

# Predizione di tasche allosteriche quali potenziali siti di legame per molecole bioattive



**SAPIENZA**  
UNIVERSITÀ DI ROMA

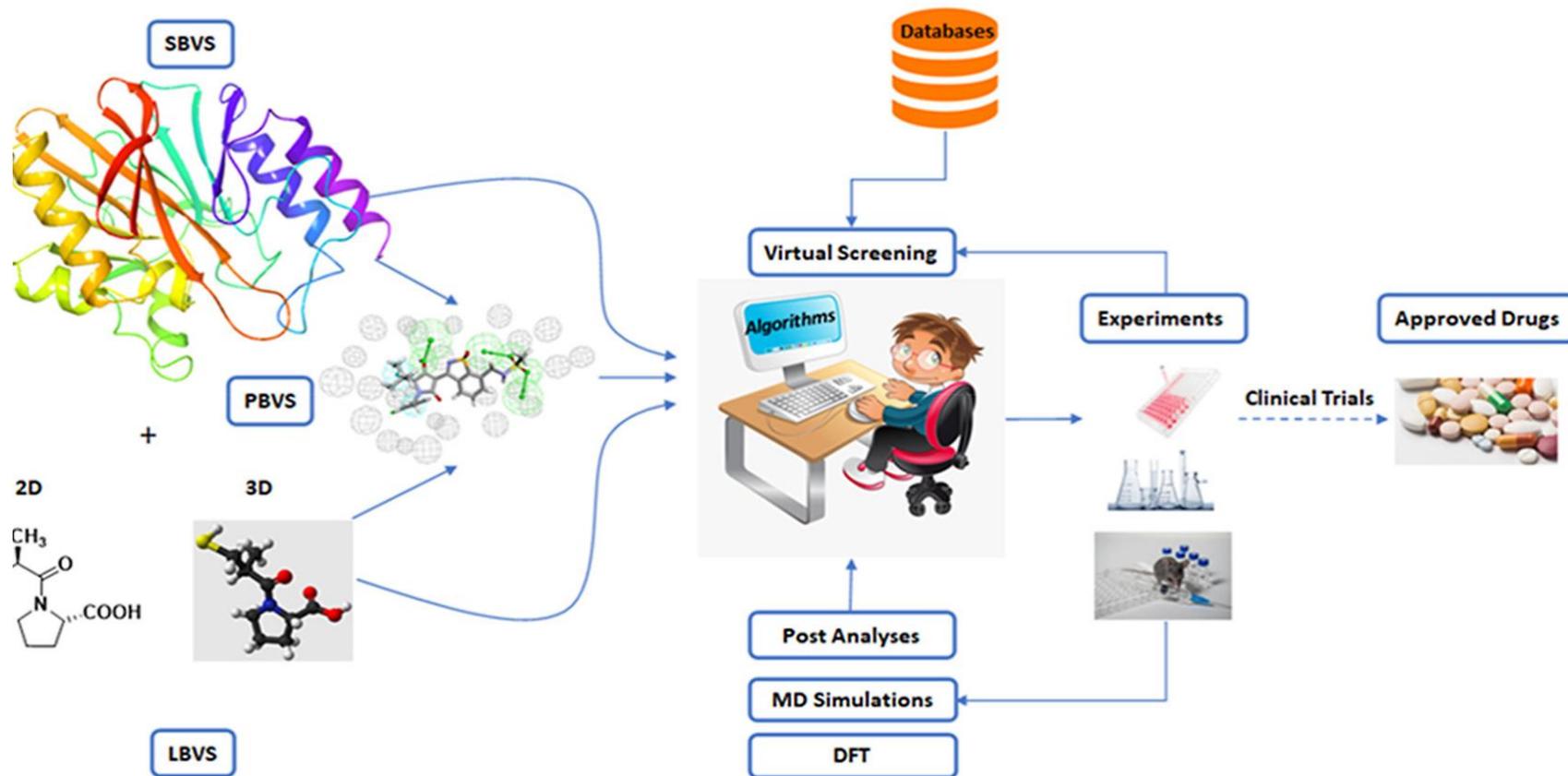
**Facoltà di Farmacia e Medicina**  
**Corso di Laurea in Chimica e Tecnologia Farmaceutiche**  
**Tesi Sperimentale in Chimica Farmaceutica**  
**a.a. 2021/2022**

