

Chimica Farmaceutica

DOCKING and Structure Based Drug Design



SAPIENZA
UNIVERSITÀ DI ROMA



SBDD methods



3-D QSAR (Molecules aligned by SB methods)

Scoring Function (A Sort of QSARs for Molecular Docking)

COMBINE (A Full SB 3-D QSAR)

Proteochemometric (A QSAR SB derived method)

Molecular Dynamics

Pharmacophoric Approaches (Using Targets Structures)

Homology Modeling (To Build Target Structures)

Structure-Based Drug Design Approaches

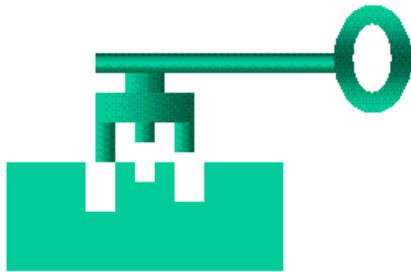
Ricognizione molecolare:

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.

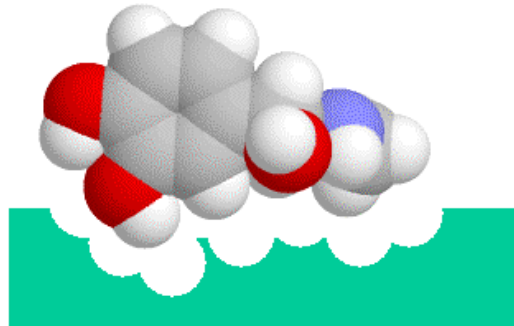
Emil Fischer(1890).



Modello del *fitting* indotto

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

Daniel Koshland 1958.



- Natura delle interazioni?
- Intensità della ricognizione molecolare?

Ricognizione molecolare

La complementarità molecolare altamente specifica tra key (ligando) e lock (recettore) gioca un ruolo chiave nei processi biologici.

La capacità del recettore di agganciarsi al suo ligando con alta specificità e affinità è dovuta alla formazione di una serie di legami deboli e interazioni favorevoli.

Interazioni specifiche ligando-recettore:

Interazioni elettrostatiche (Non-Cov)

Forze di van der Waals (Non-Cov)

Interazioni $\pi - \pi$

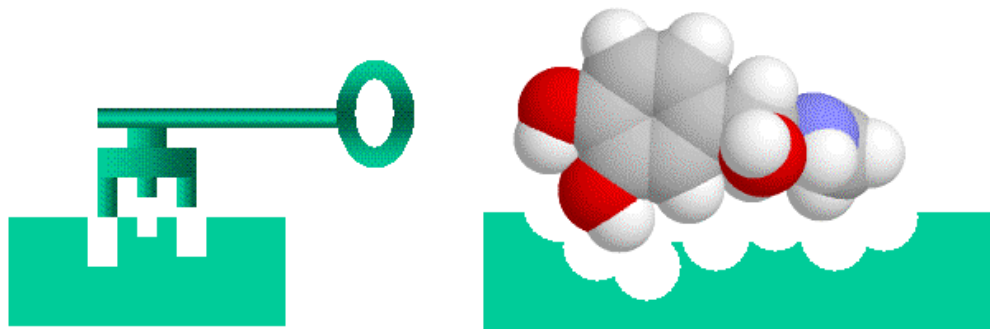
Coordinazione con Metalli

Interazioni idrofobiche

Effetti elettromagnetici



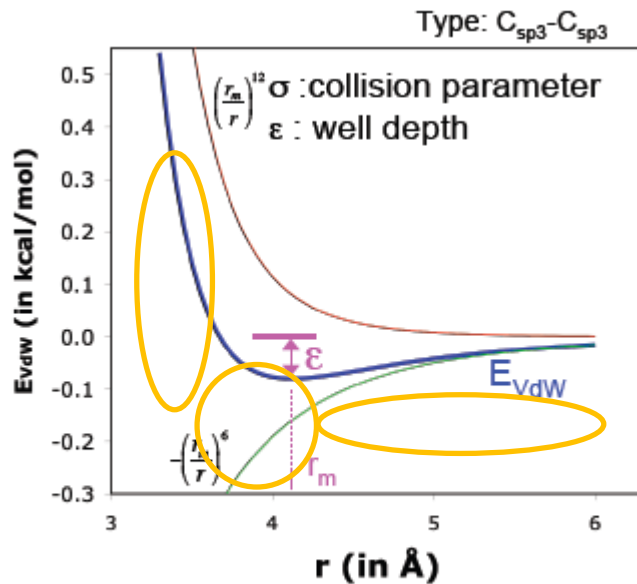
**COMPLEMENTARITA'
MOLECOLARE**



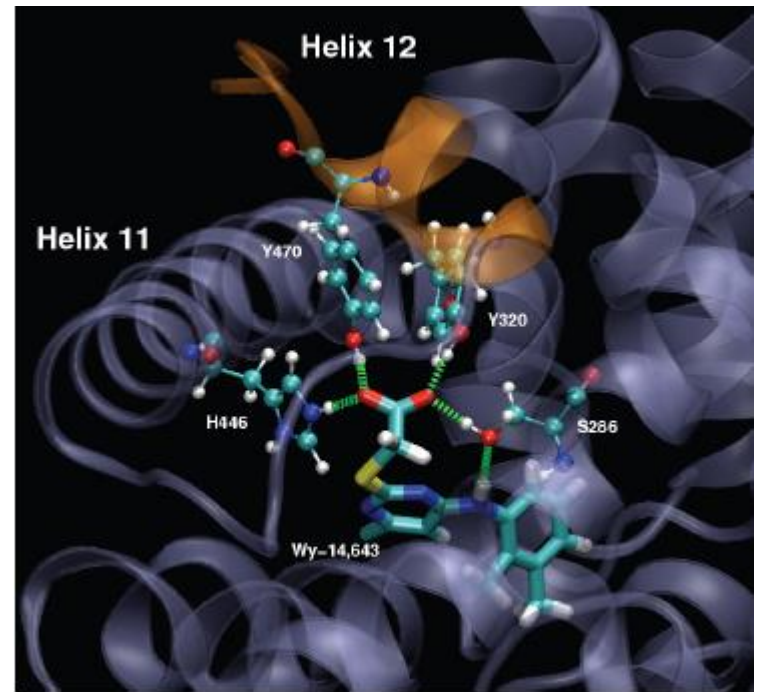
Forze di van der Waals

Debole attrazione intermolecolare causata da dipoli molecolari indotti.

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



Repulsive : Pauli exclusion principle
Attractive: induced dipole / induced dipole



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

Ad esempio, se una molecola di ossigeno si avvicina ad una molecola di acqua orientata verso di essa con l'atomo di ossigeno, la frazione di carica negativa presente su quest'ultimo respinge la nuvola elettronica della molecola di ossigeno e attira la carica positiva nucleare. Cio' determina uno spostamento del baricentro della carica negativa rispetto a quello della carica positiva nella molecola di ossigeno e quindi induce una polarita'.

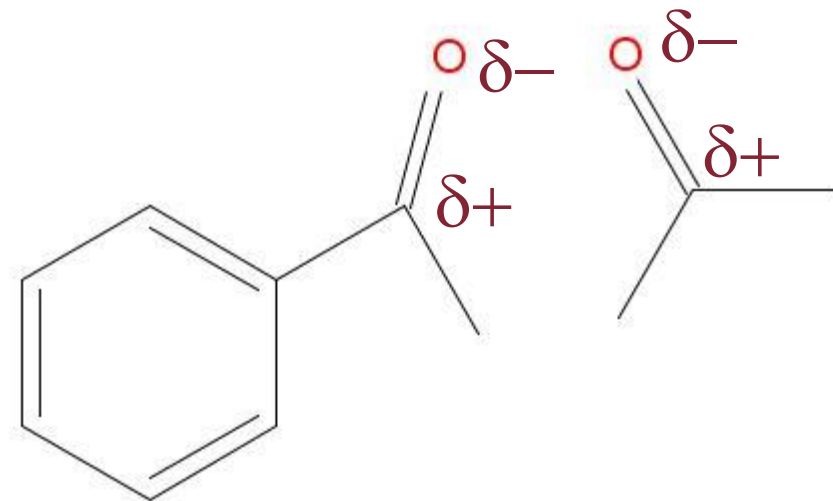
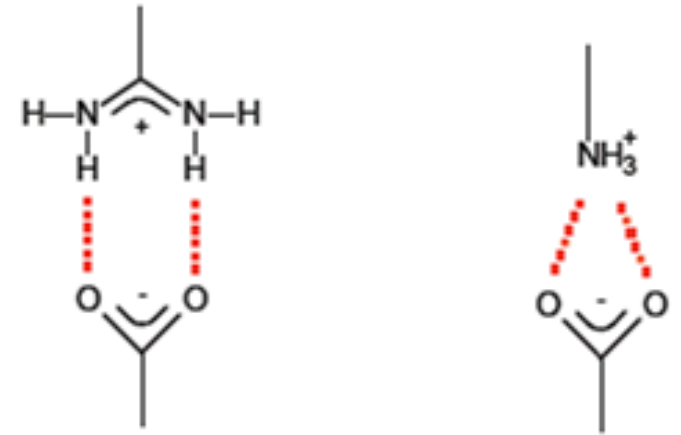
Interazioni Elettrostatiche

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:

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

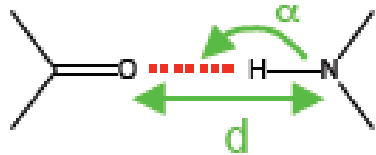
dove ϵ_r è un numero che prende il nome di costante dielettrica relativa (permittività elettrica relativa). Il prodotto $\epsilon_r \times \epsilon_0$ si indica con ϵ , e si chiama costante dielettrica del mezzo (permittività del mezzo).



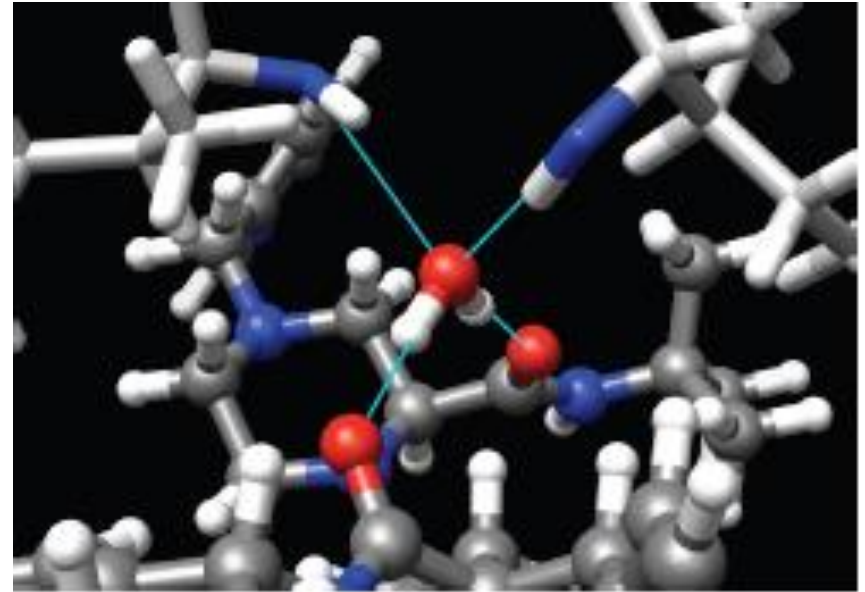
Interazioni Elettrostatiche

Il legame Idrogeno

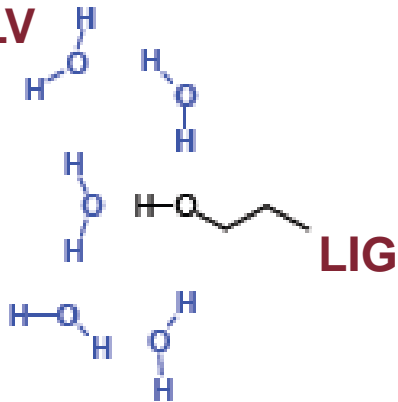
ElegH (solv) – ElegH (comp)
determina se i legami-H
contribuiscono o no all'affinità



