

Validazione del molecular docking mediante applicazione di cross-docking: valutazione dell'accuratezza di una serie di programmi mediante RMSD e molecular shape



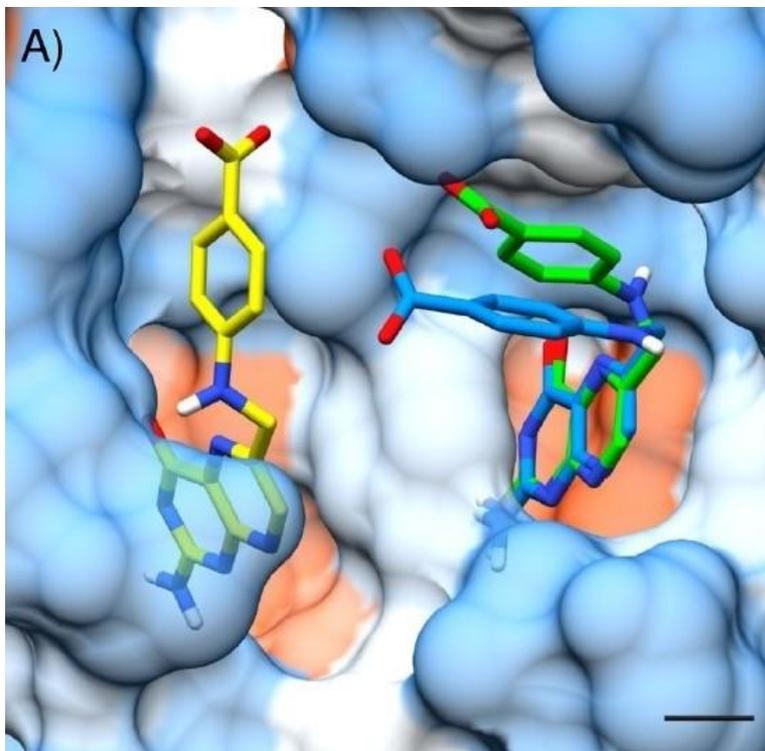
SAPIENZA
UNIVERSITÀ DI ROMA

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

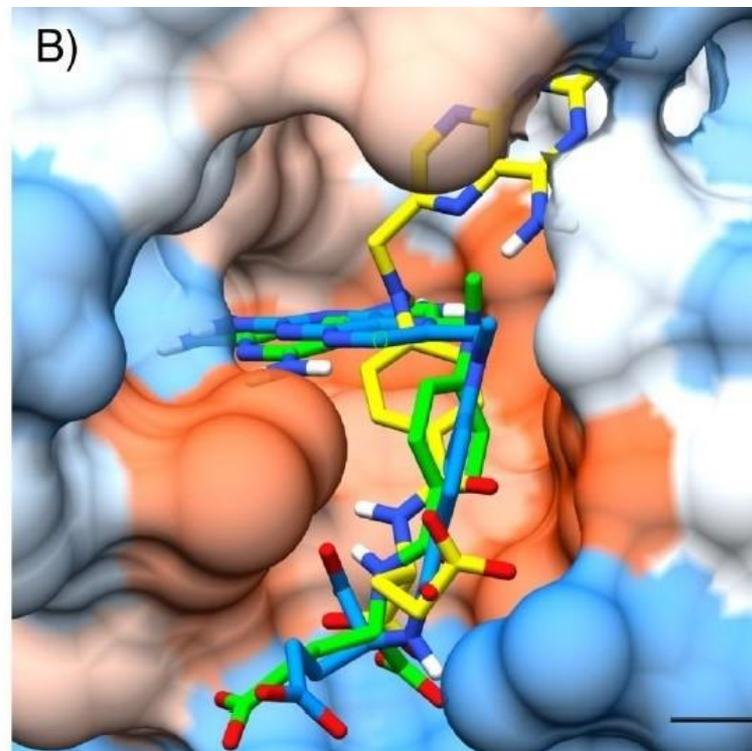
Laureando: Pietro Scio'
Matricola: 1806164