**Laureando: Giulia Marigliani**  
**Matricola: 1742149**

**Relatore: prof. Rino Ragno**



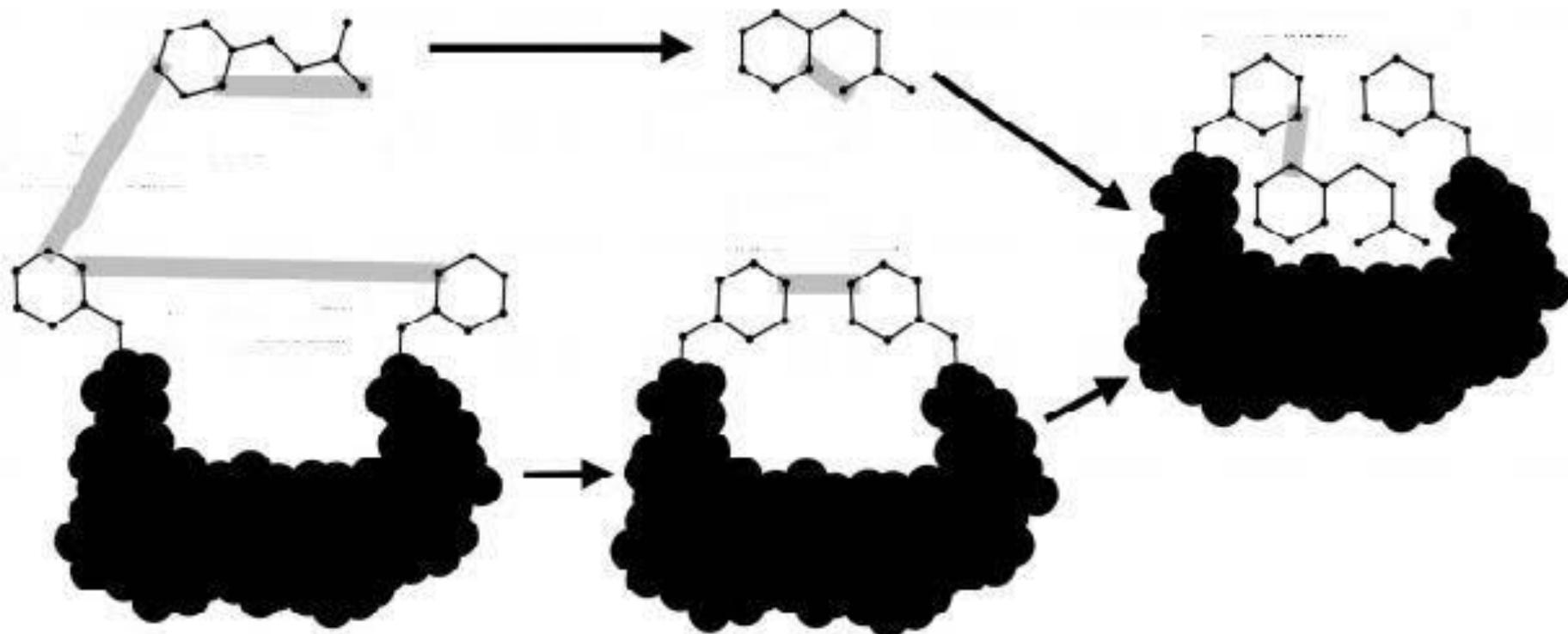
# Computer Aided Drug Design



Victor T. Sabe et al, "Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: A review", *European Journal of Medicinal Chemistry*, Volume 224,2021



