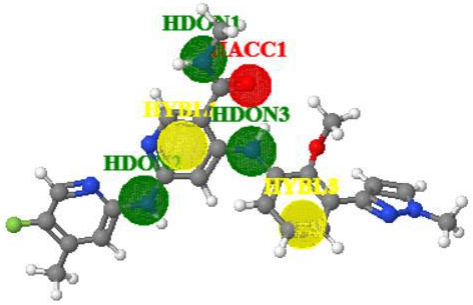
LIGANDSCOUT



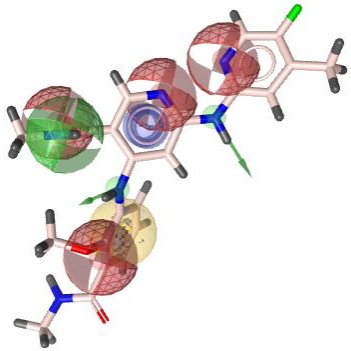
3	**INTERAZIONI LIPOFILICHE/AROMATICO (GIALLO)/(AZZURRO)**	**1**
3	**DONATORI DI LEGAMI IDROGENO (VERDE)**	**2**
1	**ACCETTORI DI LEGAMI IDROGENO (ROSSO)**	**3**

Ph4_B

PYPHARM

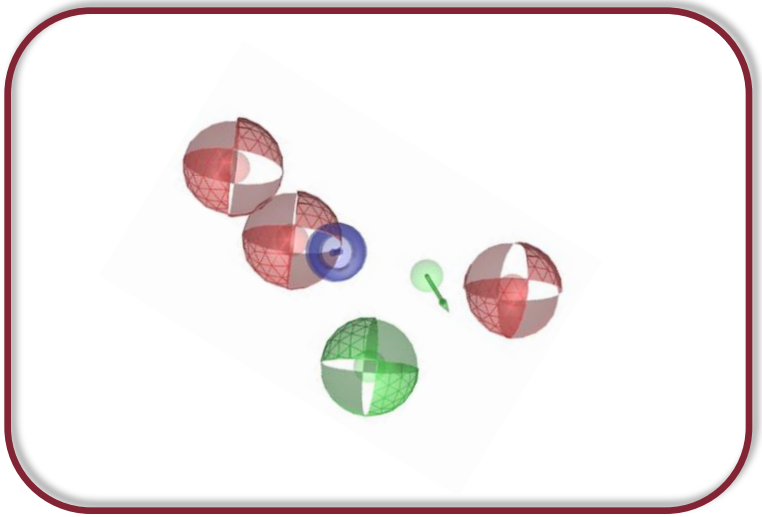


LIGANDSCOUT



2	INTERAZIONI LIPOFILICHE/AROMATICO (GIALLO)/(AZZURRO)	2
3	DONATORI DI LEGAMI IDROGENO (VERDE)	2
1	ACCETTORI DI LEGAMI IDROGENO (ROSSO)	4

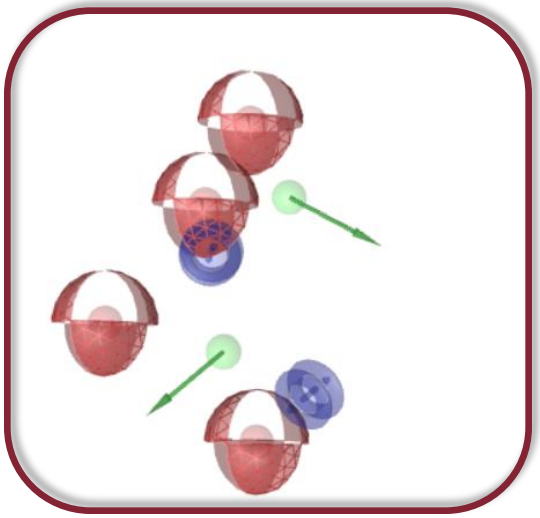
Ph4_A



Il modello Ph4_A è stato in grado di identificare 24 inibitori secondo il valore di treshold impostato