Relatore: prof. Rino Ragno

Il molecular docking e la sua validazione



PDB code: 1BR6
Ligando co-cristallizzato
Posa a minor energia secondo VINA
Posa a minor energia secondo VINARDO

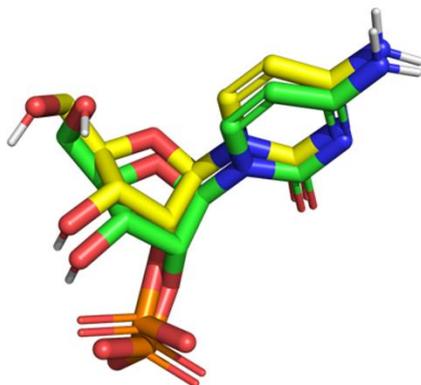


PDB code: 1DSS
Ligando co-cristallizzato
Posa a minor energia secondo VINA
Posa a minor energia secondo VINARDO

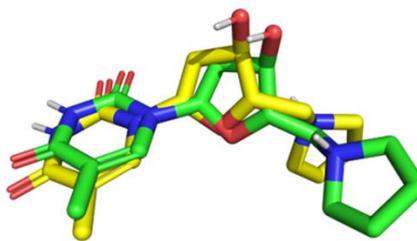
Metodiche di validazione attualmente in uso

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_{ci} - x_{di})^2 + (y_{ci} - y_{di})^2 + (z_{ci} - z_{di})^2}$$

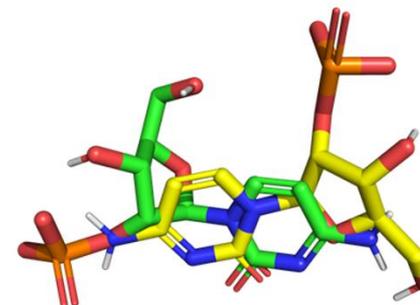
$$DA\% = \frac{f_{RMSD \leq 2} + 0,5f_{2 < RMSD \leq 3}}{f_{TOT}}$$