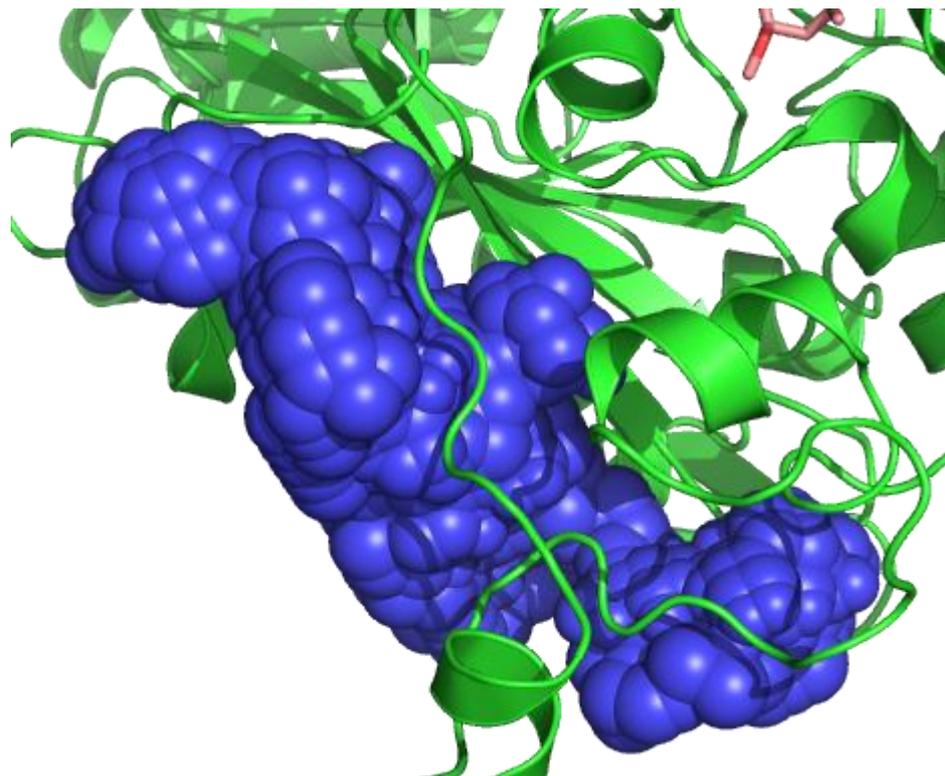
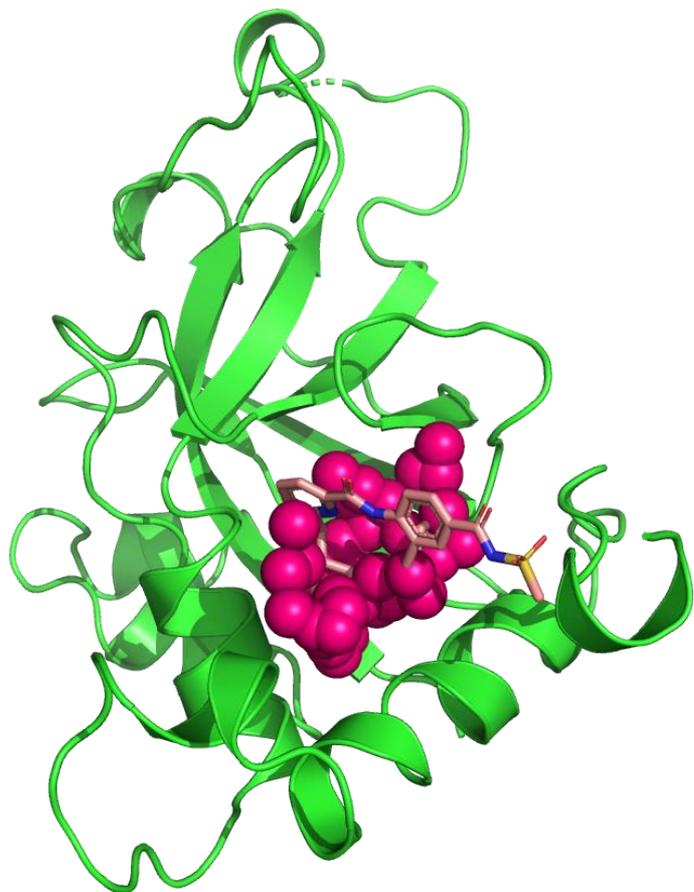
# Pocket



Morris, G. M. et al. Software news and updates AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem* (2009).