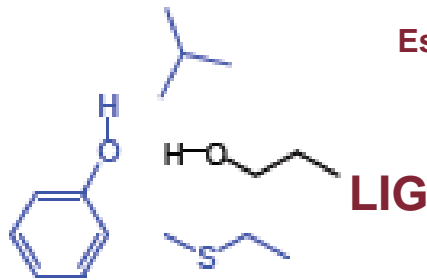
$$d = 2.5 - 3.2 \text{ \AA}$$
$$\alpha = 130 - 180^\circ$$



SOLV



solvent

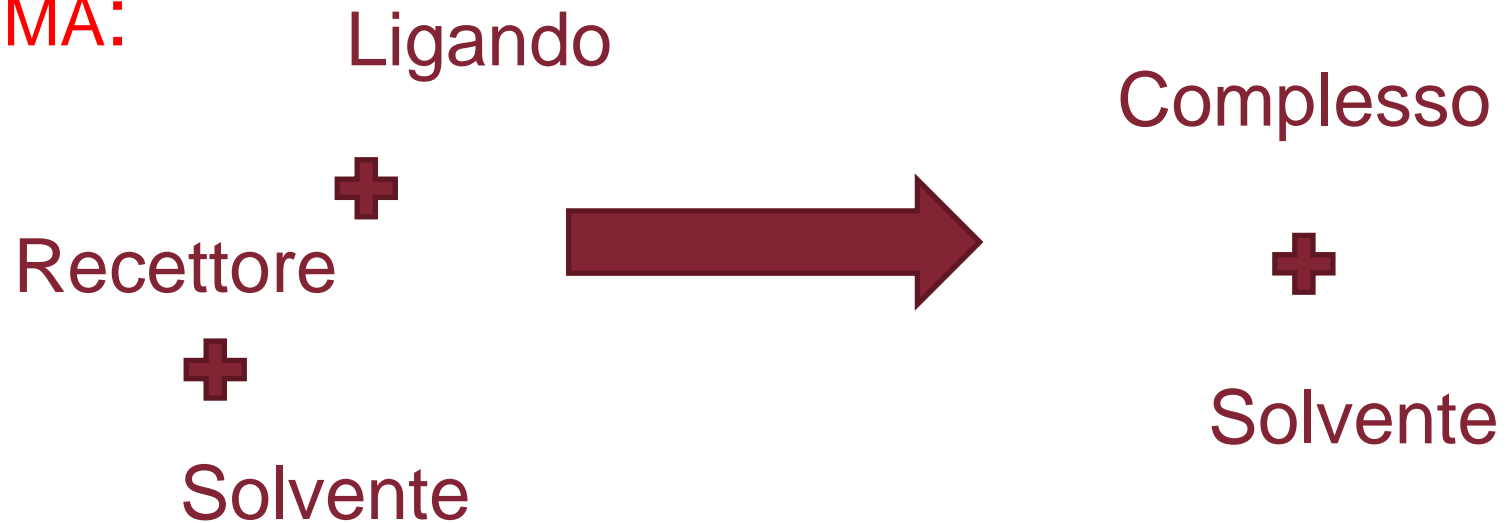


**REC
complex**

Es. Legame-H tra l'Indinavir, una molecola d'acqua e la Ile50 della proteasi dell'HIV-1 (1HSG in PDB)

Il fattore entropico ed entalpico nel binding

SISTEMA:



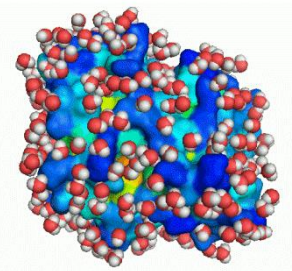
La formazione di un complesso è guidata dal cambiamento di **ENTALPIA** ed **ENTROPIA** del **sistema**

L'equazione fondamentale è:

$$\Delta G = \Delta H - T\Delta S$$

dove ΔG = variazione energia libera, ΔH = variazione entalpia, ΔS = variazione entropia, T = temperatura (K)

Effetti di solvatazione (Termine Entropico + Entalpico)



- La ricognizione molecolare tra due molecole avviene in ambiente acquoso
- Le molecole di acqua mediano l'interazione attraverso ponti idrogeno
- Rottura e formazione di legami-H
- La desolvatazione del ligando e della proteina attraverso la complessazione è:

Sfavorevole (Zone Elettrostatiche)

- Carica del solvente
- Screening delle interazioni elettrostatiche sulla superficie della macromolecola

Favorevole (zone apolari)

- Cavità nel sito attivo
- Riorganizzazione delle molecole d'acqua
- Interazioni di tipo van der Waals tra solvente e ligando
- Gli effetti di desolvatazione sono proporzionali all'area superficiale accessibile

Il fattore entropico nel binding

$$\Delta G_{\text{binding,solution}} = \Delta G_{\text{binding,vacuo}} + \Delta G_{\text{solvation}}(\text{EI}) - \Delta G_{\text{solvation}}(\text{E+I})$$

La stabilità di un complesso può essere valutata determinando la costante di equilibrio, che è correlata alla variazione di energia libera:

$$\Delta G = \Delta G^0 - RT \ln K_d$$

dove ΔG^0 =variazione dell'energia libera in condizioni standard, R = costante dei gas, T =temperatura assoluta, K_d = costante di legame espressa come costante di dissociazione con

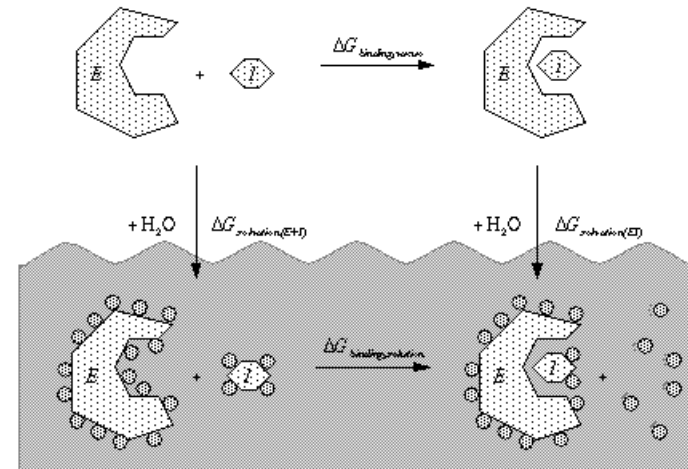


reazione di formazione del complesso è:

$$K_d = [\text{R}] [\text{L}] / [\text{RL}]$$

K_d è definita come la concentrazione del ligando per la quale il 50% dei siti del recettore sono occupati.

L'effetto idrofobico è il termine maggiormente stabilizzante per i complessi biomolecolari, mentre le interazioni coulombiane e i legami idrogeno forniscono specificità alle interazioni proteina-ligando.



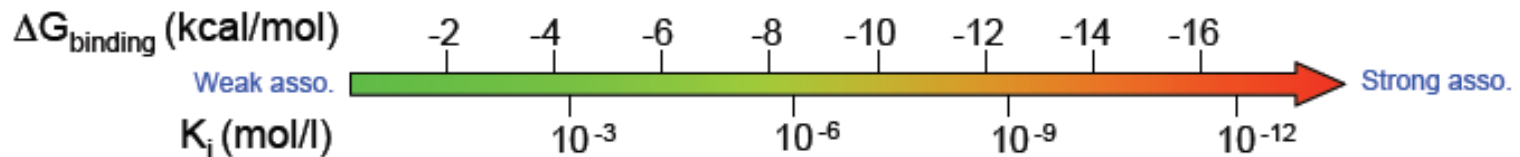
Il fattore entropico nel binding

Binding, dissociation and inhibition constants.
Binding free energy

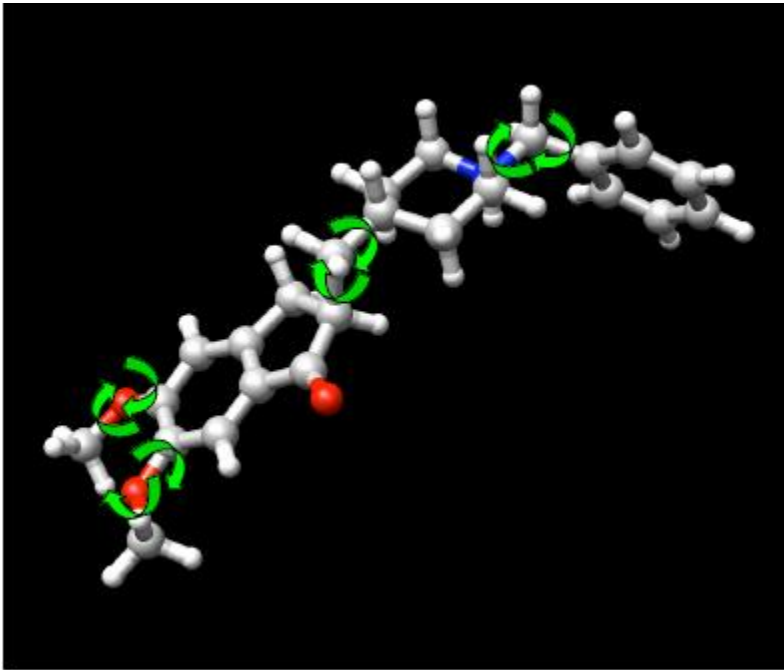


K_b : binding constant, K_d : dissociation constant, K_i : inhibition constant

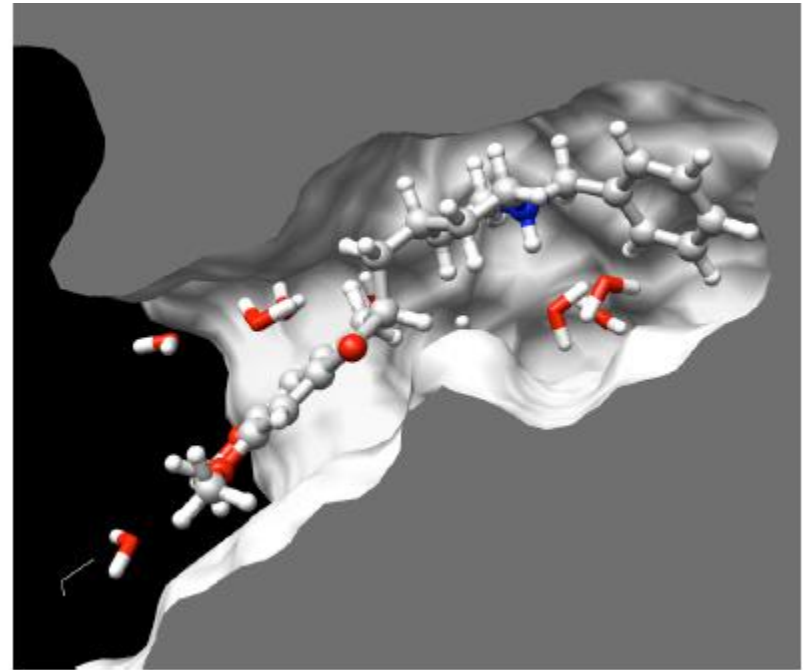
$$\Delta G_{\text{binding}} = -RT \ln K_b = RT \ln K_i = \Delta H - T\Delta S$$