RMSD = 0,69



RMSD = 2,32

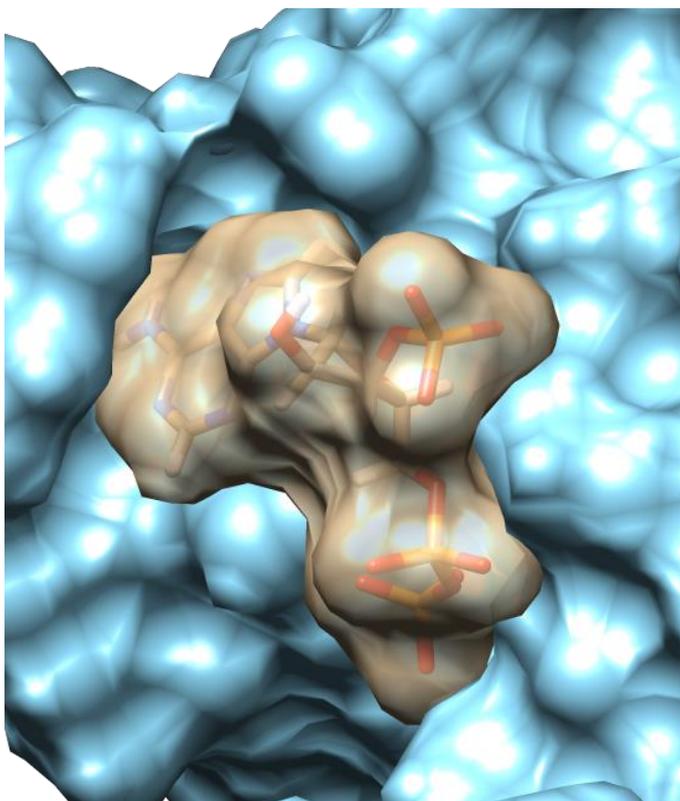


RMSD = 6,73

APPROCCIO ATOMICO

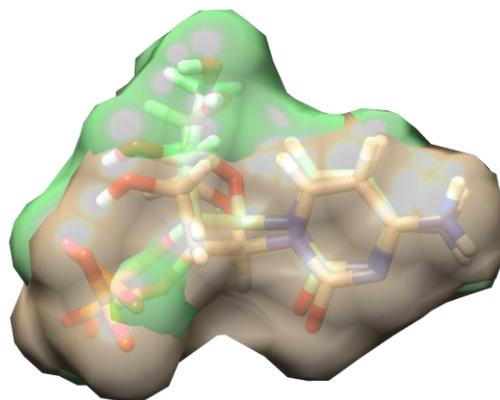
Finalità della tesi

Definizione di una nuova DA% basata sulla forma molecolare
(*molecular shape*) mediante applicazione di cross-docking

