

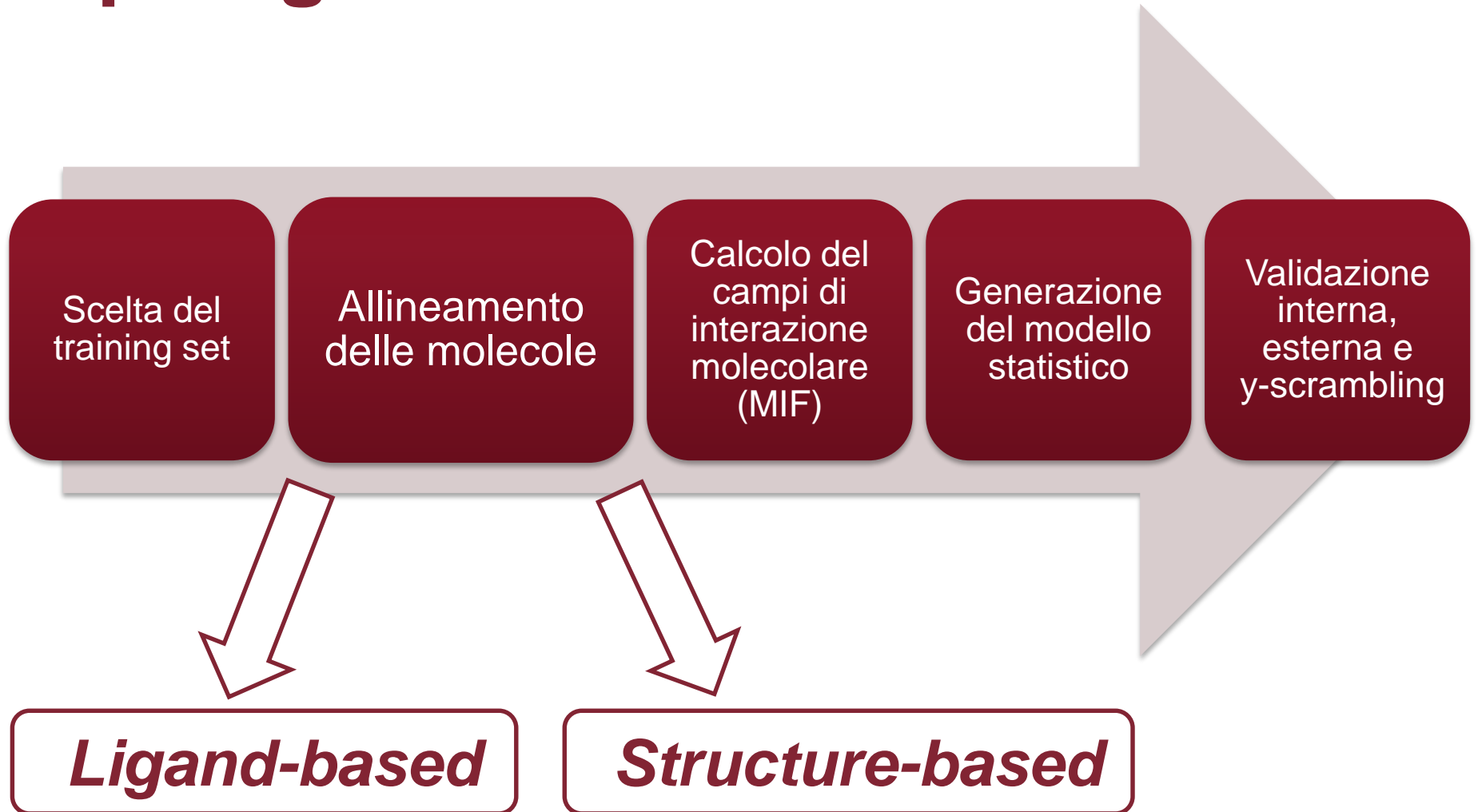
Chimica Farmaceutica

DOCKING and Structure Based Drug Design



SAPIENZA
UNIVERSITÀ DI ROMA

Tipologie di allineamento



Cos'è il docking?

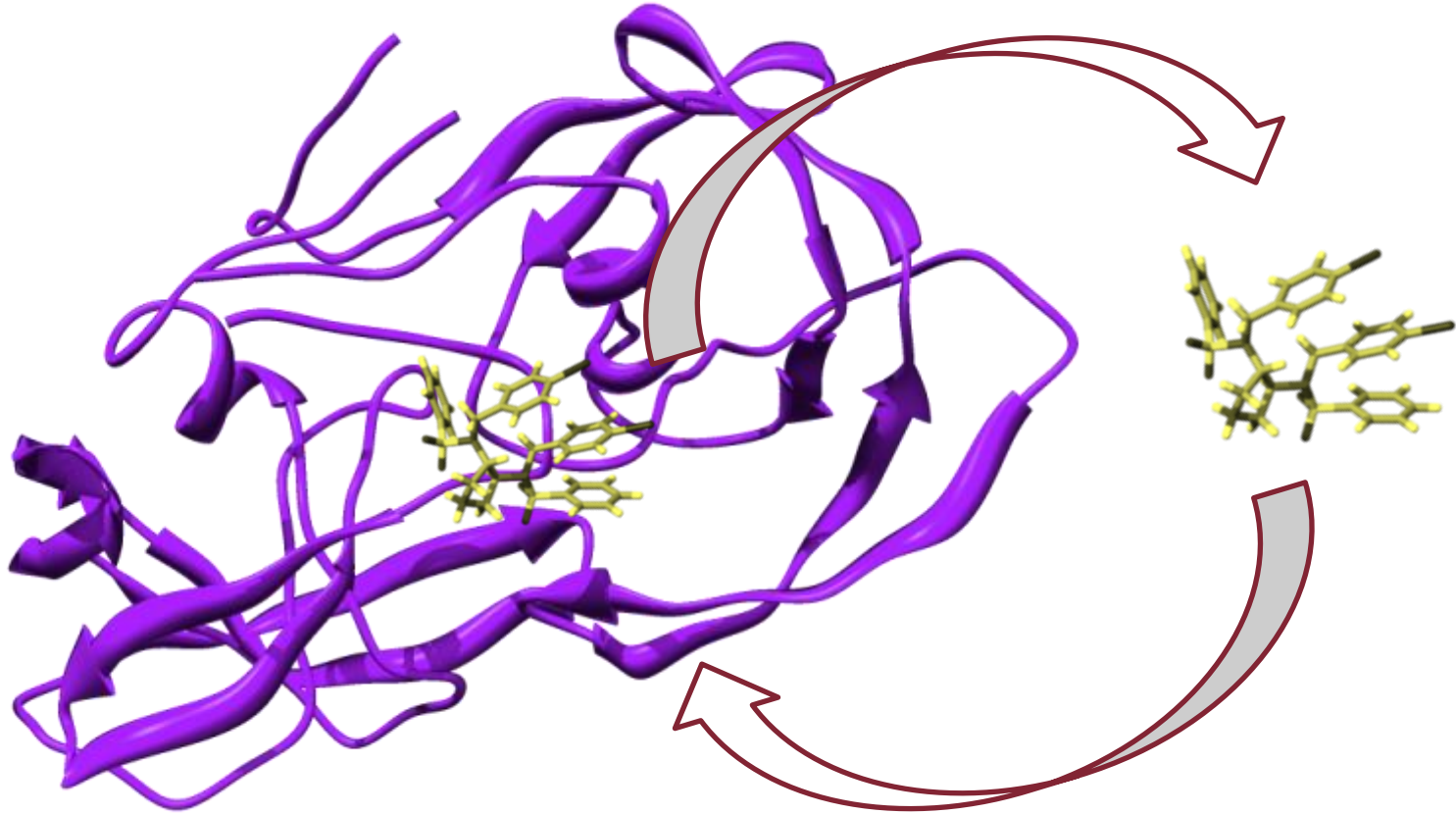
Il *docking* è una tecnica *structure-based*, mediante il quale si '*docka*' una piccola molecola all'interno della struttura *tridimensionale* della macromolecola *target*.



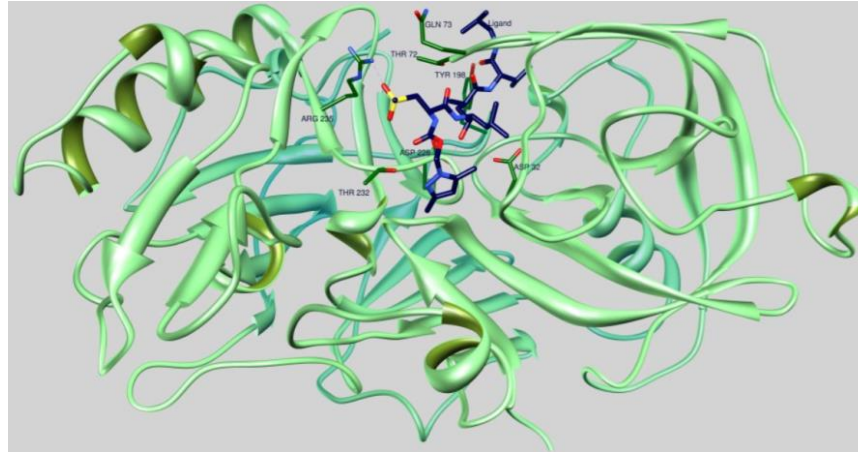
Due sono i principali scopi:

- Modelling strutturale;
- Predire l'attività biologica di un composto;

Cosa vuol dire dockare?