Il fattore entropico nel binding



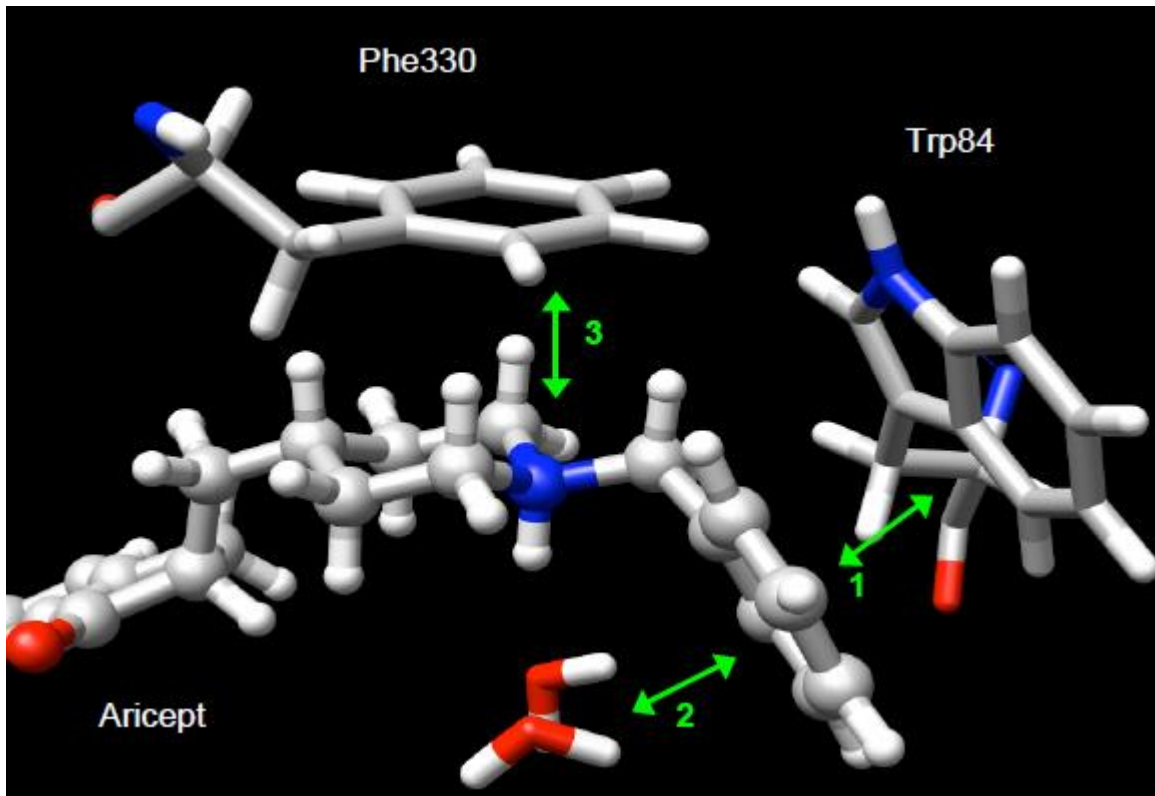
Es. Aricept in soluzione



Es. Aricept nel sito attivo
della acetilcolinesterasi

I gradi di libertà “congelati” durante la complessazione rendono il ligando sfavorito al *binding*

Interazioni π

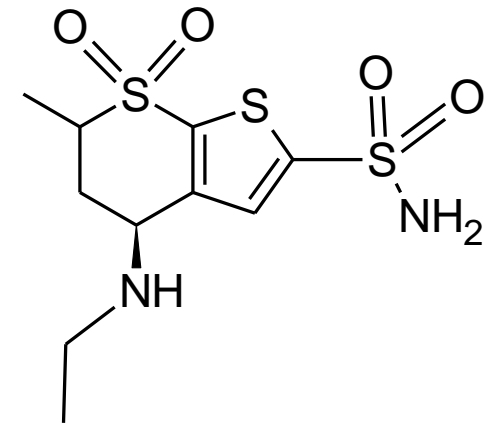
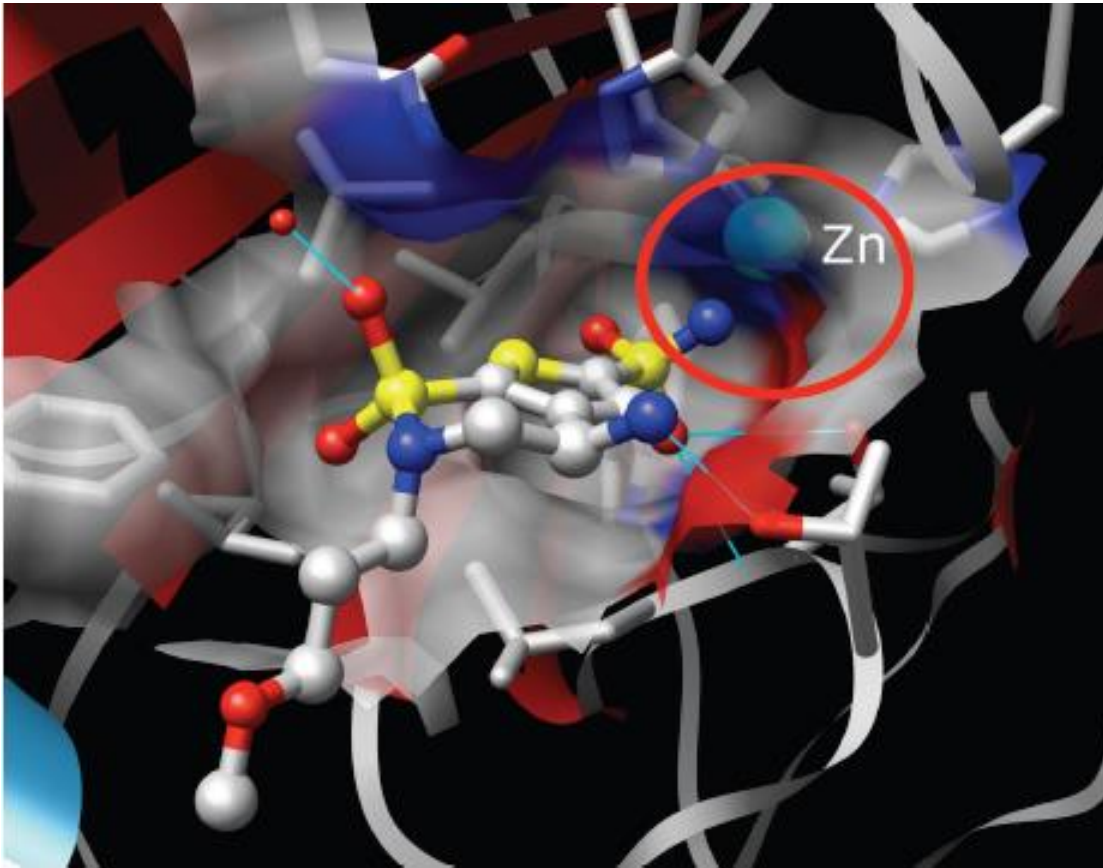


1. Interazioni π - π
2. Interazioni OH- π
3. Interazioni cationi- π

Es. Interazioni π tra l'Aricept (farmaco anti-Alzheimer approvato dalla FDA) e l'acetilcolinesterasi

Coordinazione con Metalli

Legame covalente di coordinazione



Dorzolamide: Inibitore dell'anidasi carbonica, approvato dalla FDA come agente anti-glaucoma

DOCKING: Le *Scoring Functions*

Structure-based

With 3D structure for
the targeted
macromolecule

- Physical based scoring functions :
 - Free energy simulations (FEP, TI)
 - MM-PBSA, MM-GBSA
 - Linear interaction energy (LIE)
- Empirical-based scoring function
(regression-based approaches)
Ex: Ludi score
- Knowledge-based approaches
(Potentials of Mean Force, PMF)
Ex: PMF score

Utility?

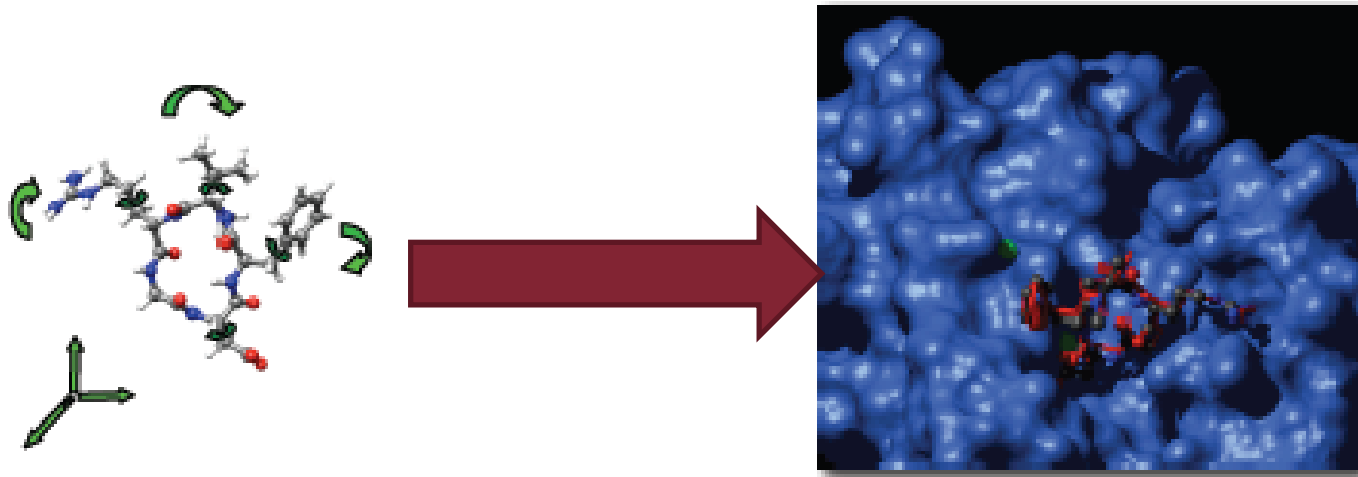
- *Virtual Screening*
- *De novo design*
- *Lead optimization*

Ligand Docking

Posa o **Binding Mode**: posizione, orientamento e conformazione di un ligando sulla superficie di macromolecole biologiche

Binding Mode sperimentale: Estratta da dati cristallografici e stimata come la migliore posa in termini di *free binding energy*.

Docking: Metodo computazionale utilizzato per la predizione di un binding mode il più vicino possibile allo sperimentale (<2 Å)



Ligand Docking

L'approccio standard:

- Campionamento dello spazio conformazionale del ligando nel binding site della macromolecola.
- Scoring di tutte le possibili conformazioni del ligando.
- Selezione delle pose con il miglior valore di energia (score) rispetto alla posa iniziale.

Livelli differenti di approssimazione:

- La proteina e il ligando sono rigidi.
- La proteina è rigida e il ligando è flessibile.
- La proteina e il ligando sono flessibili.

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

Tre tipi di ricerca conformazionale:

1. Ricerca sistematica (esplorazione combinatoriale dei gradi di libertà, costruzione incrementale)
 1. Ricerca stocastica (Monte Carlo, Algoritmo genetico)
 2. Ricerca deterministica (Minimizzazione, Dinamica Molecolare)

Details...