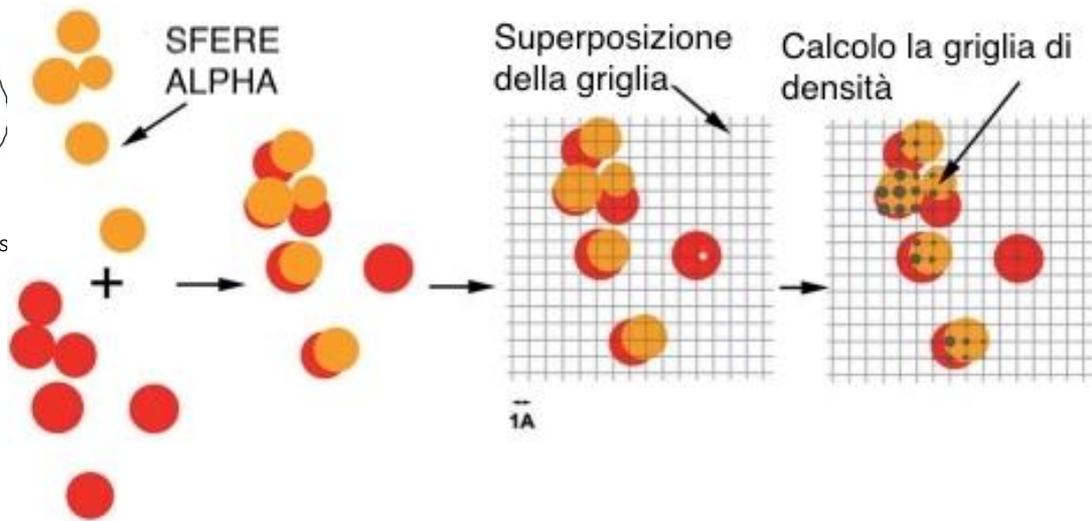
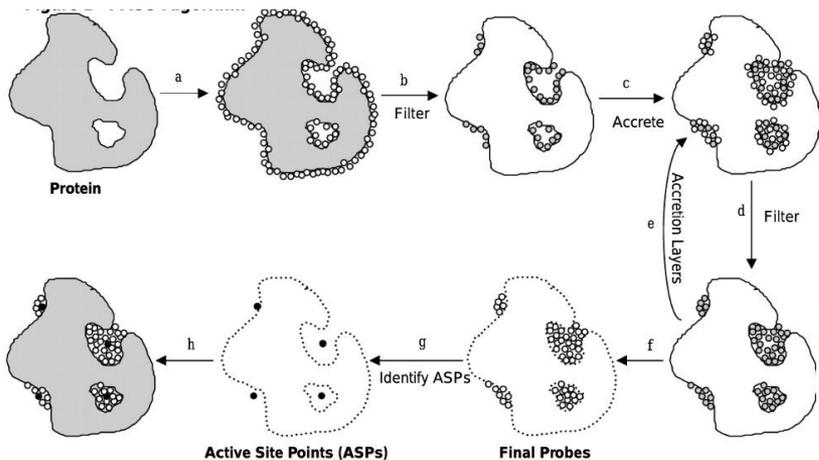


# Scopo dello studio





- Geometrici (Fpocket , Epos)
- Energetici (Autosite, Autogrid, Easymifs, Sitehound)
- Template-based
- Machine Learning e Deep Learning (P2Rank)



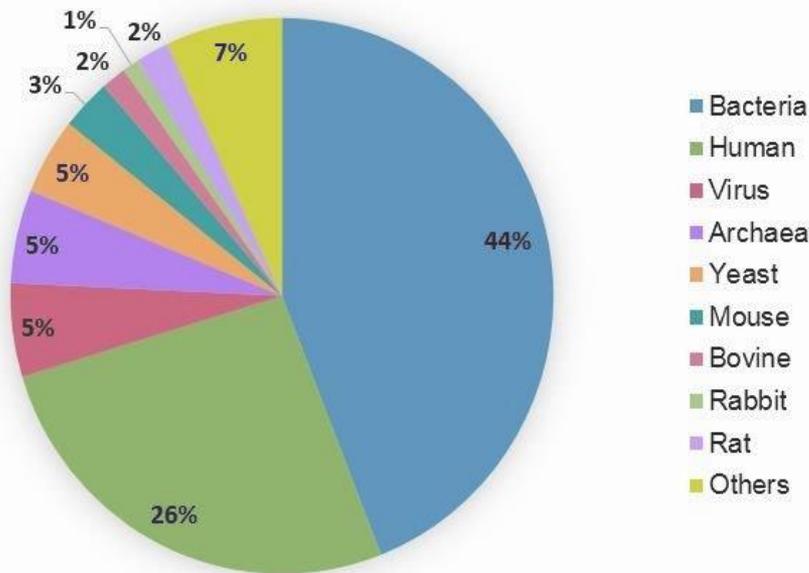
Brady, G. P. & Stouten, P. F. W. Fast prediction and visualization of protein binding pockets with PASS. *J. Comput. Aided. Mol. Des.* **14**, 383–401 (2000).