MOLECULAR DOCKING: perché viene usato



Identificazione di composti di
interesse farmaceutico

Comprensione meccanismi
interazione

Ottimizzazione del *lead*

Approccio Structure-based



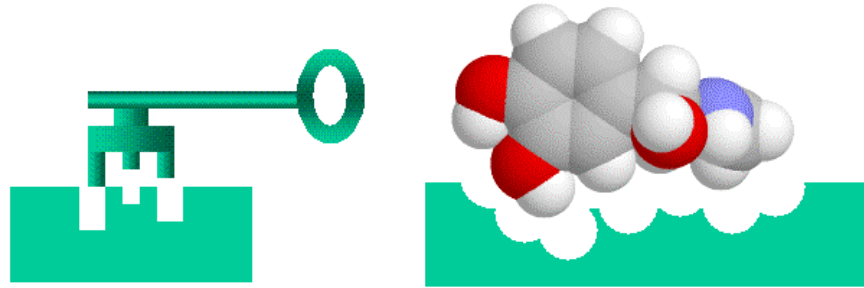
“This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.”

A screenshot of the RCSB PDB website homepage. The browser address bar shows 'www.rcsb.org/pdb/home/home.do'. The page features a navigation menu with options like 'Deposit', 'Search', 'Visualize', 'Analyze', 'Download', 'Learn', and 'More'. A search bar is prominently displayed with the text 'Search by PDB ID, author, macromolecule, sequence, or ligands'. Below the search bar, there are social media icons and a 'Feedback' button. The main content area includes a 'Welcome' section, a 'A Structural View of Biology' section with a paragraph describing the resource, and a 'November Molecule of the Month' section featuring a 3D model of a Glutamate-gated Chloride Receptor with labels for 'glutamate' and 'ivermectin'. A 'BEGIN EXPLORING' button is visible in the 'Take an Interactive Tour of the PDB' section.

Ricognizione molecolare: Insieme di interazioni tra molecole e macromolecole

Modello “*Lock and key*”

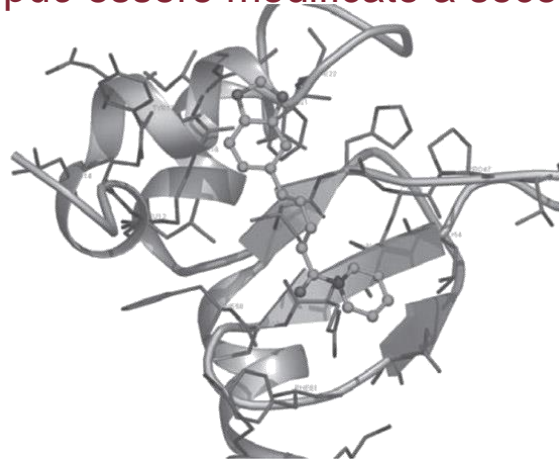
La proteina ha una sua conformazione all'interno del quale il ligando “fitta” perfettamente. Emil Fischer(1890).



Modello del *fitting* indotto

Il sito attivo della macromolecola può essere modificato a seconda di come interagisce il ligando.

Daniel Koshland 1958.



Un protocollo di docking è costituito da due fasi:

➤ Analisi del ligando e della proteina;

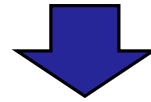
Metodi sistematici	Esplorazione combinatoriale dei gradi di libertà, Costruzione incrementale
Metodi stocastici	Monte Carlo, algoritmo genetico
Metodi di simulazione	Minimizzazione, Dinamica Molecolare

➤ Calcolo dello score dell'interazione tra ligando e proteina (Scoring function);

Force fields scoring function
Scoring function empiriche
Knowledge scoring function

ANALISI DEL LIGANDO E DELLA PROTEINA

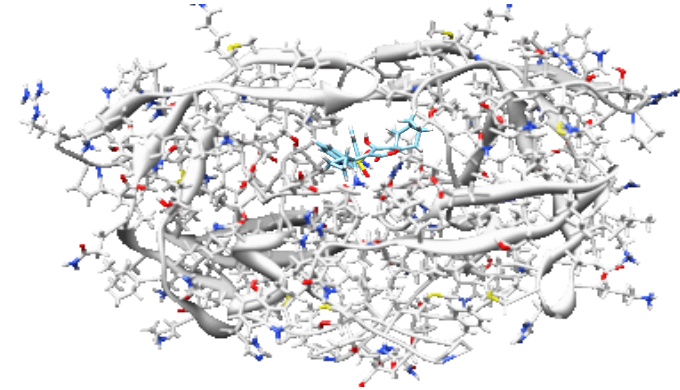
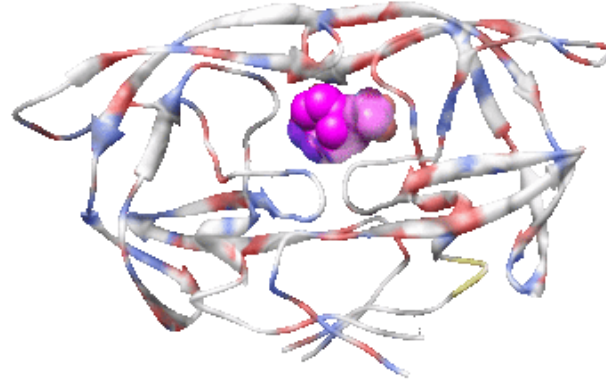
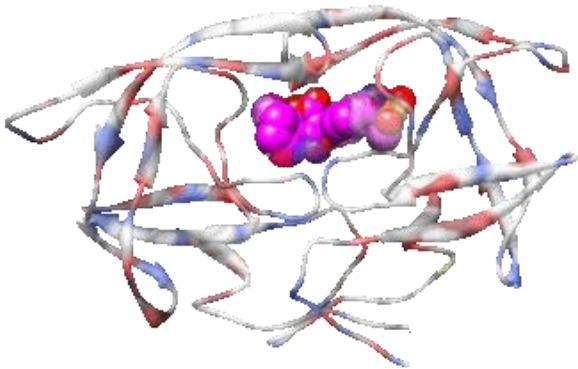
Docking



Rigido

Semi-Flessibile

Flessibile



Un protocollo di docking è costituito da due fasi:

➤ Analisi del ligando e della proteina;

Metodi sistematici	Esplorazione combinatoriale dei gradi di libertà, Costruzione incrementale
Metodi stocastici	Monte Carlo, algoritmo genetico
Metodi di simulazione	Minimizzazione, Dinamica Molecolare

- **Metodi Sistematici** → molecola viene divisa in una porzione rigida, chiamata *core* che viene dockata all'interno del sito catalitico e successivamente vengono addizionati gli altri frammenti studiati mediante l'algoritmo di Hemmerhead;

Un protocollo di docking è costituito da due fasi:

➤ Analisi del ligando e della proteina;

Metodi sistematici	Esplorazione combinatoriale dei gradi di libertà, Costruzione incrementale
Metodi stocastici	Monte Carlo, algoritmo genetico
Metodi di simulazione	Minimizzazione, Dinamica Molecolare

- **Metodi Sistematici** → molecola viene divisa in una porzione rigida, chiamata *core* che viene dockata all'interno del sito catalitico e successivamente vengono addizionati gli altri frammenti studiati mediante l'algoritmo di Hemmerhead;
- **Metodi stocastici** → Molecola viene sottoposta a cambiamenti random, e successivamente mantenuta o eliminata in base ad una funzione di probabilità;

Un protocollo di docking è costituito da due fasi:

➤ Analisi del ligando e della proteina;

Metodi sistematici	Esplorazione combinatoriale dei gradi di libertà, Costruzione incrementale
Metodi stocastici	Monte Carlo, algoritmo genetico
Metodi di simulazione	Minimizzazione, Dinamica Molecolare

- **Metodi Sistematici** → molecola viene divisa in una porzione rigida, chiamata *core* che viene dockata all'interno del sito catalitico e successivamente vengono addizionati gli altri frammenti studiati mediante l'algoritmo di Hemmerhead;
- **Metodi stocastici** → Molecola viene sottoposta a cambiamenti random, e successivamente mantenuta o eliminata in base ad una funzione di probabilità;
- **Metodi di simulazione** → Viene risolta l'equazione del moto di Newton ($F=am$) per studiare le traiettorie degli atomi all'interno di un sistema, delle diverse traiettorie si studia anche l'energia potenziale.

Un protocollo di docking è costituito da due fasi:

➤ Analisi del ligando e della proteina;

Metodi sistematici	Esplorazione combinatoriale dei gradi di libertà, Costruzione incrementale
Metodi stocastici	Monte Carlo, algoritmo genetico
Metodi di simulazione	Minimizzazione, Dinamica Molecolare

➤ Calcolo dello score dell'interazione tra ligando e proteina (Scoring function);

Force fields scoring function
Scoring function empiriche
Knowledge scoring function

Un protocollo di docking è costituito da due fasi:

- Calcolo dello score dell'interazione tra ligando e proteina (Scoring function);

Scoring function empiriche
Knowledge scoring function
Force fields scoring function

- **Scoring function empiriche** → lo score viene calcolato usando la somma di diversi parametri non correlati tra loro e sulla quale viene applicata la statistica della regressione lineare.