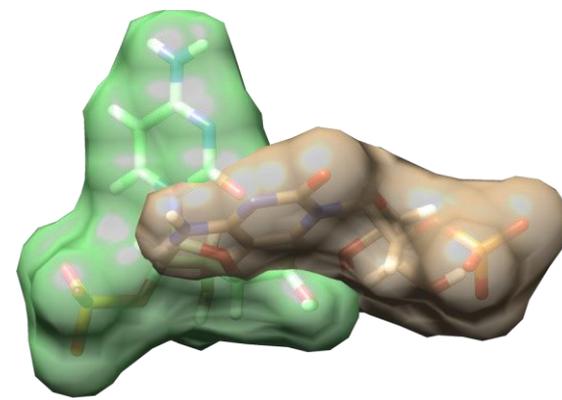


PDB code: 1AFK

$$\%VolOverlap = \frac{V_{sovrapposto}}{V_{conf.sperim}} \cdot 100$$



%VolOverlap = 96,44



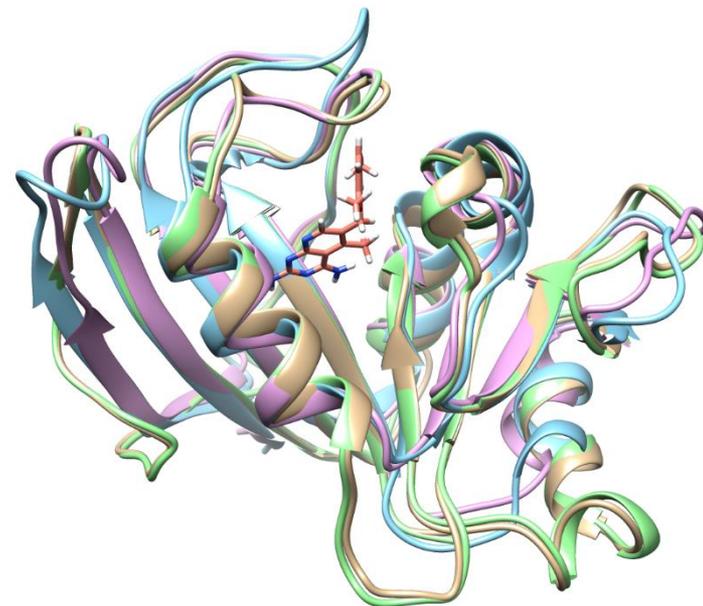
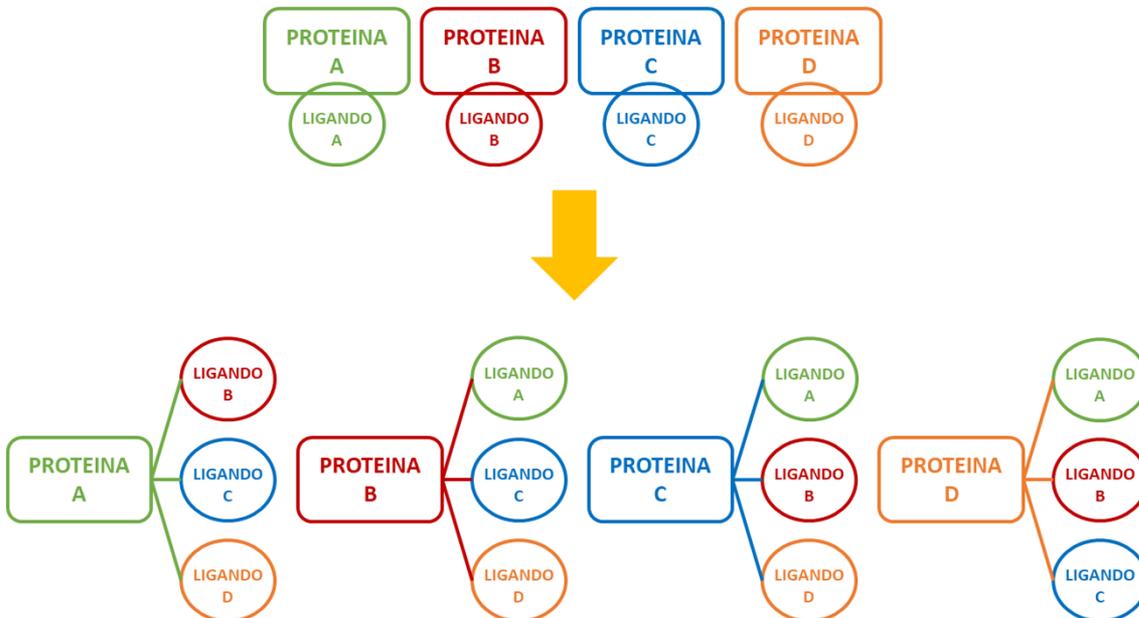
%VolOverlap = 21,42