ChEMBL ID	*SMILES*	*IC50(nM)*
CHEMBL4435170	[2H]C([2H])([2H])NC(=O)c1nnc(NC(=O)C2CC2)cc1Nc1cccc(-c2ncn(C)n2)c1OC	0.2
CHEMBL4438202	CNC(=O)c1cnc(NC(=O)C2CC2)cc1Nc1ccccc1S(C)(=O)=O	1.6
CHEMBL4440718	[2H]C([2H])([2H])NC(=O)c1nnc(Nc2ccccn2)cc1Nc1ccccc1S(C)(=O)=O	0.53
CHEMBL4542289	CCCC(=O)Nc1cc(Nc2ccccc2S(C)(=O)=O)c(C(=O)NC)cn1	2500
CHEMBL4444178	[2H]C([2H])([2H])NC(=O)c1nnc(NC(=O)C2CC2)cc1Nc1ccccc1S(C)(=O)=O	1.3
CHEMBL4530719	CS(=O)(=O)c1ccccc1Nc1cc(Nc2ccc(F)cn2)ncc1C(N)=O	0.32
CHEMBL4542289	CCCC(=O)Nc1cc(Nc2ccccc2S(C)(=O)=O)c(C(=O)NC)cn1	2500
CHEMBL4561663	NC(=O)c1ccccc1Nc1cc(Nc2ccc(F)cn2)ncc1C(N)=O	0.46
CHEMBL4571117	[2H]C([2H])([2H])NC(=O)c1cnc(NC(=O)C2CC2)cc1Nc1cccc(-c2ncn(C)n2)c1OC	13
CHEMBL4583185	[2H]C([2H])([2H])NC(=O)c1nnc(Nc2ccc(F)cn2)cc1Nc1ccccc1S(C)(=O)=O	0.58
CHEMBL4587924	CNC(=O)c1cnc(Nc2ccc(F)cn2)cc1Nc1cccc(C#N)c1OC	570
CHEMBL4776752	[2H]C([2H])([2H])NC(=O)c1nnc(NC(=O)C2CC2)cc1Nc1cccc(-c2nccc(OC)n2)c1OC	1.4
CHEMBL4777175	[2H]C([2H])([2H])NC(=O)c1nnc(NC(=O)C2CC2)cc1Nc1cccc(-c2ncc(C(C)(O)C(F)(F)F)cn2)c1OC	0.44
CHEMBL4777528	[2H]C([2H])([2H])NC(=O)c1nnc(NC(=O)C2CC2)cc1Nc1cccc(-c2cnc(C(=O)N(C)C)cn2)c1OC	0.25
CHEMBL4777648	[2H]C([2H])([2H])NC(=O)c1nnc(NC(=O)C2CC2)cc1Nc1cccc(-c2ccncn2)c1OC	1.3
CHEMBL4780057	[2H]C([2H])([2H])NC(=O)c1nnc(NC(=O)C2CC2)cc1Nc1cccc(-c2nccc(C)n2)c1OC	4.5
CHEMBL4780233	[2H]C([2H])([2H])NC(=O)c1cnc(NC(=O)C2CC2)cc1Nc1cccc(-c2cnc(C)cn2)c1OC	0.27
CHEMBL4780315	[2H]C([2H])([2H])NC(=O)c1cnc(NC(=O)C2CC2)cc1Nc1cccc(-c2cnc(OC)cn2)c1OC	0.54
CHEMBL4782449	[2H]C([2H])([2H])NC(=O)c1nnc(NC(=O)C2CC2)cc1Nc1cccc(-c2cnccn2)c1OC	0.5
CHEMBL4784391	[2H]C([2H])([2H])NC(=O)c1nnc(NC(=O)C2CC2)cc1Nc1cccc(-c2ncc(F)cn2)c1OC	1.9
CHEMBL4784838	[2H]C([2H])([2H])NC(=O)c1nnc(NC(=O)C2CC2)cc1Nc1cc(F)cc(-c2ncc(C(C)(C)O)cn2)c1OC	0.51
CHEMBL4785512	[2H]C([2H])([2H])NC(=O)c1nnc(NC(=O)C2CC2)cc1Nc1cccc(-c2ccc(C(=O)N(C)C)cn2)c1OC	0.25
CHEMBL4785957	[2H]C([2H])([2H])NC(=O)c1nnc(NC(=O)C2CC2)cc1Nc1cccc(-c2ncc(C)cn2)c1OC	1.2
CHEMBL4788122	[2H]C([2H])([2H])NC(=O)c1nnc(NC(=O)C2CC2)cc1Nc1cccc(-c2cccnn2)c1OC	1.4
CHEMBL4789015	[2H]C([2H])([2H])NC(=O)c1cnc(NC(=O)C2CC2)cc1Nc1cccc(-c2ncccn2)c1OC	0.25

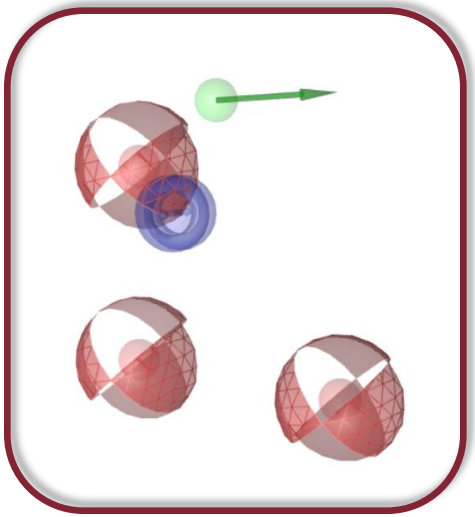
Ph4_B



ChEMBL ID	*SMILES*	*IC50(nM)*
CHEMBL4476830	CNC(=O)c1cnc(Nc2ccc(F)cn2)cc1Nc1ccc(F)c(F)c1OC	0.5