Un protocollo di docking è costituito da due fasi:

- Calcolo dello score dell'interazione tra ligando e proteina (Scoring function);

Scoring function empiriche
Knowledge scoring function
Force fields scoring function

- **Scoring function empiriche** → lo score viene calcolato usando la somma di diversi parametri non correlati tra loro e sulla quale viene applicata la statistica della regressione lineare.
- **Knowledge scoring function** → in base alle osservazioni fatte sui cristalli presenti in bancadati riguardo i parametri caratterizzanti un certo sistema si cerca di determinare l'affinità tra ligando-recettore.

Un protocollo di docking è costituito da due fasi:

- Calcolo dello score dell'interazione tra ligando e proteina (Scoring function);

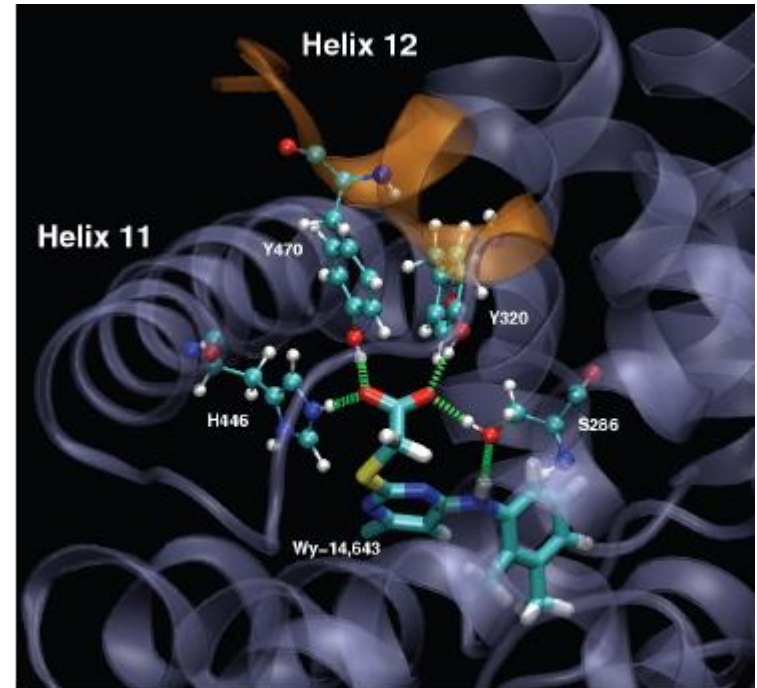
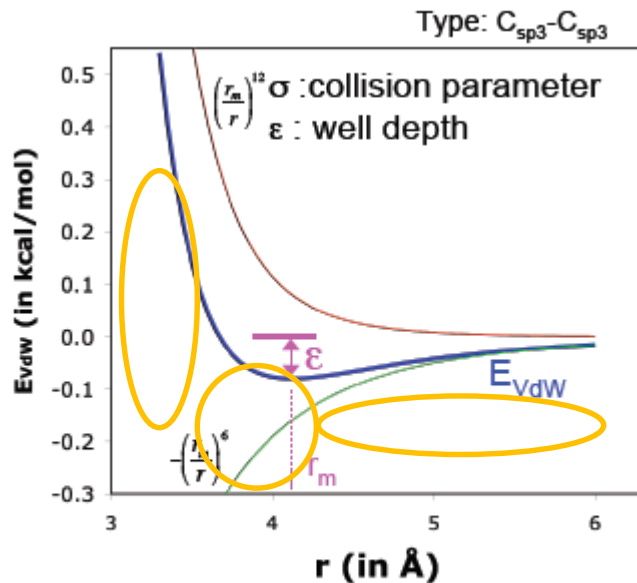
Scoring function empiriche
Knowledge scoring function
Force fields scoring function

- **Scoring function empiriche** → lo score viene calcolato usando la somma di diversi parametri non correlati tra loro e sulla quale viene applicata la statistica della regressione lineare.
- **Knowledge scoring function** → in base alle osservazioni fatte sui cristalli presenti in bancadati riguardo i parametri caratterizzanti un certo sistema si cerca di determinare l'affinità tra ligando-recettore.
- **Force field scoring function** → quantifica la somma di due energie, l'energia di interazione tra recettore e ligando e l'energia interna del ligando.

Forze di van der Waals

Debole attrazione intermolecolare causata da dipoli molecolari indotti.

$$E_{\text{vdW}} = \epsilon \left[\left(\frac{r_m}{r} \right)^{12} - 2 \left(\frac{r_m}{r} \right)^6 \right]$$



Es. Interazione tra il ligando Wy-14,643 e il sito attivo della PPAR α

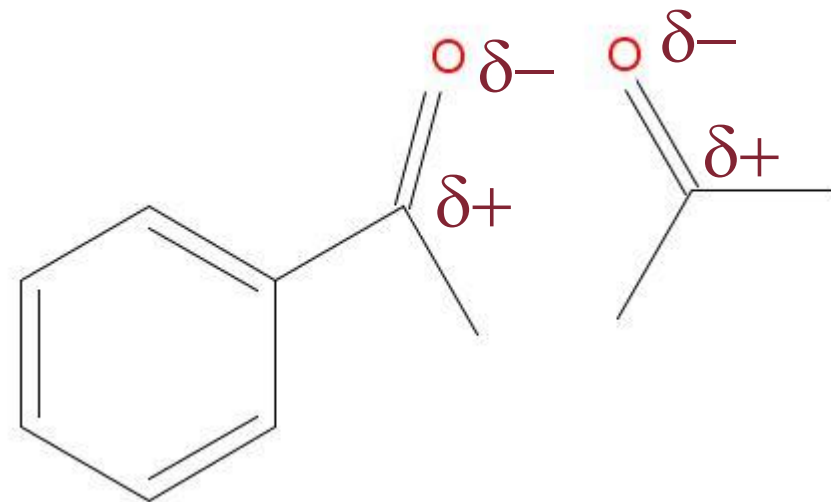
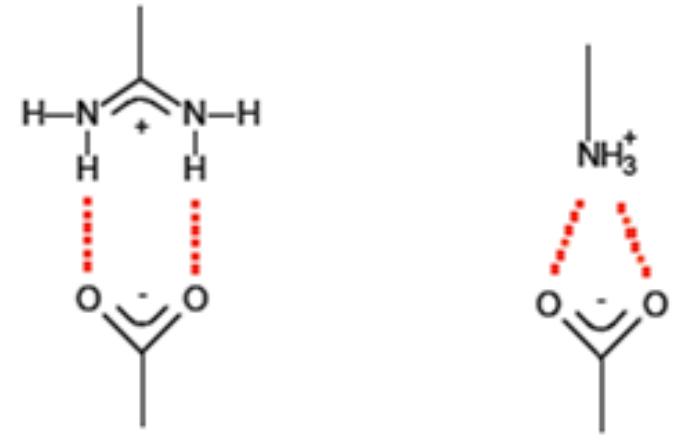
Interazioni Elettrostatiche

La legge di Coulomb

In un mezzo isolante diverso dall'aria la forza F , a parità di cariche e di distanza, risulta generalmente minore. La forza di Coulomb si scrive allora:

$$\vec{F} = \frac{1}{4\pi \cdot \epsilon_0 \cdot \epsilon_r} \cdot \frac{q_1 \cdot q_2}{r^2} \cdot \hat{r}$$

dove ϵ_r è un numero che prende il nome di costante dielettrica relativa (permittività elettrica relativa). Il prodotto $\epsilon_r \times \epsilon_0$ si indica con ϵ , e si chiama costante dielettrica del mezzo (permittività del mezzo).



DOCKING

Program		Vendor	Ligand flexibility		Protein flexibility	Scoring function
Name	Version		Type	Algorithm		
Dock	5.4	UCSF	Systematic	Incremental build	-	Force field or contact score (grid)
Autodock	3.0/4.0	Scripps	Stochastic	Genetic Algorithm	-	Force field (AMBER)
Gold		CCDC	Stochastic	Genetic Algorithm	Flexible OH function of Ser, Thr, Tyr and NH3+ of Lys	Empirical score (Chemscore, Goldscore)
FlexX FlexE	2.0.2	BioSolveIT	Systematic	Incremental build	Ensemble of protein structures	Empirical score (Adapted LUDI score)
Glide		Schrodinger	Systematic	Exhaustive search	-	Empirical score
ICM	3	Molsoft	Stochastic / Deterministic	Pseudo-Brownian sampling and local minimization	Flexible side chains	Force field and Empirical score (Grid)
QXP			Stochastic	Monte Carlo	-	Force field (grid)
Hammerhead			Systematic	Incremental build	-	Empirical score

Come si sviluppa un protocollo di Docking?