Campionamento dello spazio configurazionale

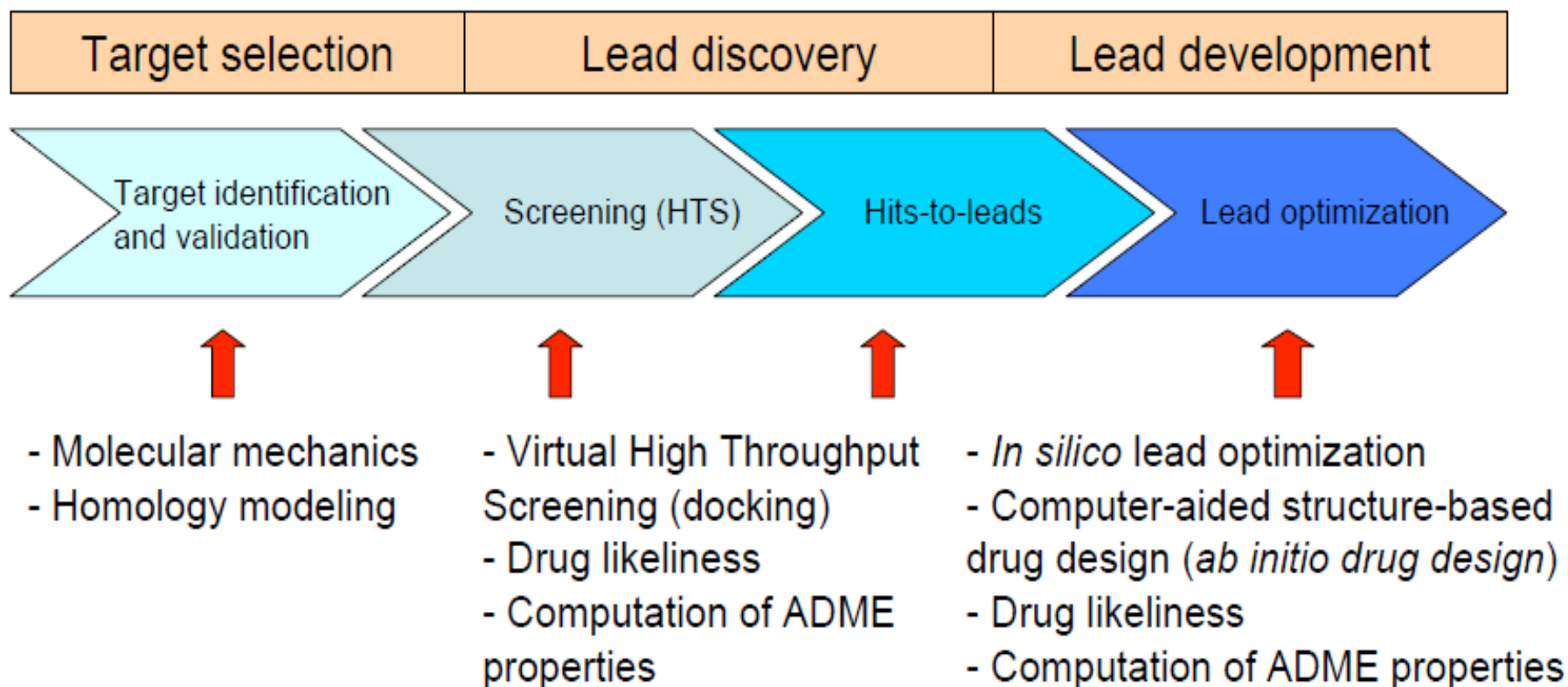
Diversi algoritmi sviluppati per trovare la miglior conformazione del ligando nel sito di legame.

Scoring dei complessi proteina-ligando

Scopo dello scoring: identificare il modo di binding corretto attraverso la sua minore energia, se **la configurazione ad energia minore è assunta essere quella “corretta”**.

Sono state sviluppate diverse funzioni per misurare l'affinità di binding ligando-recettore negli algoritmi di docking → **Funzioni di scoring = funzioni designate a classificare complessi proteina-ligando differenti, a seconda delle loro affinità di binding.**

Drug Design



...and applications

662

J. Chem. Inf. Model. **2010**, *50*, 662–676

Combining 3-D Quantitative Structure–Activity Relationship with Ligand Based and Structure Based Alignment Procedures for *in Silico* Screening of New Hepatitis C Virus NS5B Polymerase Inhibitors

Ira Musmuca,[†] Antonia Caroli,[†] Antonello Mai,[†] Neerja Kaushik-Basu,[‡] Payal Arora,[‡] and Rino Ragno^{*†}

Istituto Pasteur-Fondazione Cenci Bolognetti, Dipartimento di Chimica e Tecnologie del Farmaco, Sapienza Università di Roma, P. le A. Moro 5, 00185, Rome, Italy and Department of Biochemistry and Molecular Biology, UMDNJ-New Jersey Medical School, 185 South Orange Avenue, Newark, NJ

Received December 4, 2009

Workflow del protocollo di allineamento: L'allineamento *Structure-Based*

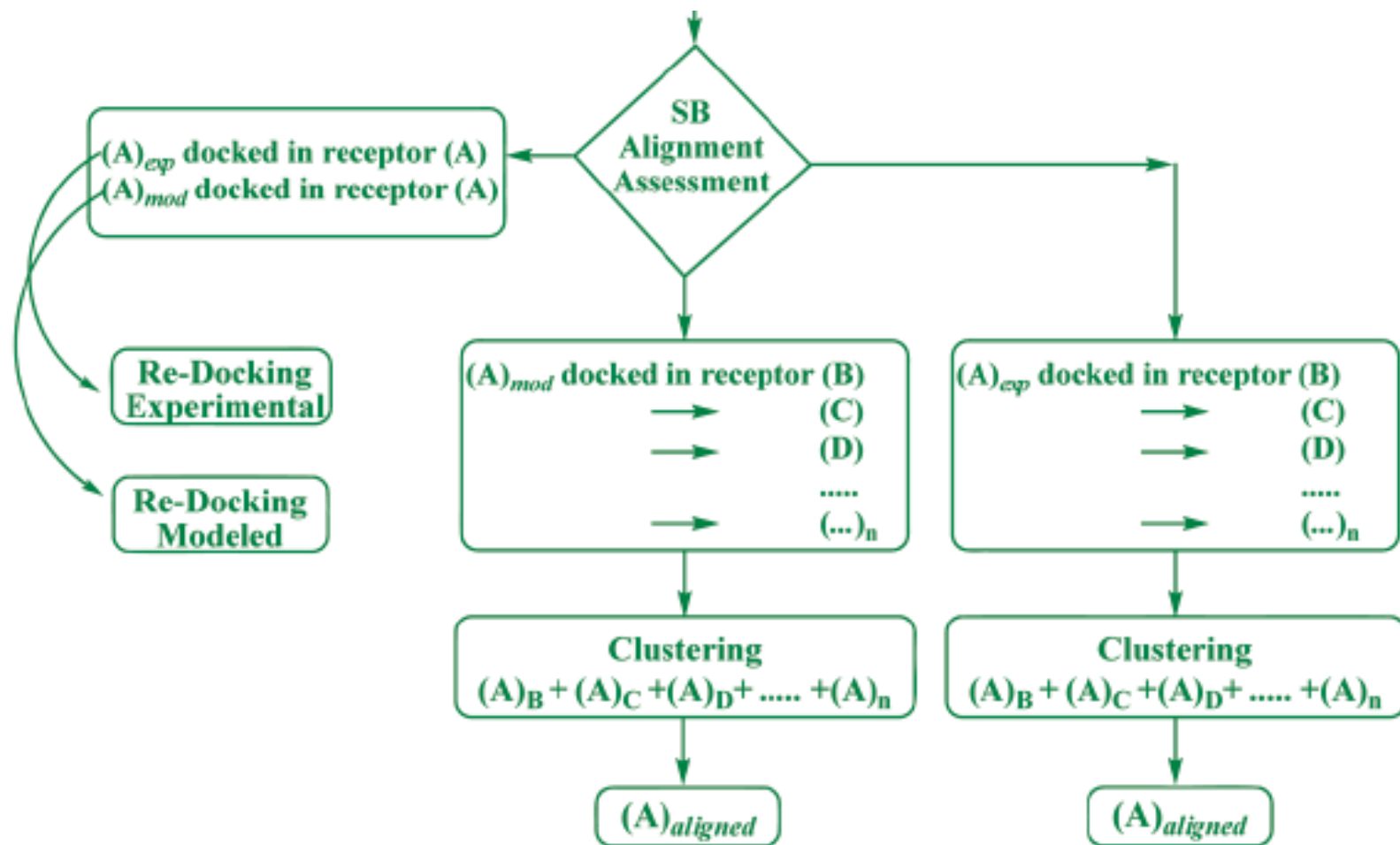


Table 5. Assessment of the Autodock Program in the Redocking Stage^a

binding site	PDB	ligand entry	best docked (rmsd)	best cluster		best fitted cluster		
				cluster N°	rmsd	cluster N°	rmsd	
thumb	1NHU	1	3.17	3	2.17	3	0.89	
	1NHV	2	4.13	2	4.75	26	1.86	
	1OSS	13	3.46	1	3.46	9	1.50	
	1YVX	3	3.81	1	3.81	6	1.58	
	1YVZ	4	3.74	4	1.92	4	0.78	
	2D3U	6	0.71	1	0.71	1	0.44	
	2D3Z	7	0.75	1	0.75	1	0.60	
	2D4I	8	1.43	1	1.43	1	0.58	
	2GIR	5	5.70	2	1.12	2	0.70	
	2HAI	14	2.05	1	2.05	1	0.92	
	2HWH	9	9.86	2	2.13	2	0.79	
	2HWI	10	0.34	1	0.34	1	0.24	
	2IIR	11	5.84	2	1.67	2	0.73	
	2JC0	15	0.85	1	0.85	1	0.68	
	2OSD	12	5.74	3	2.78	7	1.28	
		average rmsd		3.44		2.00		0.90
palm	1YVF	17	3.38	4	1.12	4	0.93	
	1Z4U	16	0.89	1	0.89	1	0.57	
	2AWZ	18	3.53	3	1.72	3	1.11	
	2AX0	19	0.84	1	0.84	1	0.52	
	2AX1	20	3.24	2	0.99	2	0.61	
	2FVC	21	1.04	1	1.04	1	0.74	
	2GC8	24	2.03	1	2.03	2	1.82	
	2GIQ	22	3.49	5	1.91	5	1.81	
	2JC0	15	0.65	1	0.65	1	0.42	
	2JC1	23	0.74	1	0.74	1	0.47	
		average rmsd		1.98		1.19		0.9

^a rmsd values for the first ranked pose (best docked), the lowest energy docked conformation of the most populated cluster (best cluster), and the one closest to the experimentally bound conformation (best fitted cluster).

BD= La conformazione con il minor valore di *binding energy* generata da Autodock

BC= La conformazione a più bassa energia del cluster più popolato

BF= La conformazione con il minor valore di rmsd

STRUCTURE-BASED DRUG DESIGN (SBDD).