APPROCCIO VOLUMETRICO

«**SUPER PROGRAMMA**» DI VALIDAZIONE
SELEZIONE AUTOMATICA DEL MIGLIOR SOFTWARE DI DOCKING



Cross-docking



**VIENE APPROSSIMATA
LA FLESSIBILITA' DEL
TARGET PROTEICO**

Panoramica generale

Costruzione di un **dataset di complessi**



Cross-docking

- 4 programmi diversi
 - 8 SFs totali
- Conformazione sperimentale e randomica



Definizione della
DA% volumetrica



Validazione
secondo la DA% volumetrica



Validazione
secondo la DA% atomica



Costruzione di un dataset di complessi



Refined Set
1412 targets
5316 complessi

Selezione targets
con >10 complessi

101 targets
2552 complessi



Clustering per similarità
dei ligandi co-cristallizzati

101 targets
404 complessi

Selezione dei
targets monomerici

43 targets
172 complessi

Selezione targets a
struttura primaria
minore

DATASET FINALE

20 targets
80 complessi

$$20 \text{ targets} \times 3 = 60 \text{ targets} \\ 80 \text{ complessi} \times 3 = 240 \text{ complessi}$$

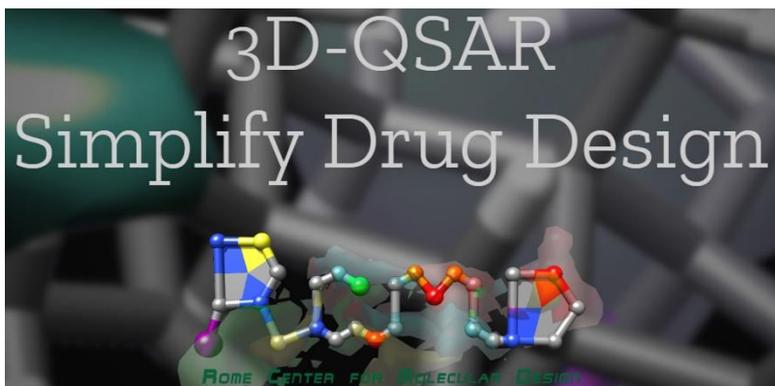
Esecuzione del cross-docking

PLANTS
plp
plp95
chemplp

SMINA
vinardo
vina
ad4

GALAXYDOCK3

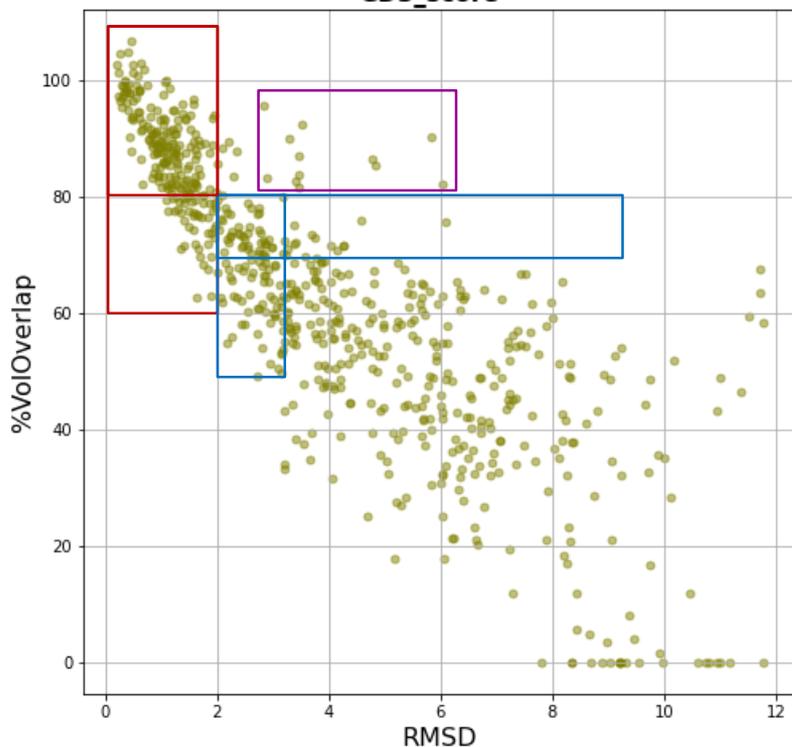
LEDOCK





Risultati

GalaxyDock3
ECCD
RMSD vs %VolOverlap
GD3_score

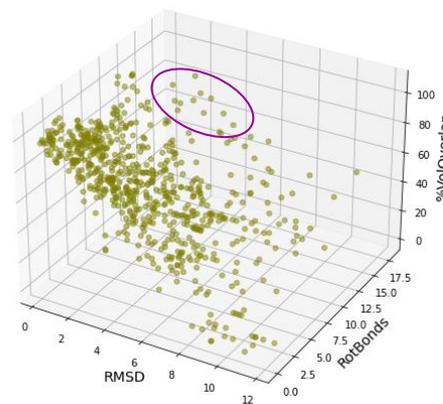


$$DA\% = \frac{f_{RMSD \leq 2} + 0,5f_{2 < RMSD \leq 3}}{f_{TOT}}$$

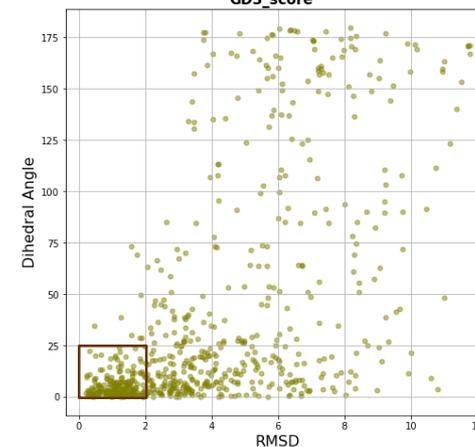


$$DA\% = \frac{f_{\%Vol.Overlap \geq 80} + 0,75f_{70 \leq \%Vol.Overlap < 80}}{f_{TOT}}$$