ECRD

RCRD

ECCD

RCCD

- Plants (chemplp, plp, plp95)
- AutoDock
- Vina
- Surflex
- ParaDock (PMF, Pscore)

BC

BF

BD

- Clustering

Download, modelling,
preparazione della lock e
minimizzazione dei complessi

Allineamento SB mediante
MatchMaker

Docking

Analisi dei risultati

PULIZIA DELLE PROTEINE

Eliminazione solvente

Eliminazione metalli strutturali

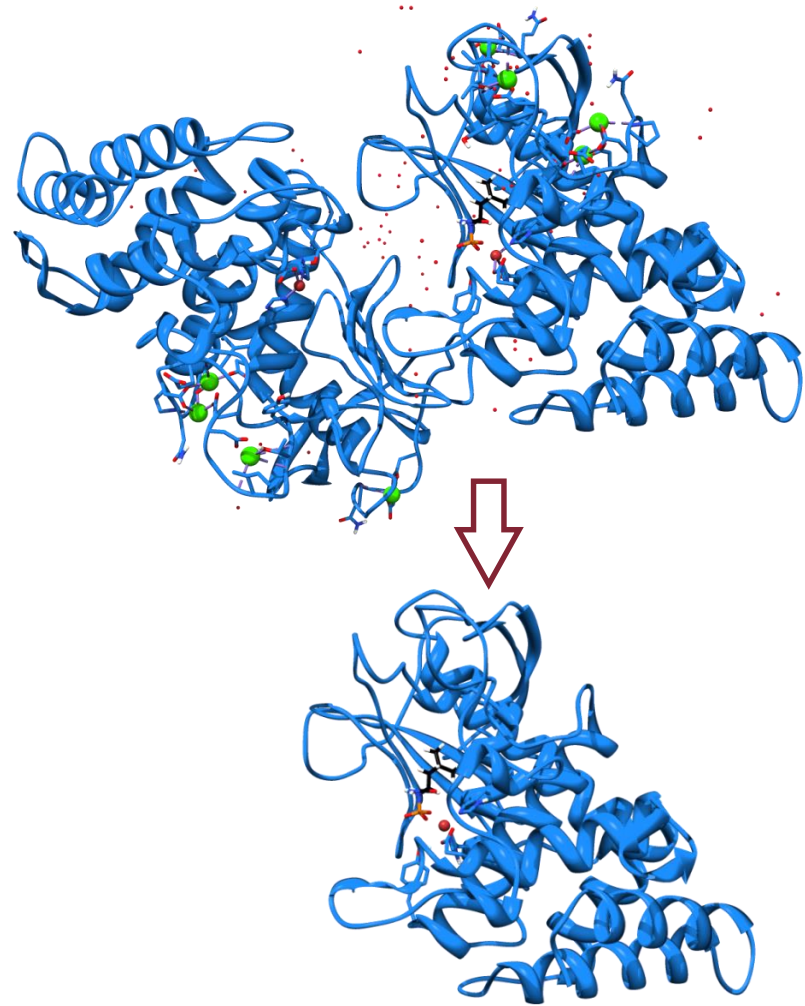
Eliminazione catene inutili

Ricostruzione amminoacidi tagliati

Protonazione

Cariche amminoacidi

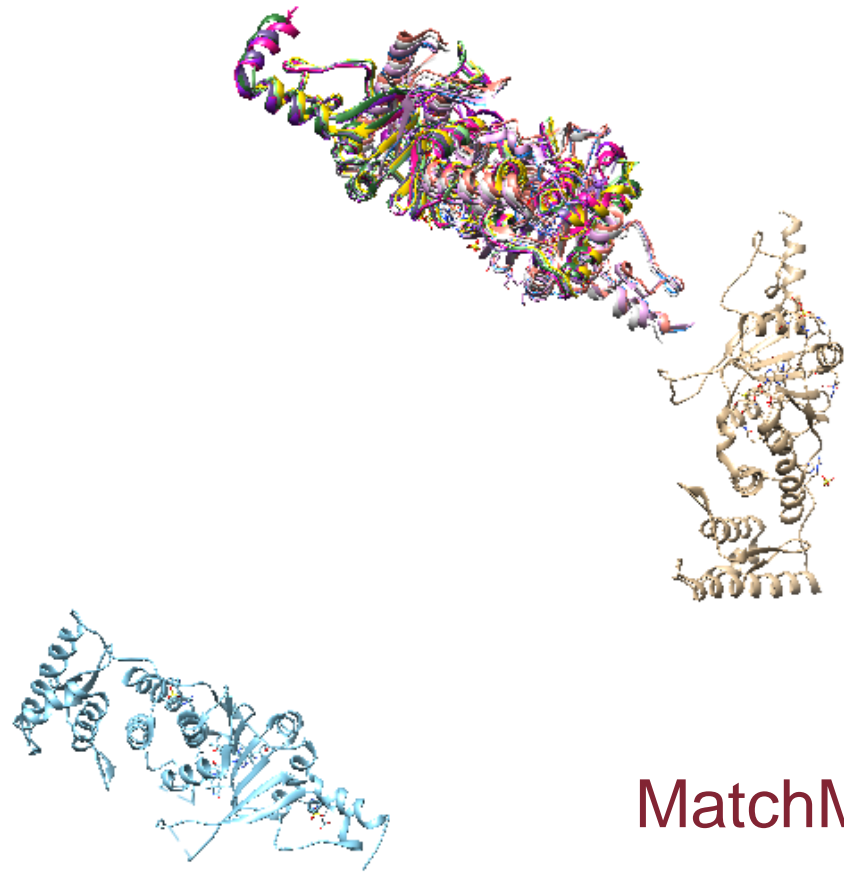
Allineamento Complessi



Allineamento SB mediante MatchMaker

DOT1L Disruptor of Telomeric Silencing 1-like

Codice ID	Risoluzione (Å)
1NW3	2.50
3QOW	2.10
3QOX	2.30
3SR4	2.50
3SX0	2.28
3UWP	2.05
4EK9	2.50
4EKG	2.80
4EKI	2.85
4ER5	2.57
4EQZ	2.15
4ER0	2.50
4ER6	2.30
4ER7	2.20
4HRA	3.15



MatchMaker

Come si sviluppa un protocollo di Docking?

ECRD

RCRD

ECCD

RCCD

- Plants (chemplp, plp, plp95)
- AutoDock
- Vina
- Surflex
- ParaDock (PMF, Pscore)

BC

BF

BD

- Clustering

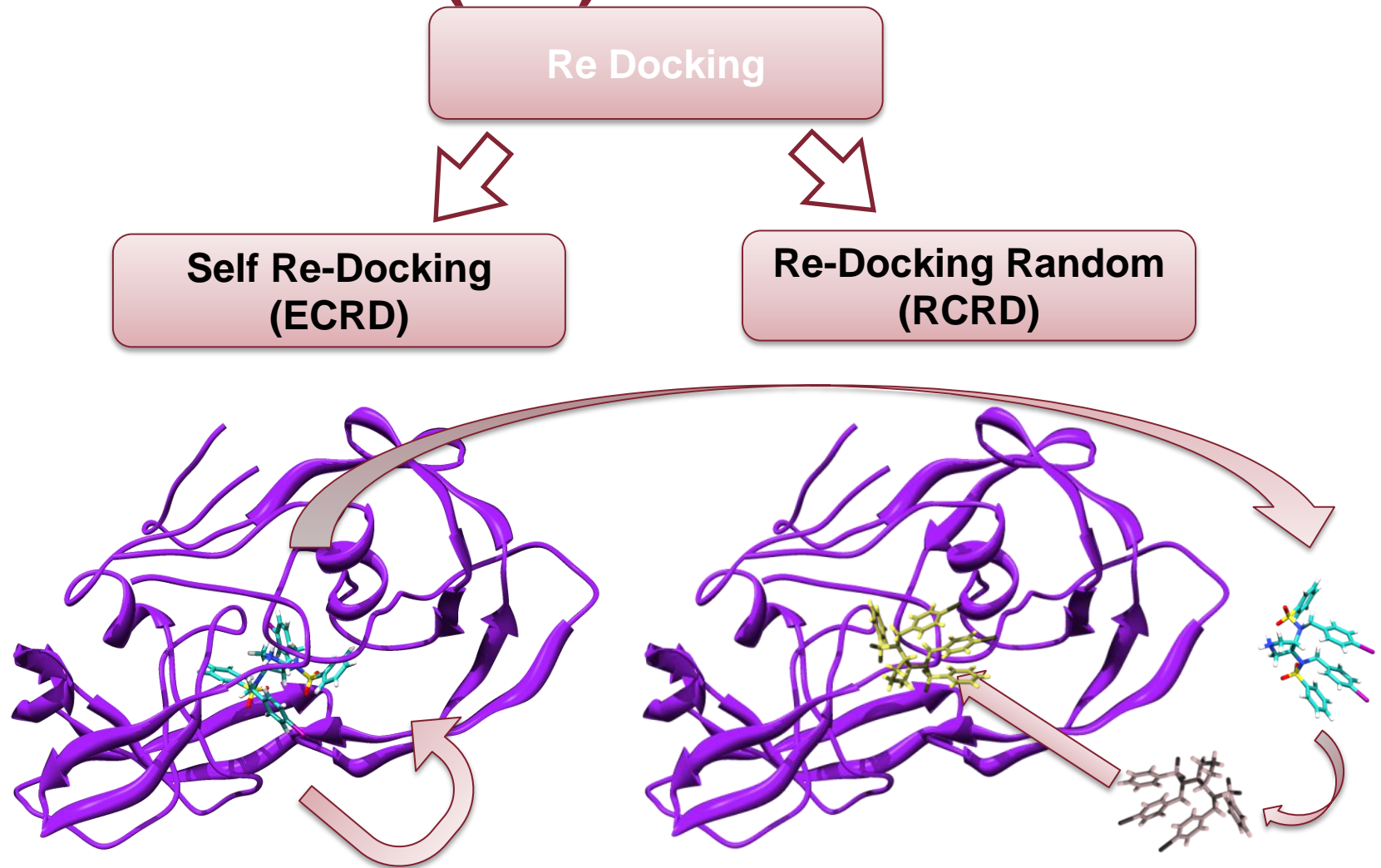
Download, modelling e
minimizzazione dei complessi