Le Guilloux, V., Schmidtke, P. & Tuffery, P. Fpocket: An open-source platform for ligand pocket detection. *BMC Bioinformatics*, 1–11 (2009).

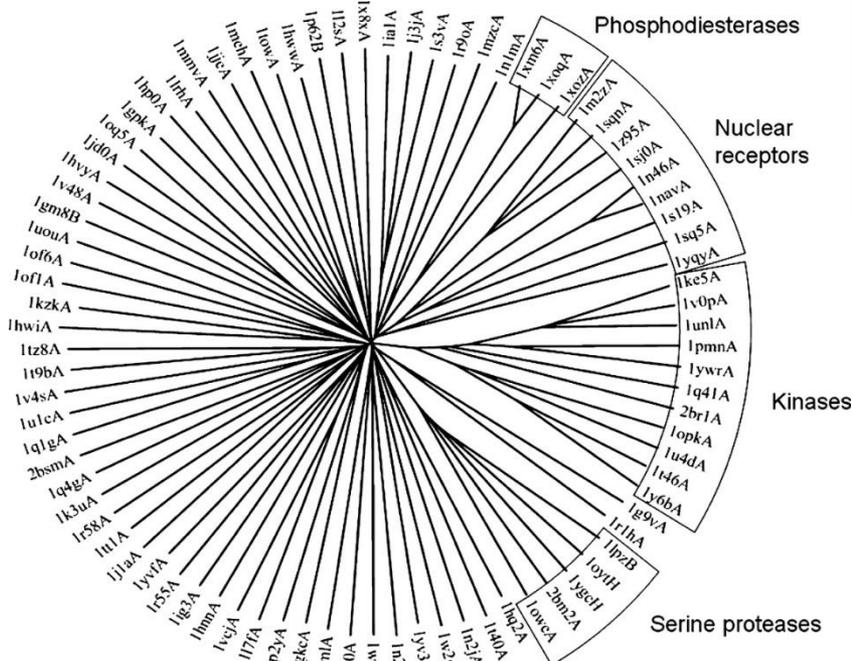


# Dataset

- ❖ ASBench (180 PDB)
- ❖ ASTEX (77 PDB)
- ❖ CHEN11 (165 PDB)
- ❖ COACH420 (601 PDB)
- ❖ HCV NS3 (65 PDB)
- ❖ Drug-Target 198 (143 PDB)
- ❖ UB-48



Huang, Z. et al. ASD v2.0: Updated content and novel features focusing on allosteric regulation. *Nucleic Acids Res.*



Hartshorn, M. J. et al. Diverse, high-quality test set for the validation of protein-ligand docking performance. *J. Med. Chem.* (2007)



# Metriche



ROME CENTER FOR MOLECULAR DESIGN

- **OLP** (Overlap between the predicted binding site and the ligand)

$$OLP = \frac{|P \cup L| - |P \cap L|}{|P \cup L|}$$

$$OLP = \frac{|P \cup P'| - |P \cap P'|}{|P \cup P'|}$$

- **DCA** (The minimal distance between the center of the predicted binding site and any atom of the ligand)

$$DCA = \min \left( \sqrt{(\bar{P}_x - L_x)^2 + (\bar{P}_y - L_y)^2 + (\bar{P}_z - L_z)^2} \right) \quad DCA = \min \left( \sqrt{(\bar{P}_x - P'_x)^2 + (\bar{P}_y - P'_y)^2 + (\bar{P}_z - P'_z)^2} \right)$$