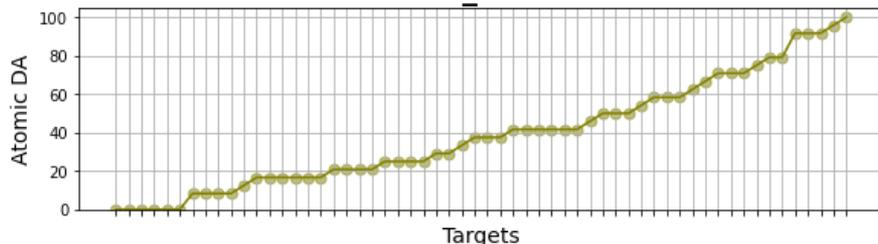
GalaxyDock3
ECCD
RMSD vs %VolOverlap vs Rot.Bonds
GD3_score



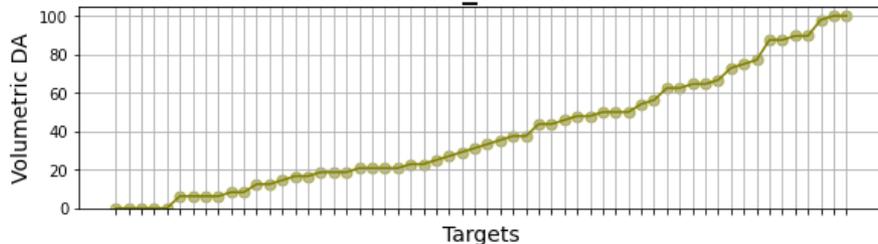
GalaxyDock3
ECCD
RMSD vs Dihedral Angle
GD3_score



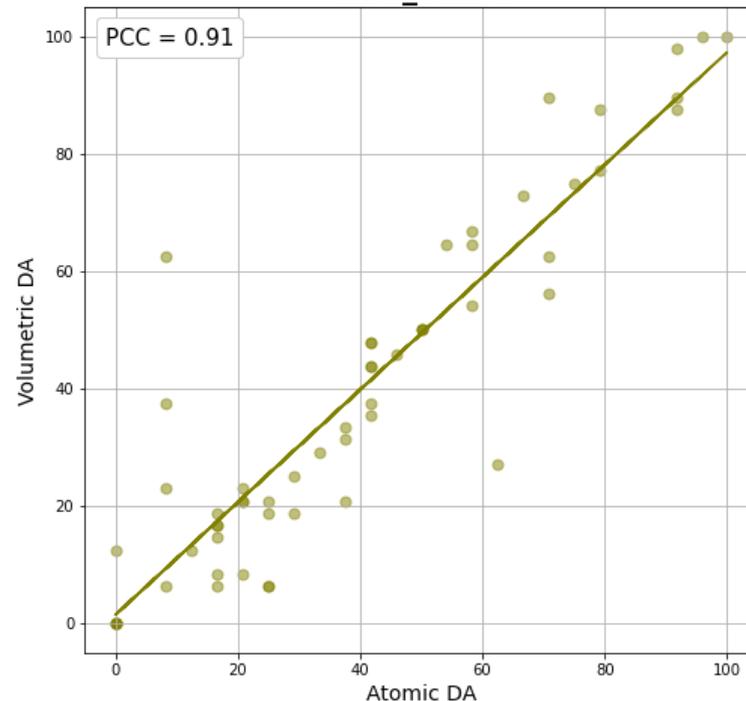
GalaxyDock3
 ECCD
 AtomicDA
 GD3_score



GalaxyDock3
 ECCD
 VolumetricDA
 GD3_score



GalaxyDock3
 ECCD
 VolumetricDA vs AtomicDA
 GD3_score



PCC MEDIO = 0,85

COEFFICIENTE DI CORRELAZIONE DI PEARSON (PCC)