Allineamento SB mediante
MatchMaker

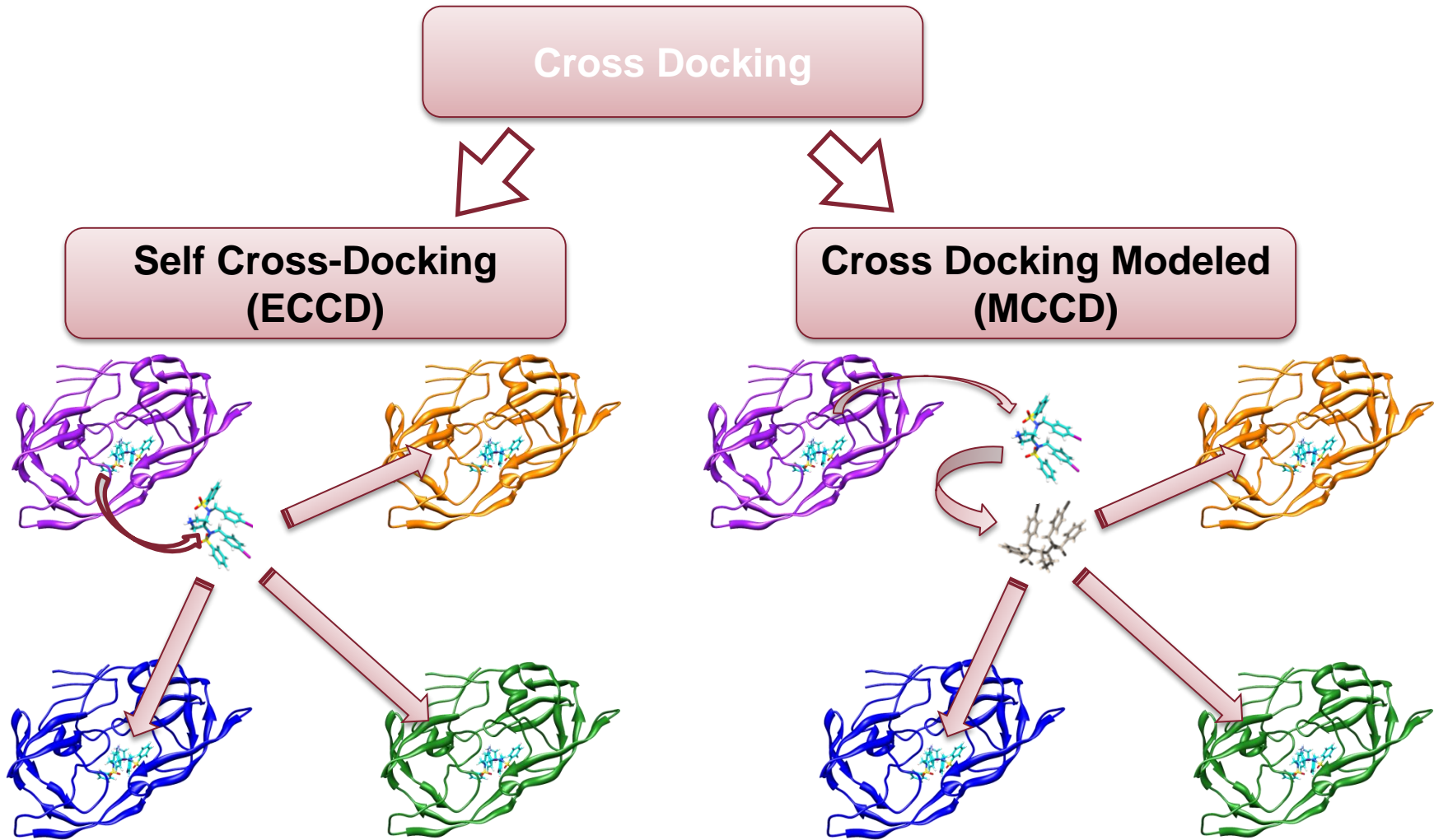
Docking

Analisi dei risultati

RE-DOCKING (RD)



CROSS-DOCKING (CD)