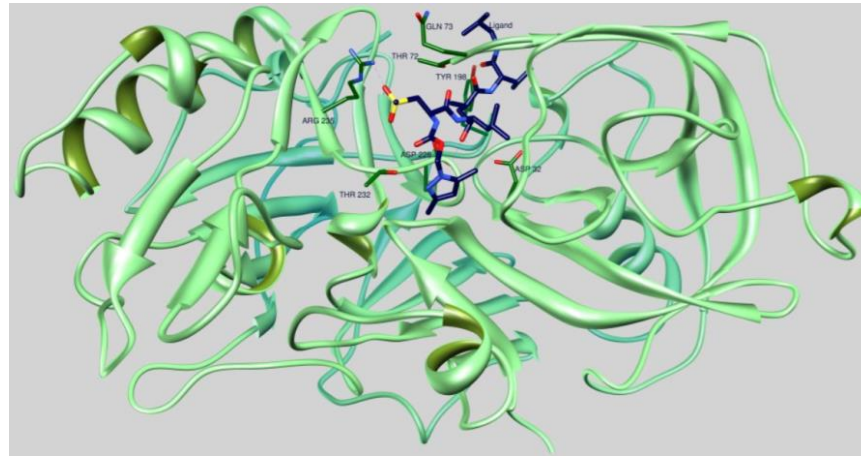
**UN' EFFICIENTE OTTIMIZZAZIONE MEDIANTE CROSS-
DOCKING**

MOLECULAR DOCKING

DOCKING MOLECOLARE



- SAMPLING
- SCORING FUNCTION



Identificazione di composti di
interesse farmaceutico

Comprensione meccanismi
interazione

Ottimizzazione del *lead*

Bioremediation

MOLECULAR DOCKING

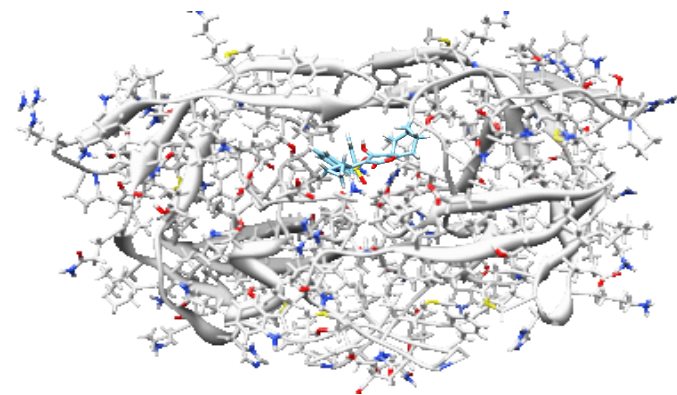
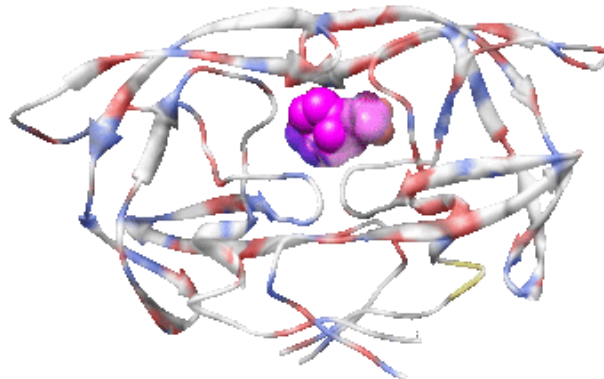
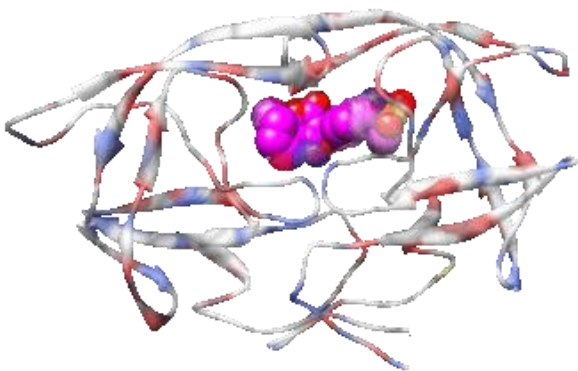
Docking