$$PCC = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

Atomic DA
ECCD

Volumetric DA
ECCD

	Atomic DA ECCD									Volumetric DA ECCD								
	MeanRot Bonds	plp	plp95	chemplp	vinardo	vina	ad4	GD3_score	LeDock_score	MeanRot Bonds	plp	plp95	chemplp	vinardo	vina	ad4	GD3_score	LeDock_score
0																		
CASEIN_KINASE_II_ALPHA_SUBUNIT_1M2P	0.25	8.33	0.00	8.33	0.00	4.17	0.00	8.33	16.67	0.25	16.67	22.92	22.92	22.92	25.00	12.50	37.50	37.50
GALACTOCEREBROSIDASE	1.00	58.33	58.33	100.00	75.00	62.50	4.17	91.67	66.67	1.00	100.00	100.00	100.00	100.00	97.92	85.42	97.92	95.83
BETA-MANNOSIDASE	2.00	91.67	83.33	83.33	87.50	75.00	54.17	91.67	91.67	2.00	87.50	79.17	81.25	70.83	72.92	58.33	89.58	95.83
PROTEIN_POLYBROMO-1	2.25	83.33	70.83	62.50	83.33	87.50	70.83	58.33	58.33	2.25	70.83	52.08	58.33	77.08	79.17	58.33	54.17	62.50
HEAT_SHOCK_PROTEIN_HSP90-ALPHA	2.25	8.33	4.17	8.33	8.33	8.33	4.17	16.67	16.67	2.25	16.67	20.83	18.75	25.00	25.00	14.58	16.67	37.50
HISTONE-LYSINE_N-METHYLTRANSFERASE_NSD3	2.25	45.83	33.33	58.33	33.33	45.83	33.33	58.33	83.33	2.25	58.33	58.33	66.67	54.17	62.50	64.58	64.58	97.92
ALPHA-MANNOSIDASE_II	2.25	87.50	75.00	62.50	33.33	50.00	12.50	8.33	33.33	2.25	83.33	75.00	70.83	45.83	50.00	43.75	62.50	50.00
TRANSPORTER	2.50	41.67	45.83	41.67	33.33	29.17	4.17	41.67	45.83	2.50	39.58	45.83	39.58	22.92	25.00	22.92	47.92	47.92
ANDROGEN_RECEPTOR	2.50	41.67	58.33	41.67	54.17	58.33	33.33	54.17	66.67	2.50	54.17	64.58	60.42	58.33	66.67	60.42	64.58	70.83
POLY-ADP-RIBOSE_POLYMERASE_1	2.75	58.33	58.33	75.00	45.83	37.50	41.67	70.83	83.33	2.75	64.58	60.42	62.50	62.50	50.00	45.83	56.25	89.58
CEREBLON_ISOFORM_4	2.75	58.33	58.33	62.50	50.00	58.33	45.83	50.00	75.00	2.75	50.00	50.00	50.00	62.50	50.00	60.42	50.00	58.33
METALLO-BETA-LACTAMASE_TYPE_2	3.00	8.33	4.17	8.33	4.17	4.17	0.00	0.00	33.33	3.00	12.50	0.00	0.00	0.00	0.00	0.00	0.00	31.25
(...)	(...)									(...)								

**UTILIZZO CONGIUNTO DELLE DUE DA%
PER OTTENERE I MIGLIORI RISULTATI**

**I DATI PRODOTTI DAL MODELLO
SONO AFFIDABILI?**



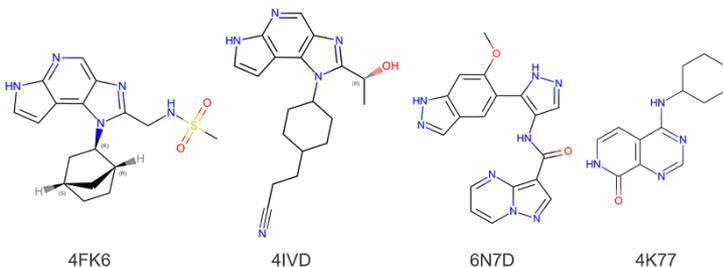
Test di affidabilità dei dati

Volumetric DA
RCCD

	MeanRot Bonds	plp	plp95	chemplp	vinardo	vina	ad4	GD3_score	LeDock_score
0									
TYROSINE-PROTEIN_KINASE_JAK1	3.00	54.17	54.17	72.92	72.92	68.75	77.08	64.58	58.33