PANORAMICA

Surflex

Dock

Paradocks

Autodock

Plants

AutodockVina

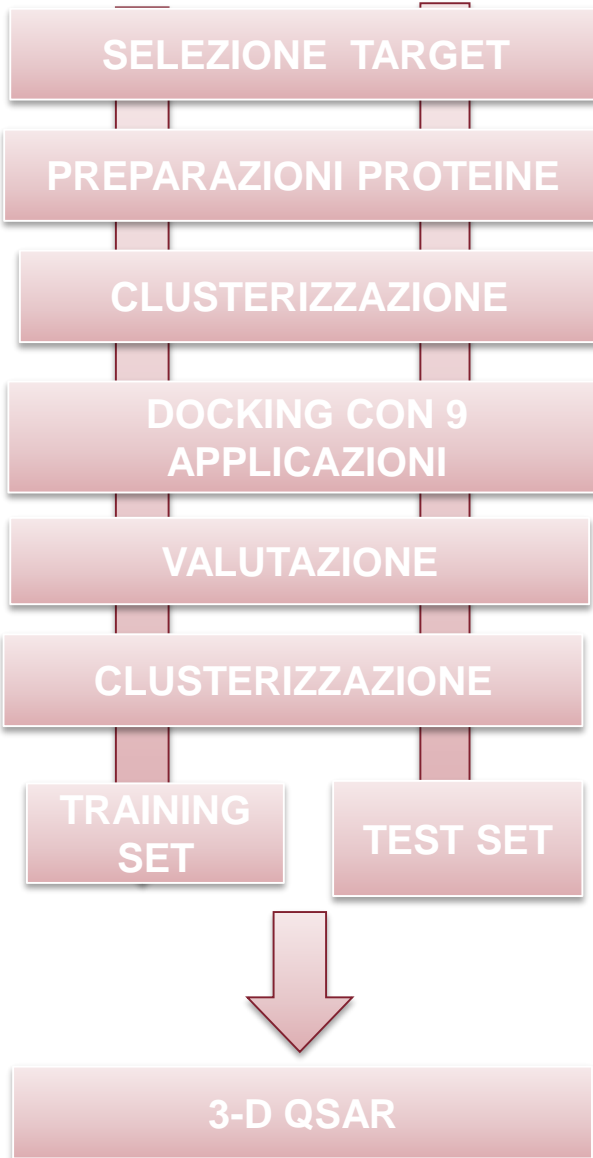
pmf04

p-score

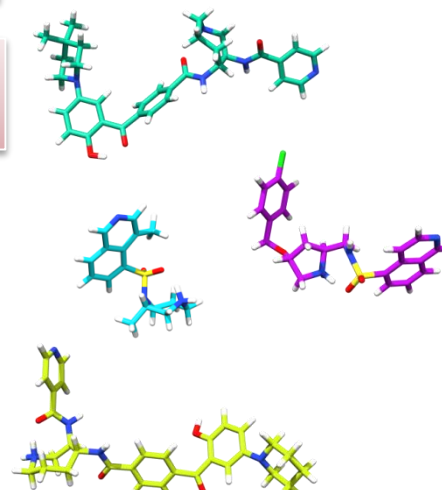
plp

chemplp

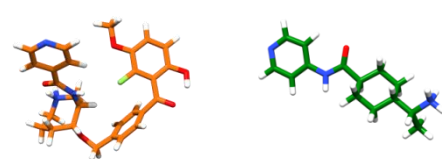
plp95



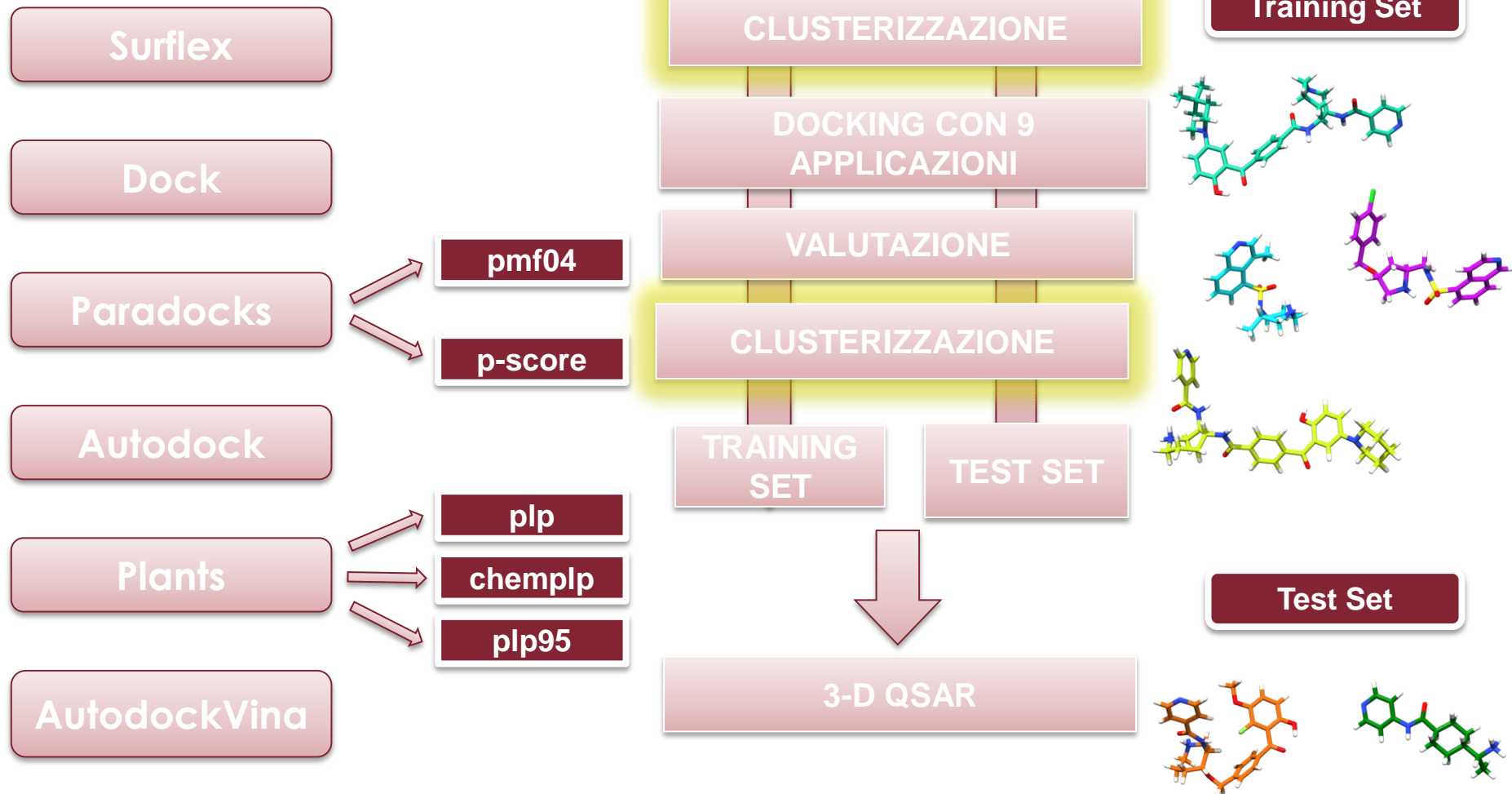
Training Set



Test Set



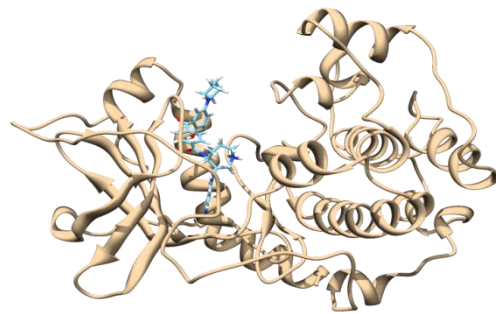
PANORAMICA



ALLINEAMENTO/CLUSTER

cAMP-dependent protein Kinase catalytic subunit alpha

Riunificazione
lock/key nel
complesso

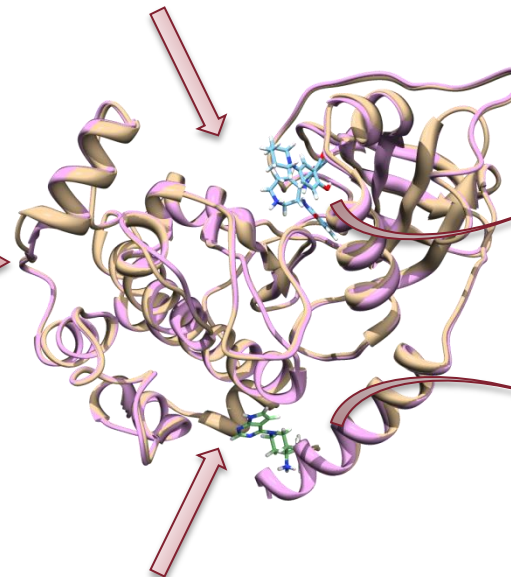


Allineamento



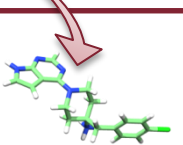
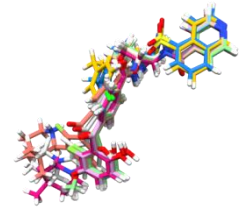
Proteine Pulite

Binding site 1



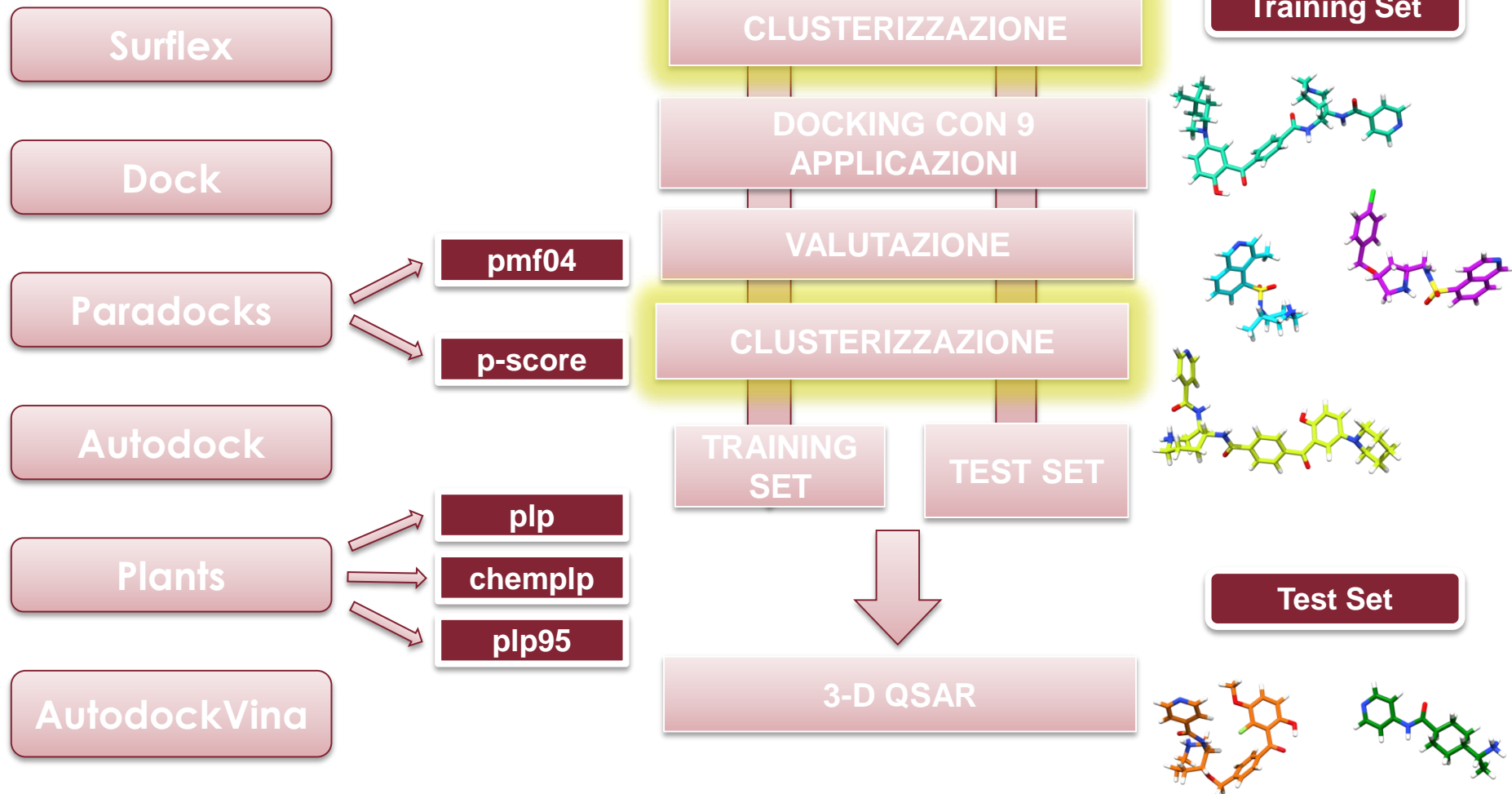
Binding site 2

Cluster 1



Cluster 2

PANORAMICA



CLUSTERIZZAZIONE

Predizione Binding mode



Clusterizzazione

➤ BF (Best Fitted)

➤ BC (Best Clustered)

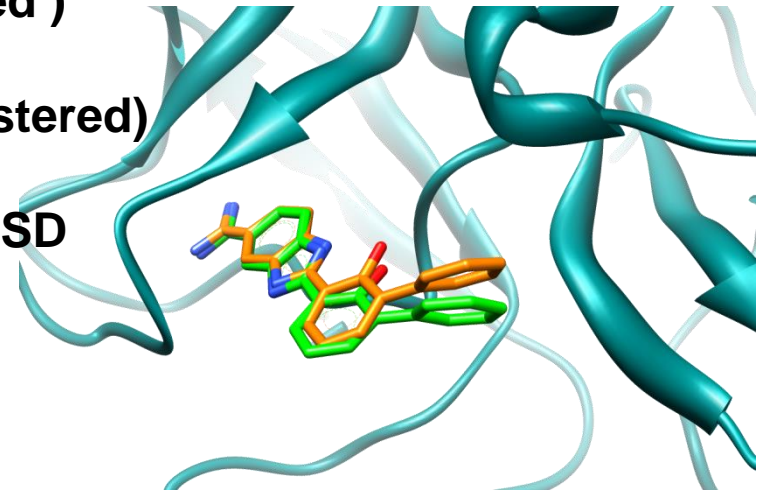
➤ BD (Best Docked)