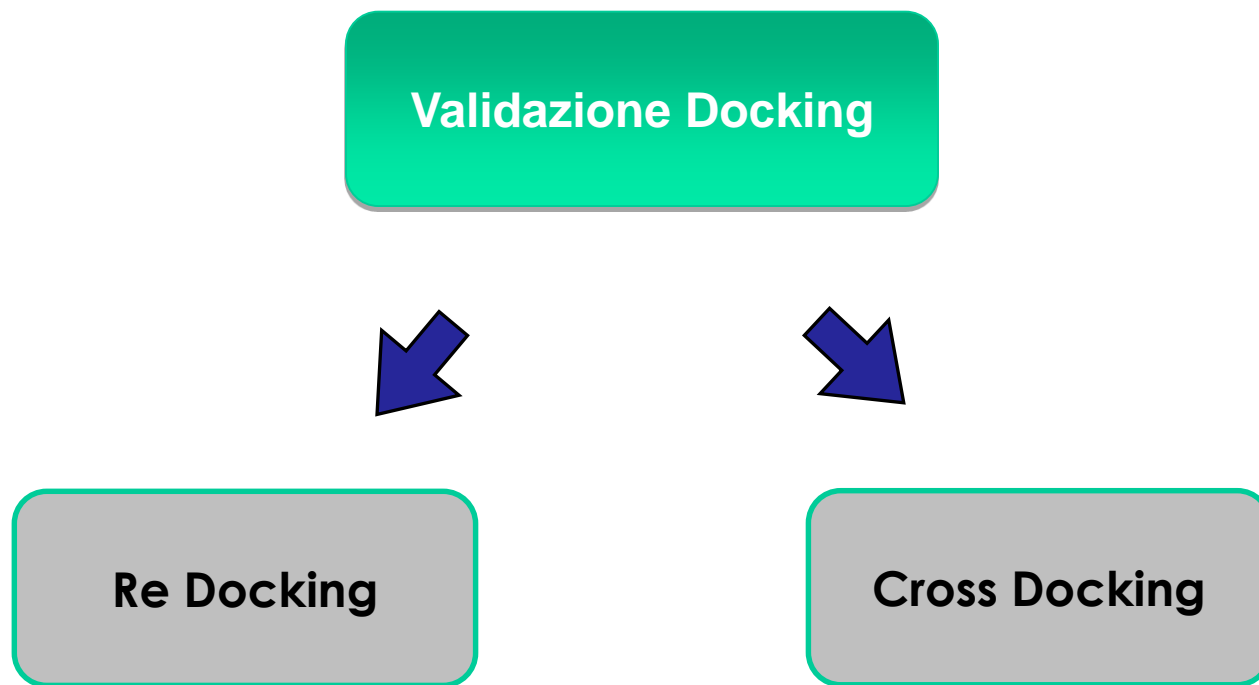
Rigido

Semi-Flessibile

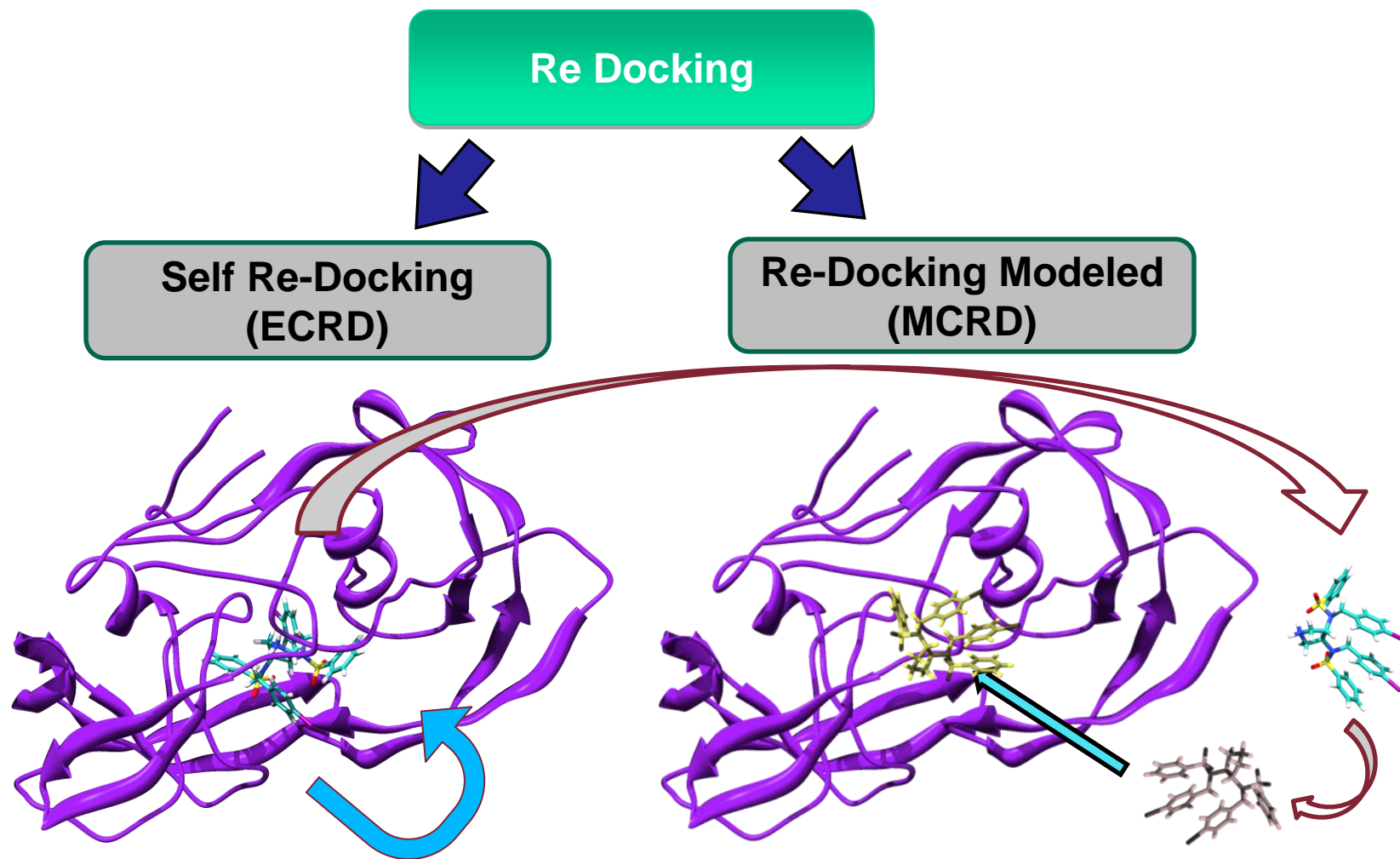
Flessibile



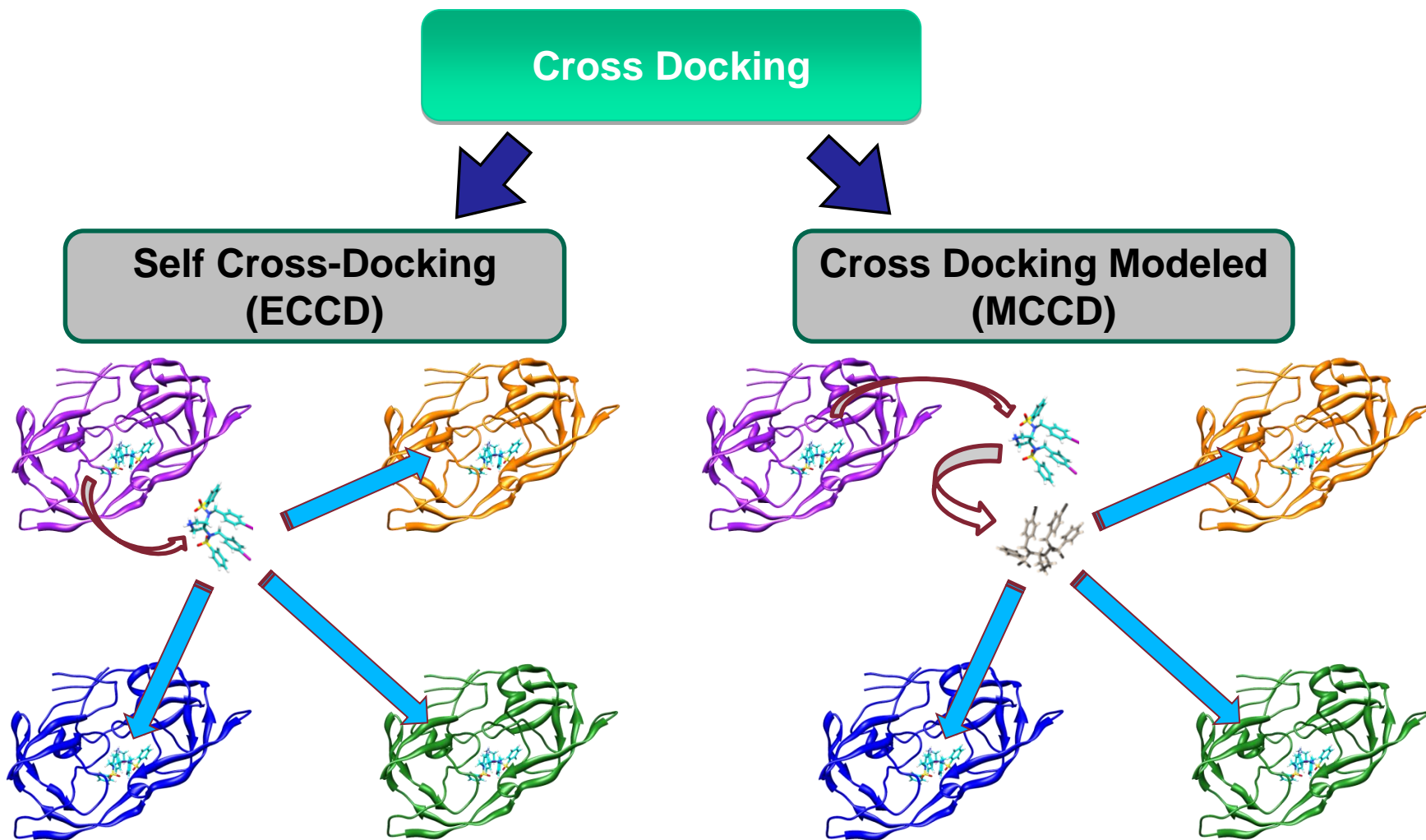
VALIDAZIONE Docking



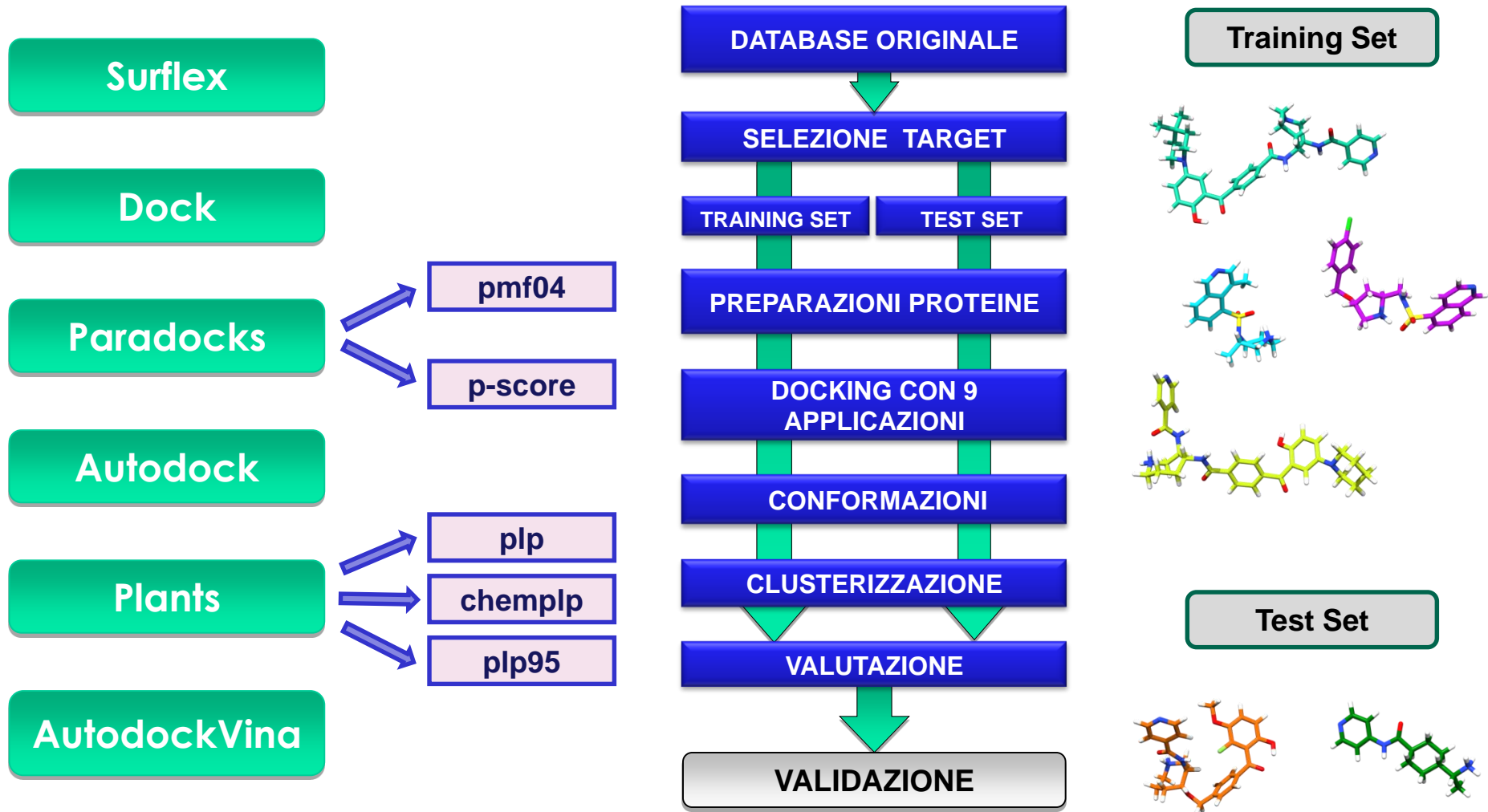
RE-docking (RD)



Cross-docking (CD)



panoramica



Procedura

Metodo unificato

Automatizzazione

Scelta Programma

- **Parametri base**
- **Spazio di ricerca 5Å**
- **Stesse molecole di partenza**
- **Metodica unificata per la valutazione dei risultati ottenuti**