LIGANDI NON CONTENUTI NEL DATASET



	4FK6_key	4IVD_key	6N7D_key	4K77_key
	%VolOverlap	%VolOverlap	%VolOverlap	%VolOverlap
4E4L_lock	77,26	86,45	85,72	58,26
4EHZ_lock	89,58	83,1	88,08	84,13
4IVC_lock	79,69	93,43	80,58	42,92
6N78_lock	80,93	73,14	81,26	54,54

$$\%VolDA = \frac{10+0,75 \cdot 3}{16} = 76,56$$

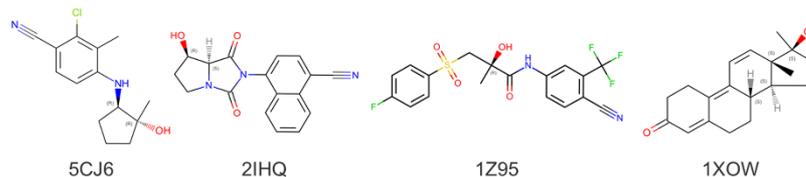
**DIFFERENZA
PERCENTUALE = 0,5%**

Volumetric DA
RCCD

	MeanRot Bonds	plp	plp95	chemplp	vinardo	vina	ad4	GD3_score	LeDock_score
0									
ANDROGEN_RECEPTOR	2.50	62.50	70.83	64.58	64.58	66.67	50.00	60.42	58.33



LIGANDI NON CONTENUTI NEL DATASET



	5CJ6_key	2IHQ_key	1Z95_key	1XOW_key
	%VolOverlap	%VolOverlap	%VolOverlap	%VolOverlap
1E3G_lock	76,65	80,6	95,05	76,69
1I37_lock	73,64	68,92	92,37	75,48
3B66_lock	62,06	65,16	92,52	57,88
3G0W_lock	83,73	84,54	96,22	78,16

$$\%VolDA = \frac{7+0,75 \cdot 5}{16} = 67,19$$

**DIFFERENZA
PERCENTUALE = 3,6%**

Le potenzialità di 3D-QSAR.com

Applications

Click on any button to start using the platform. If you want to know a bit more about the application, visit the link below to see all you will be able to do.

Explore Apps

- Py-MolEdit
- Py-ConfSearch
- Py-Align
- Py-CoMFA
- Py-ComBinE
- Py-Docking
- Py-Pharm NEW
- Py-PDB NEW
- Py-Graph NEW

- ✓ Preparazione automatica dei ligandi e delle proteine
- ✓ Editor per la correzione delle strutture cristallografiche
- ✓ Esecuzione del docking e calcolo dei parametri di validazione
 - ✓ Organizzazione dei risultati prodotti

**OTTIMIZZAZIONE E SEMPLIFICAZIONE
DEL CROSS-DOCKING**

CONCLUSIONI

$$DA\% = \frac{f_{\%Vol.Overlap \geq 80} + 0,75f_{70 \leq \%Vol.Overlap < 80}}{f_{TOT}}$$

DA% VOLUMETRICA
BASATA SULLA FORMA MOLECOLARE



VALIDAZIONE PIU' ACCURATA
DEL MOLECULAR DOCKING



**VIRTUAL
SCREENING**
AUMENTO DELL'HIT RATE



**RELAZIONI
STRUTTURA-ATTIVITA'**
STRUCTURE BASED DRUG DESIGN



**TOSSICOLOGIA
PREDITTIVA**
MINORI EFFETTI AVVERSI

WORK IN PROGRESS
«SUPER PROGRAMMA» DI VALIDAZIONE

**GRAZIE
PER L'ATTENZIONE**