➤ BF (Best Fitted)

➤ BC (Best Clustered)

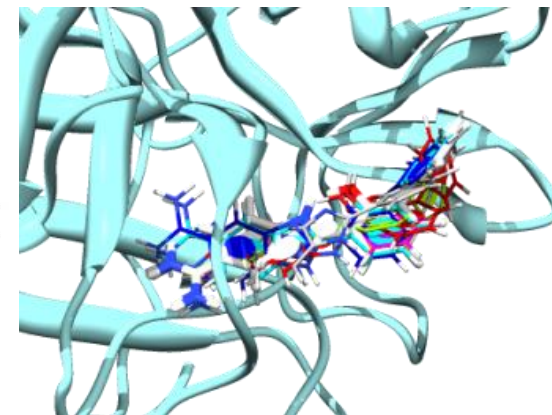
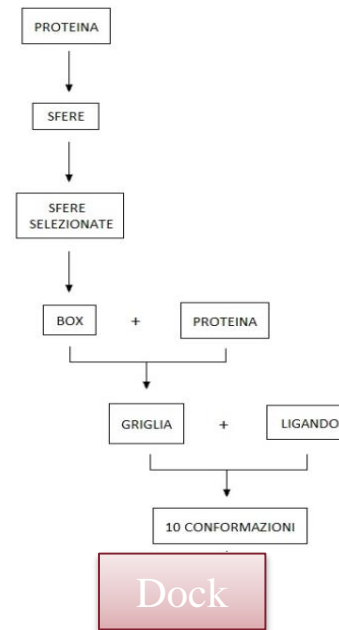
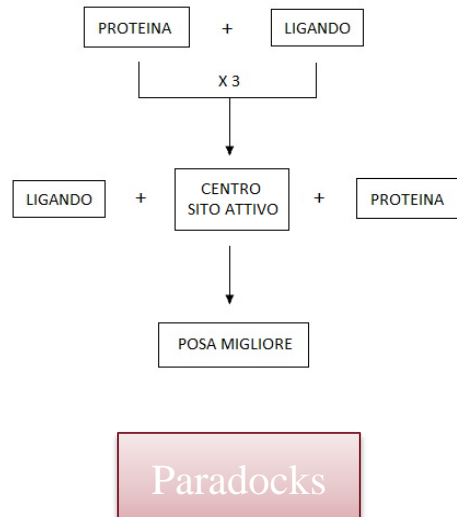
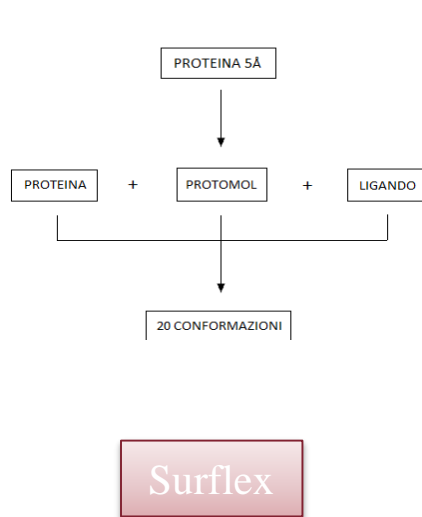
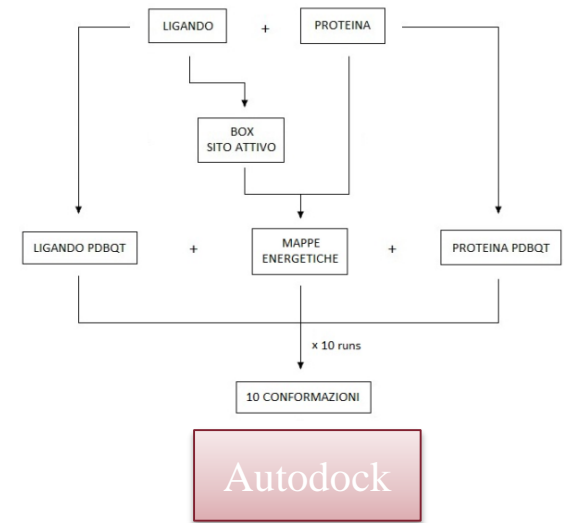
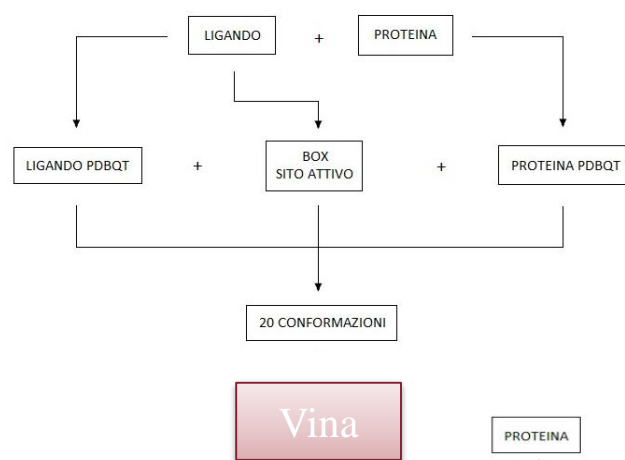
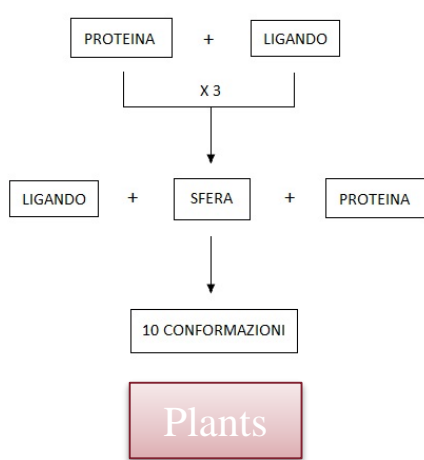
➤ Energia + RMSD

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^n (x_{1,i} - x_{2,i})^2}{n}}$$



■ originale ■ docking

PROGRAMMI USATI



EXPERIMENTAL vs MODELED

$$DA \text{ (Docking Accuracy)} = f_{RMSD \leq 2} + 0.5 (f_{RMSD \leq 3} - f_{RMSD \leq 2})$$

	ECRD	MCRD	ECRD vs MCRD
Software	DA%		$\Delta DA\%$
Surflex	32.75	18.20	-14.56
Paradocks pmf04	27.22	15.19	-12.03
Paradocks p-score	35.28	22.78	-12.50
Vina	46.84	35.08	-11.76
Dock	41.93	36.08	-5.85
Autodock	49.05	41.11	-7.94
Plants chemplp	57.91	47.31	-10.60
Plants plp	53.64	45.25	-8.39
Plants plp95	52.06	42.88	-9.18
Media	44.08	33.76	-10.31

EXPERIMENTAL vs MODELED

$$DA \text{ (Docking Accuracy)} = f_{RMSD \leq 2} + 0.5 (f_{RMSD \leq 3} - f_{RMSD \leq 2})$$

	ECCD		MCCD		ECCD vs MCCD
	BD	BC	BD	BC	BD
Software	DA%		DA%		Δ DA%
Surflex	52.85	49.84	48.10	45.41	-4.75
Paradocks pmf04	46.84	47.63	41.46	41.77	-5.38
Paradocks p-score	60.92	60.60	49.84	52.37	-11.08
Vina	69.78	57.59	65.19	53.64	-4.59
Dock	53.64	53.48	52.37	52.37	-1.27
Autodock	70.09	65.03	65.98	62.97	-4.11
Plants chemplp	72.47	70.73	68.35	64.40	-4.12
Plants plp	70.41	66.14	65.51	61.71	-4.91
Plants plp95	69.15	70.25	66.61	66.61	-2.53
Media	62.90	60.14	58.16	55.69	-4.75

RE-DOCKING vs CROSS-DOCKING

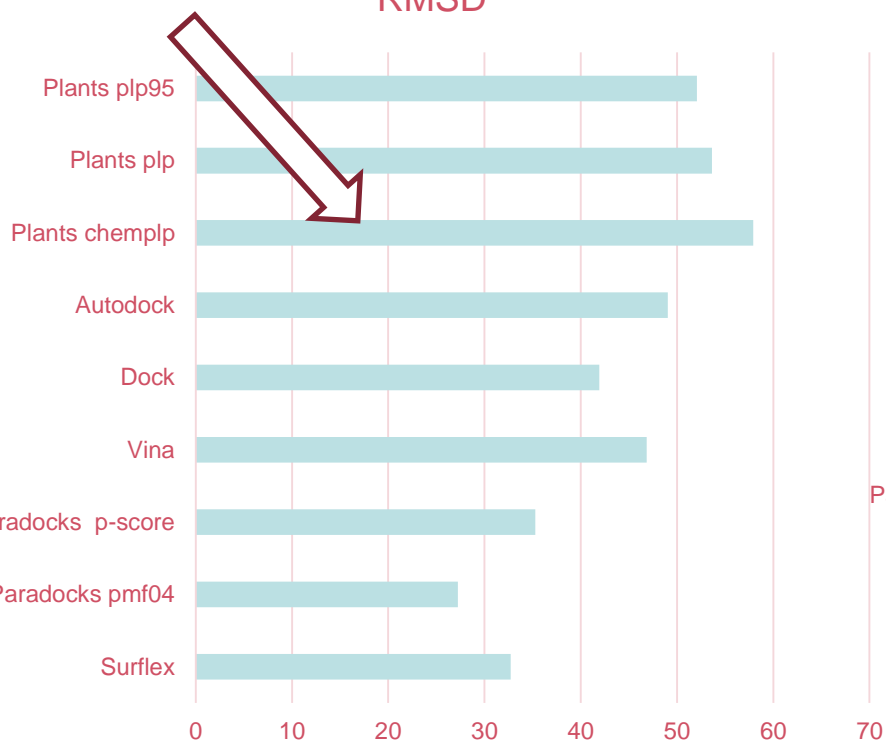
$$DA \text{ (Docking Accuracy)} = f_{RMSD \leq 2} + 0.5 (f_{RMSD \leq 3} - f_{RMSD \leq 2})$$