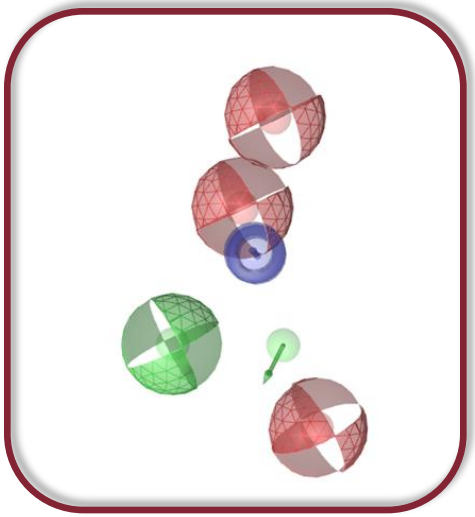
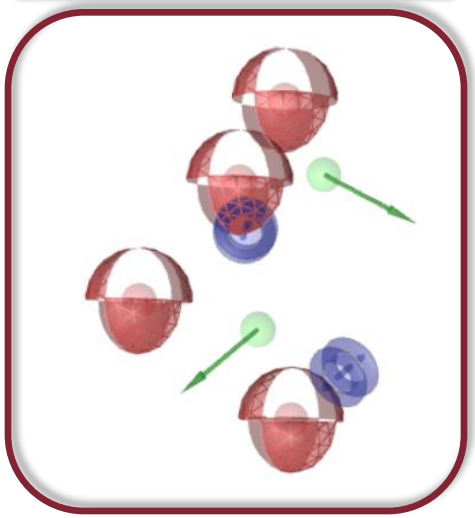
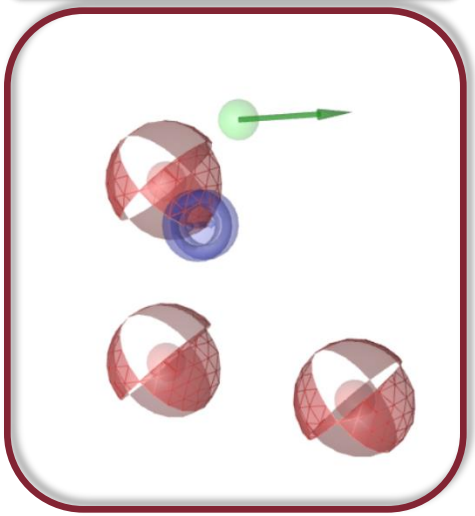
Il modello Ph4_B è stato in grado di identificare 1 solo inibitore secondo il valore di treshold impostato

Shared A&B



Il modello Shared A&B è stato in grado di identificare 5 inibitori secondo il valore di treshold impostato

ChEMBL ID	*SMILES*	*IC50(nM)*
CHEMBL4544320	CNC(=O)c1cnc(Nc2ccccn2)cc1Nc1ccccc1S(C)(=O)=O	0.95
CHEMBL4537678	CNC(=O)c1cnc(Nc2ccc(F)cn2)cc1Nc1ccccc1C(N)=O	0.3
CHEMBL4469812	CNC(=O)c1cnc(Nc2ccc(F)cn2)cc1Nc1ccccc1SC	0.8
CHEMBL4465409	CNC(=O)c1cnc(Nc2ccccn2)cc1Nc1ccccc1C(N)=O	0.6
CHEMBL4444169	[2H]C([2H])([2H])NC(=O)c1cnc(Nc2ccc(F)cn2)cc1Nc1ccccc1S(C)(=O)=O	0.49

Ph4_A	Ph4_B	Shared A&B
		
Buone capacità predittive	Scarse capacità predittive	Capacità predittive intermedie

CONCLUSIONI

- Il codice PyPharm ha mostrato buone potenzialità per approcciarsi a metodiche di modeling farmacoforico ligand-based.
- Rispetto a LigandScout la scelta ponderata della molecola reference potrebbe essere determinante nella realizzazione di modelli farmacoforici validi.
- La suddivisione preliminare delle molecole con annessa clusterizzazione ha portato all'identificazione del modello Ph4_A che mostrato buone capacità predittive.
- L'implementazione di una funzionalità di screening in PyPharm permetterebbe un ulteriore confronto con Ligand Scout.

RINGRAZIAMENTI

Prof. Rino Ragno

Dott. Lorenzo Antonini

Dott.ssa Manuela Sabatino

Andrea Di Giulio

Julian Politsch

Grazie per l'attenzione!