Pulizia delle Proteine

Eliminazione solvente

Eliminazione metalli strutturali

Eliminazione catene inutili

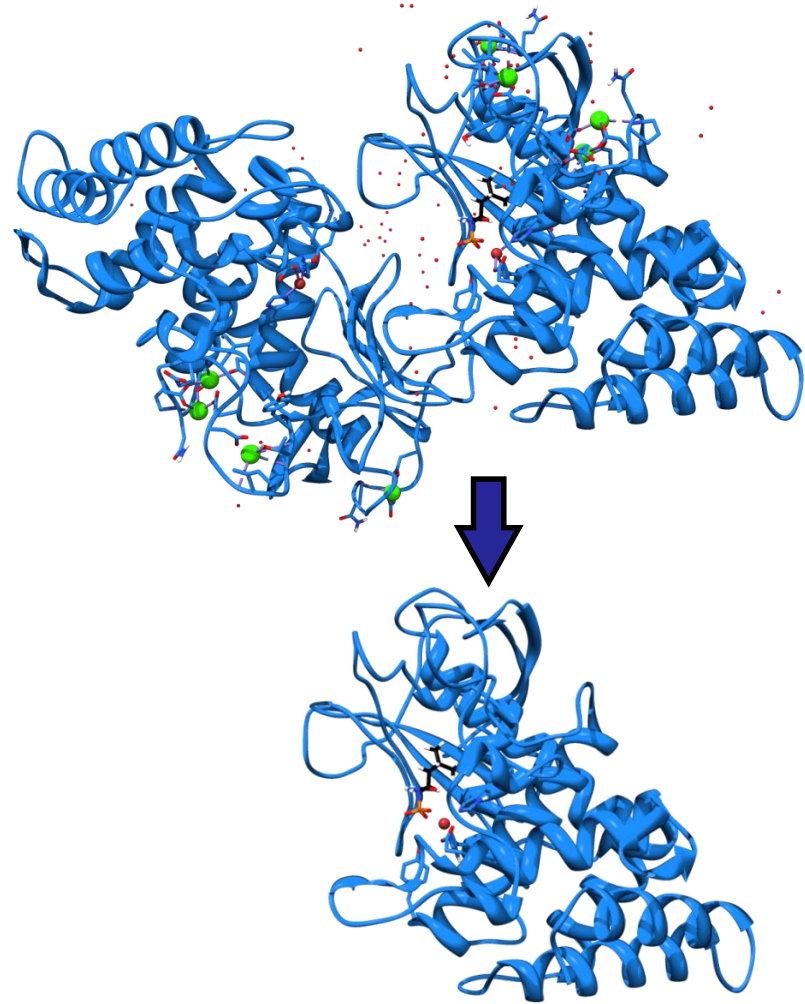
Ricostruzione aminoacidi tagliati

Protonazione

Cariche aminoacidi

Allineamento Complessi

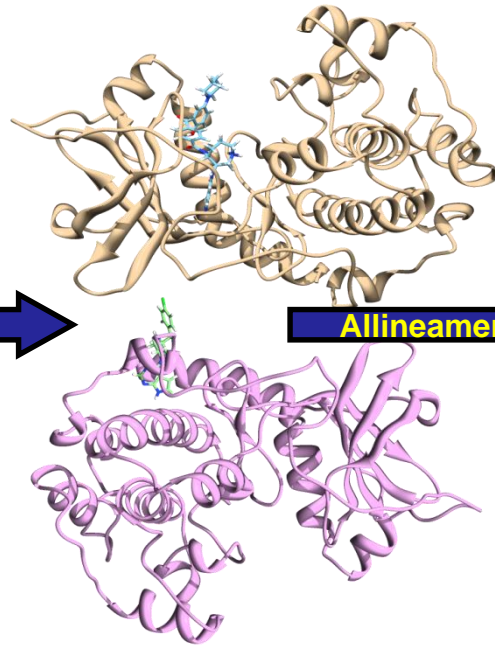
Divisione in Cluster



ALLINEAMENTO/CLUSTER

cAMP-dependent protein Kinase catalytic subunit alpha

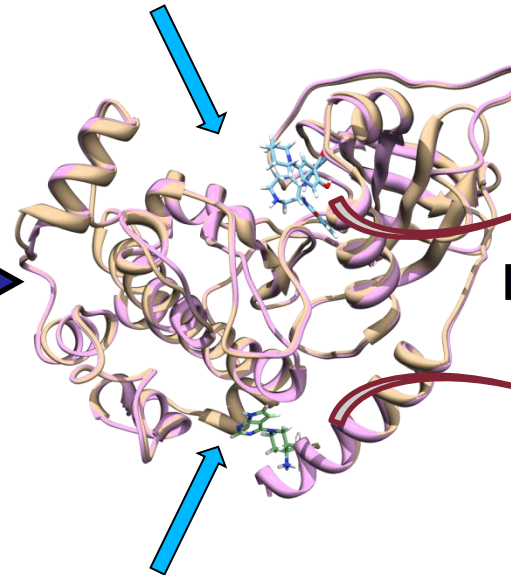
Riunificazione
lock/key nel
complesso



Proteine Pulite

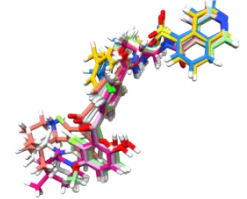
Allineamento

Binding site 1

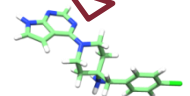


Binding site 2

Cluster 1

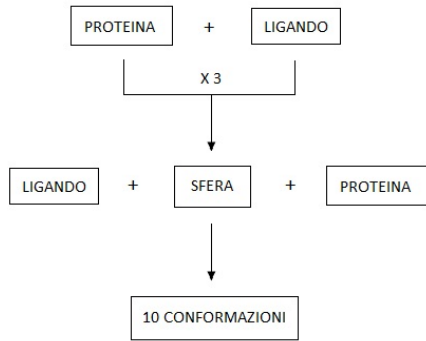


Divisione
complessi in
lock e key

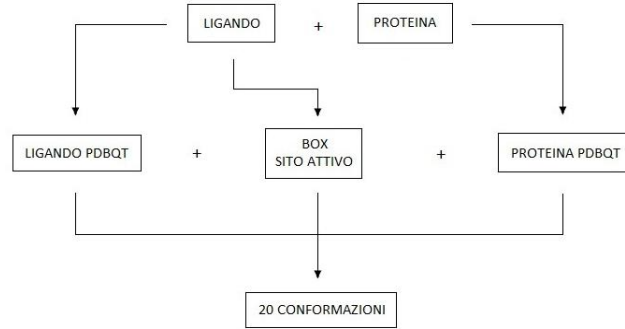


Cluster 2

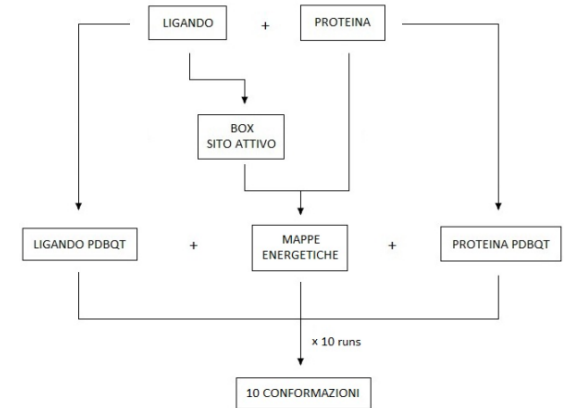
PROGRAMMI USATI



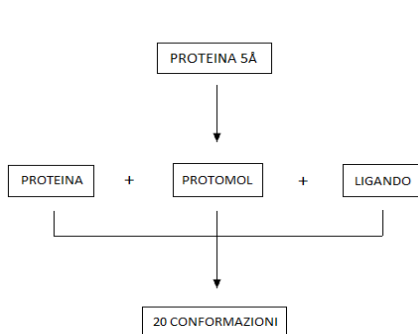
Plants



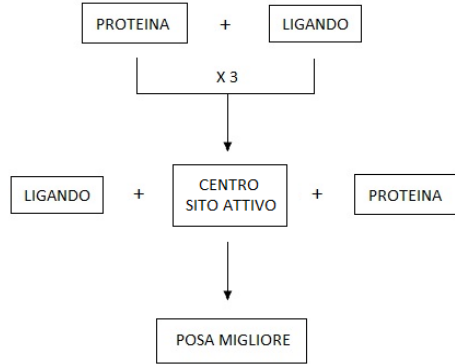
Vina



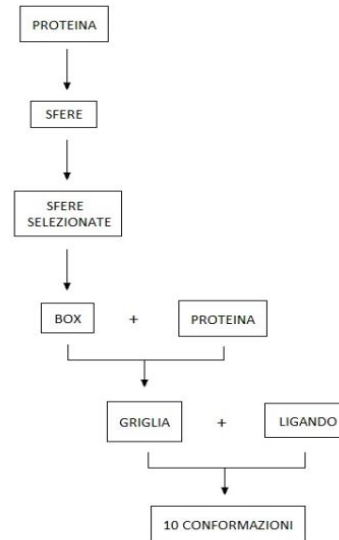
Autodock



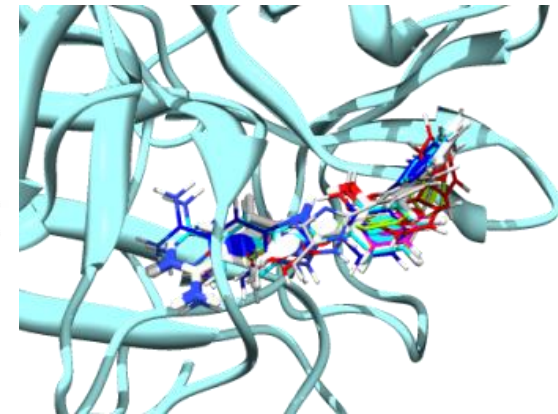
Surflex



Paradocks

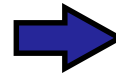


Dock



Clusterizzazione

Predizione Binding mode

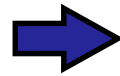


Clusterizzazione

- Rapido
- Automatizzabile

➤ BD (Best Docked)

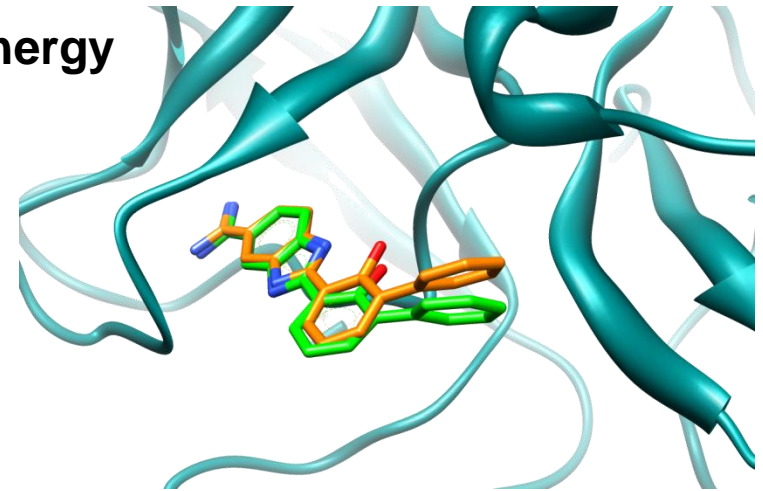
➤ BC (Best Cluster)



Binding Energy

RMSD

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



■ originale ■ docking

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%		Δ 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

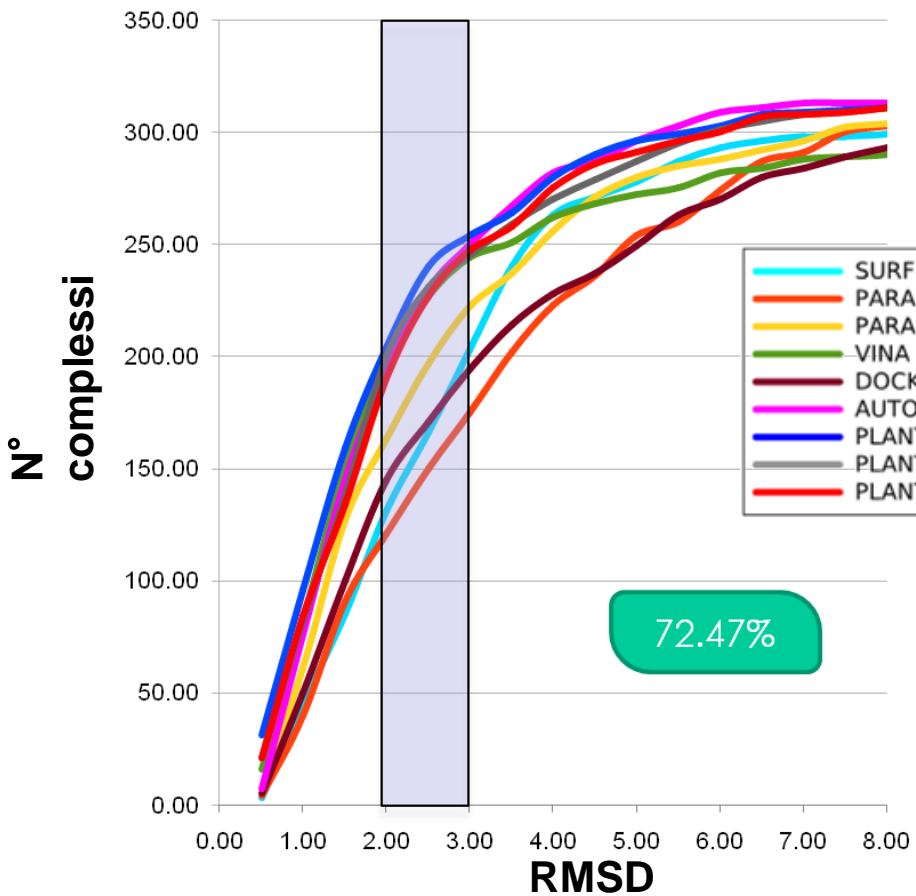
$$DA \text{ (Docking Accuracy)} = f_{RMSD \leq 2} + 0.5 (f_{RMSD \leq 3} - f_{RMSD \leq 2})$$

	ECRD	vs	BD-ECCD
Software			$\Delta DA\%$
Surflex			+20.09
Paradocks pmf04			+19.62
Paradocks p-score			+25.63
Vina			+22.94
Dock			+11.71
Autodock			+21.04
Plants chemplp			+14.56
Plants plp			+16.77
Plants plp95			+17.09
Media			+18.83

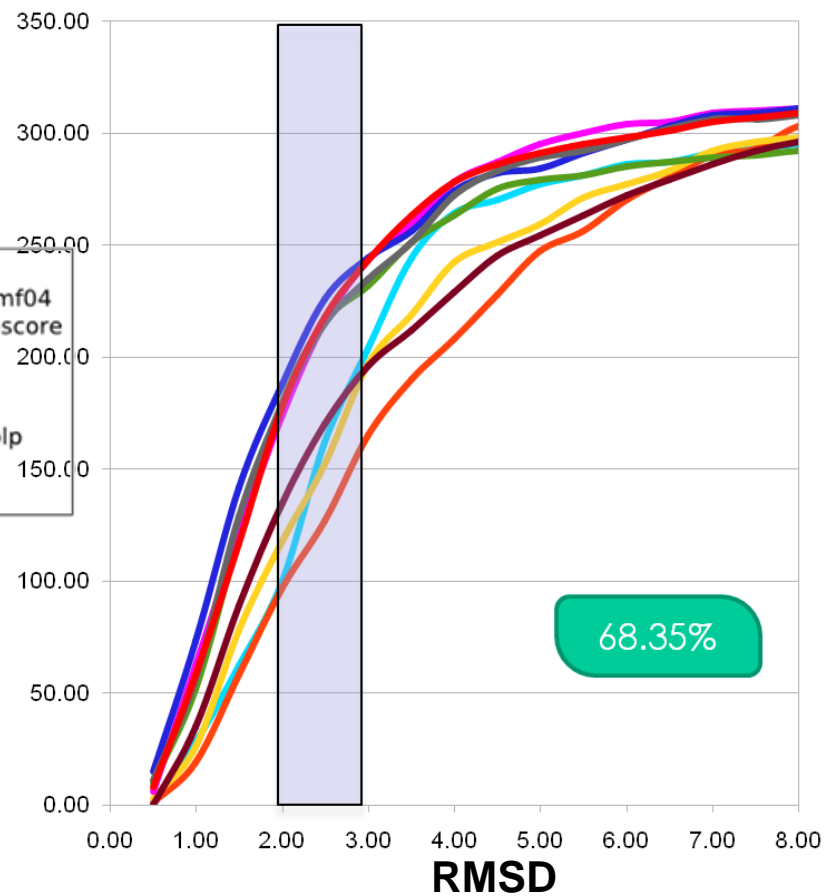
	MCRD	vs	BD-MCCD
Software			$\Delta DA\%$
Surflex			+29.91
Paradocks pmf04			+26.27
Paradocks p-score			+27.06
Vina			+30.11
Dock			+16.30
Autodock			+24.87
Plants chemplp			+21.04
Plants plp			+20.25
Plants plp95			+23.73
Media			+24.36