	ECRD	vs	BD-ECCD		MCRD	vs	BD-MCCD
Software			$\Delta DA\%$	Software			$\Delta DA\%$
Surflex			+20.09	Surflex			+29.91
Paradocks pmf04			+19.62	Paradocks pmf04			+26.27
Paradocks p-score			+25.63	Paradocks p-score			+27.06
Vina			+22.94	Vina			+30.11
Dock			+11.71	Dock			+16.30
Autodock			+21.04	Autodock			+24.87
Plants chemplp			+14.56	Plants chemplp			+21.04
Plants plp			+16.77	Plants plp			+20.25
Plants plp95			+17.09	Plants plp95			+23.73
Media			+18.83	Media			+24.36

Prestazioni singoli programmi

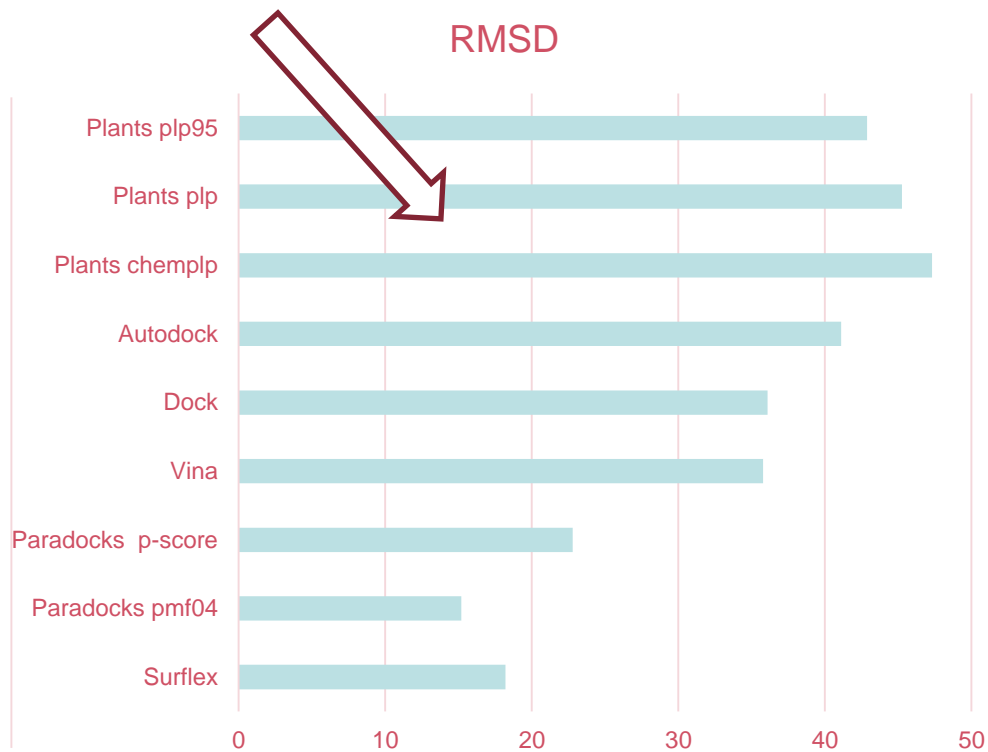
Re-Docking

RMSD



Re -Docking Modeled

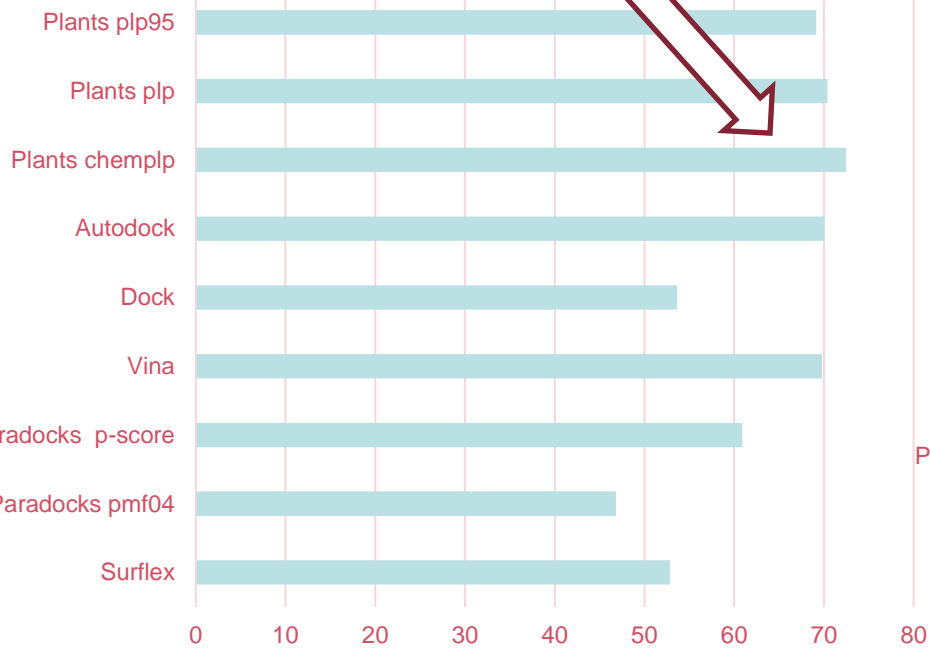
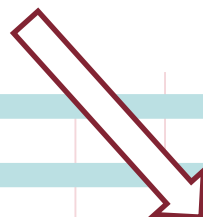
RMSD



Prestazioni singoli programmi

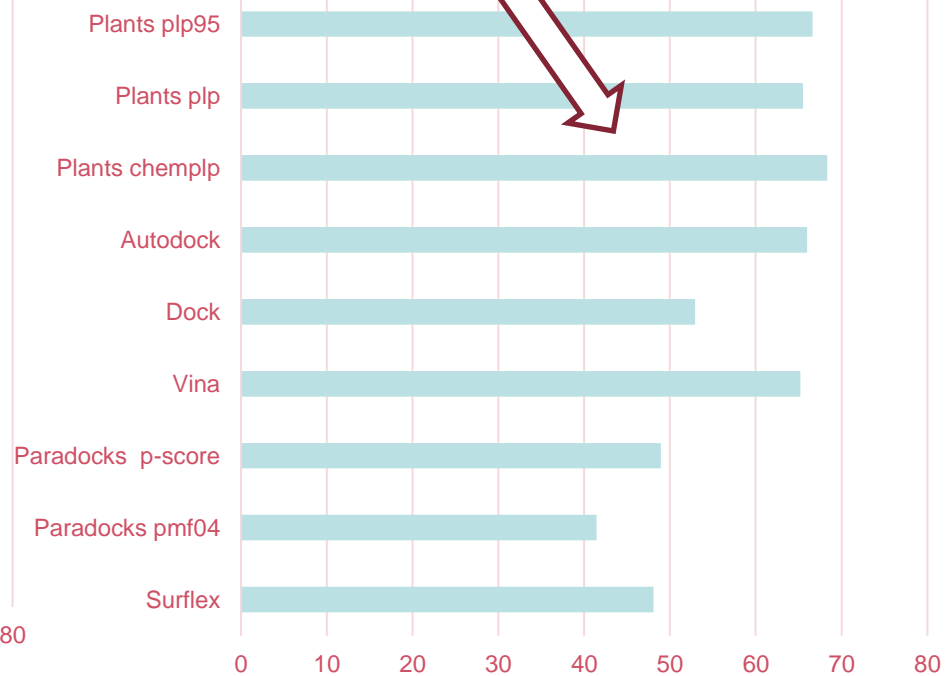
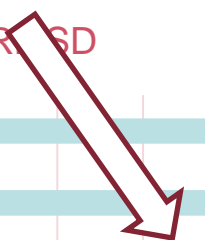
Cross-Docking

RMSD



Cross -Docking Modeled

RMSD



conclusioni

- RD soffre della conformazione di input maggiormente del CD
- In termini di *docking accuracy* il CD garantisce prestazioni maggiori di RD
- La scelta del programma mediante validazione con CD è risultata effettivamente predittiva
- Limitazioni
- Ulteriori studi sono in corso su un più vasto numero di complessi al fine di ulteriormente confermare i risultati ottenuti