Prestazioni singoli programmi

Cross-Docking



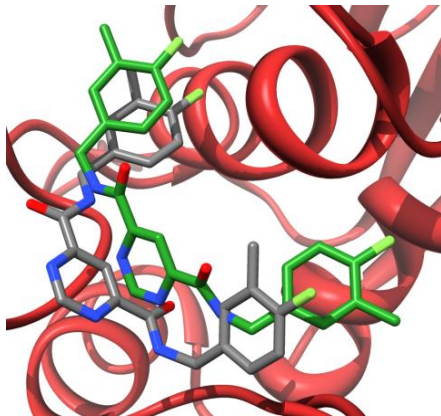
Cross -Docking Modeled



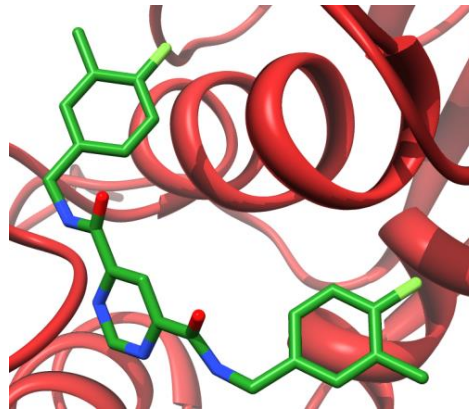
Prima scelta

RMSD medio

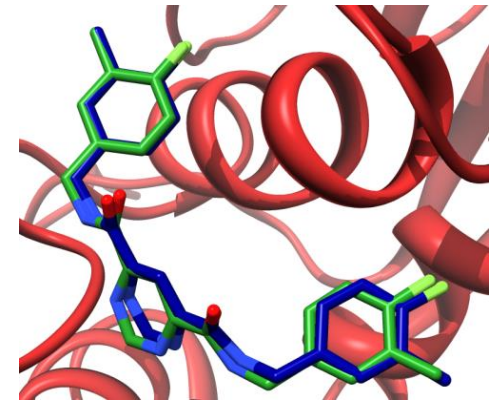
Deviazione
Standard RMSD



Autodock



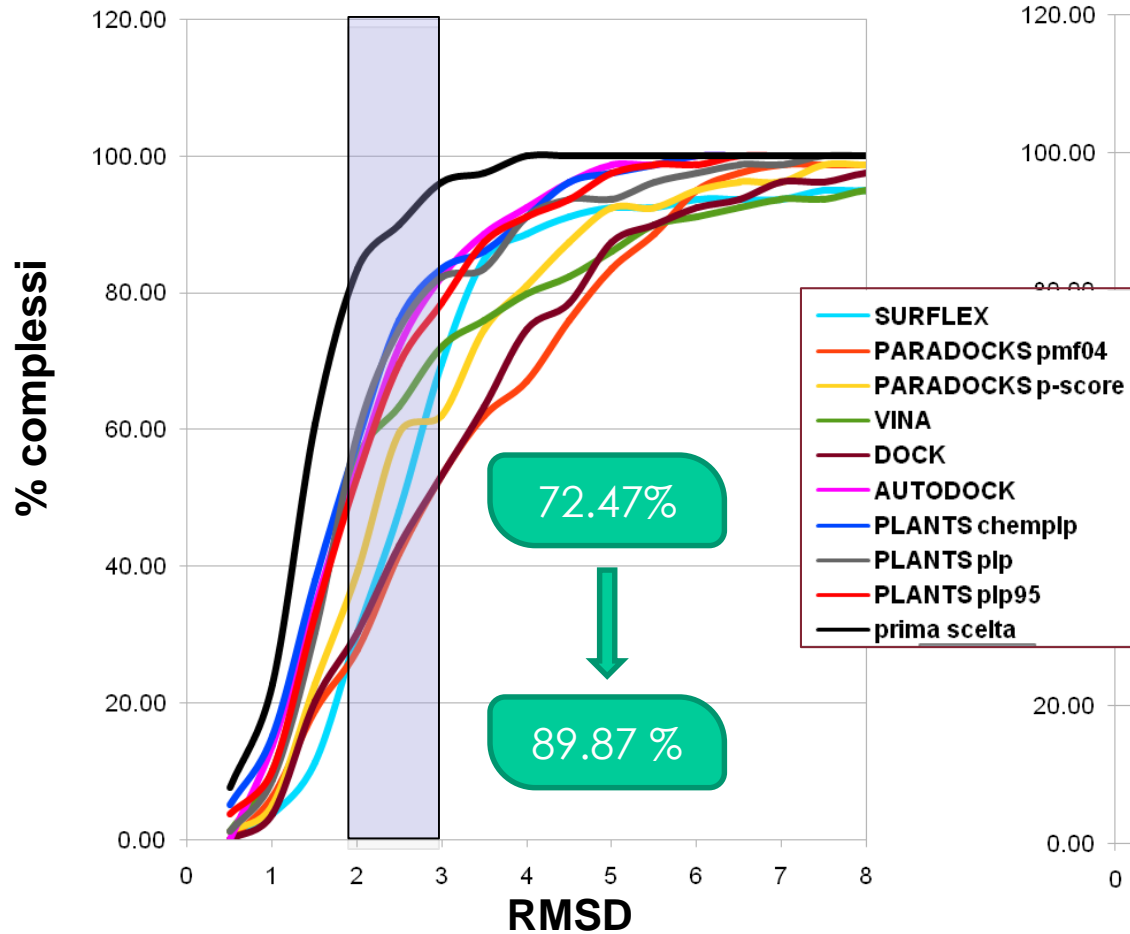
Sperimentale



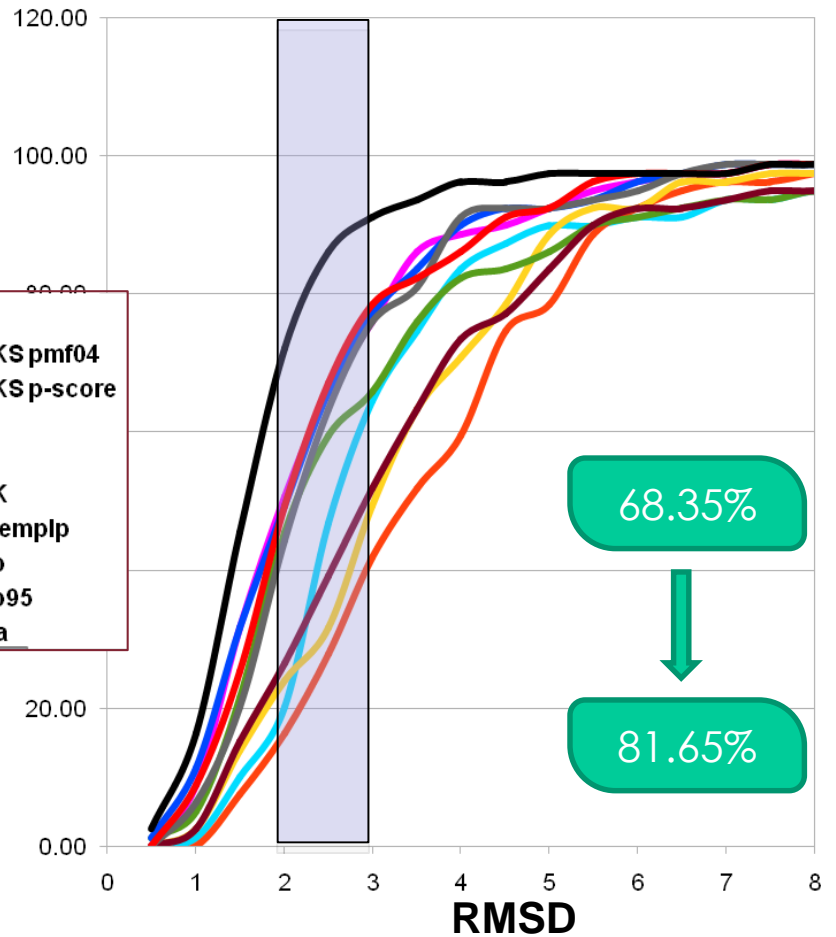
PLANTS chemplp

Prestazioni Prima scelta

Cross-Docking

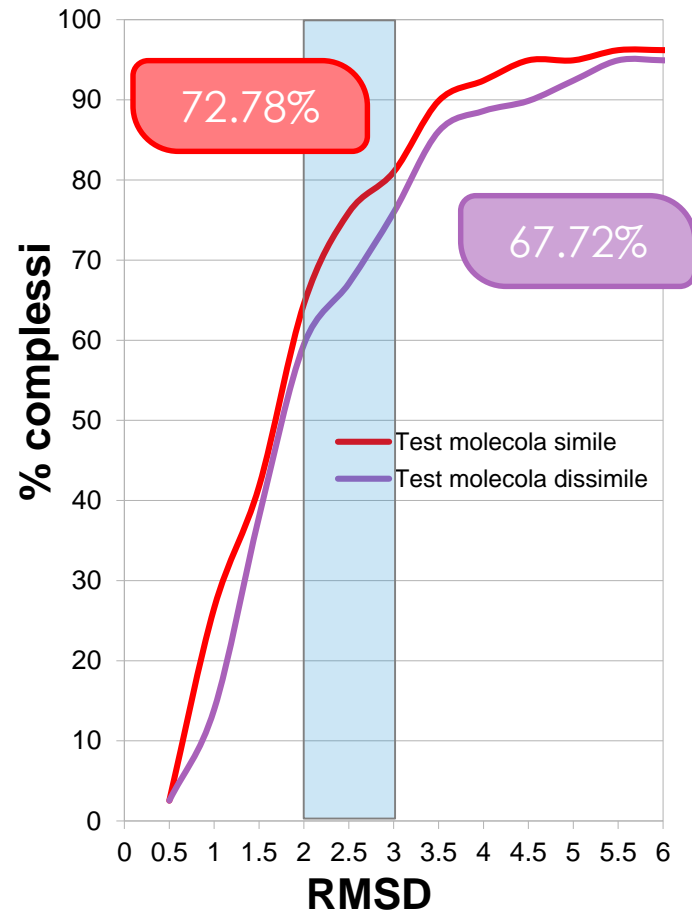
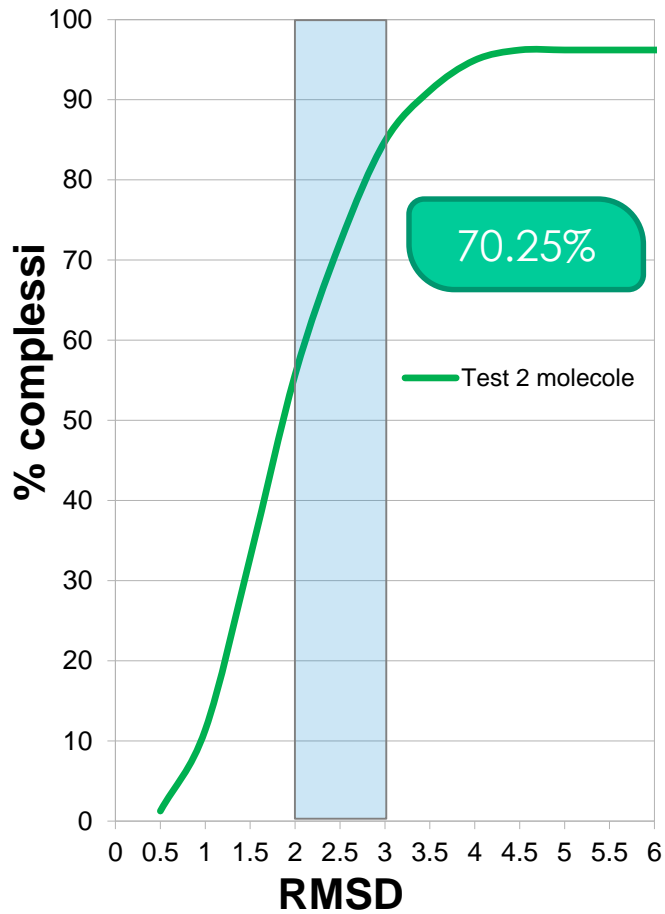


Cross -Docking Modeled



Validazione esterna

Validazione Cross Docking Modeled



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