- **DCC** (The distance from the center of the predicted binding site to the center of the ligand)

$$DCC = \sqrt{(\bar{P}_x - \bar{L}_x)^2 + (\bar{P}_y - \bar{L}_y)^2 + (\bar{P}_z - \bar{L}_z)^2}$$

$$DCC = \sqrt{(\bar{P}_x - \bar{P}'_x)^2 + (\bar{P}_y - \bar{P}'_y)^2 + (\bar{P}_z - \bar{P}'_z)^2}$$

- **LCV / LCR** (Ligand Coverage Volume / Ligand Coverage Ratio)

$$LCR = \frac{|P \cap L|}{|L|}$$

$$LCR = \frac{|P \cap P'|}{|P'|}$$

# Percentuali di pocket identificate correttamente da ogni software

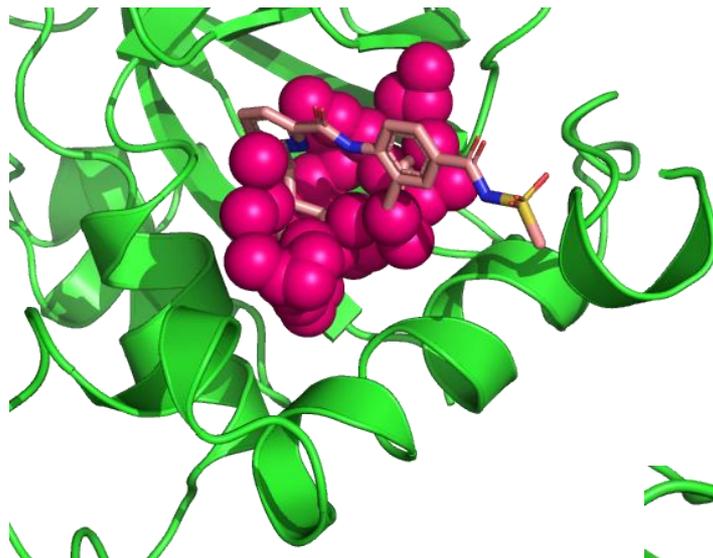
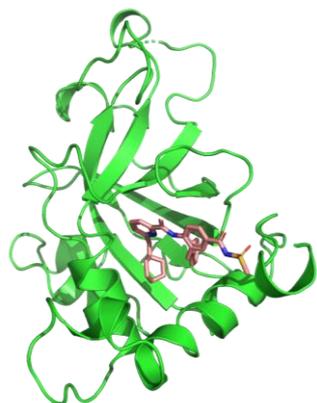


- ❖ Dal lavoro precedente si è evidenziato che il software che identificava con maggior successo la tasca di legame per il ligando farmacologico è P2Rank.

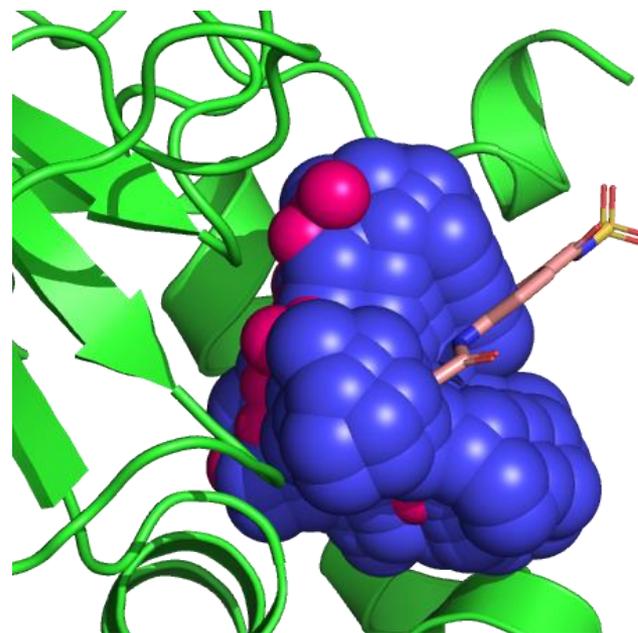
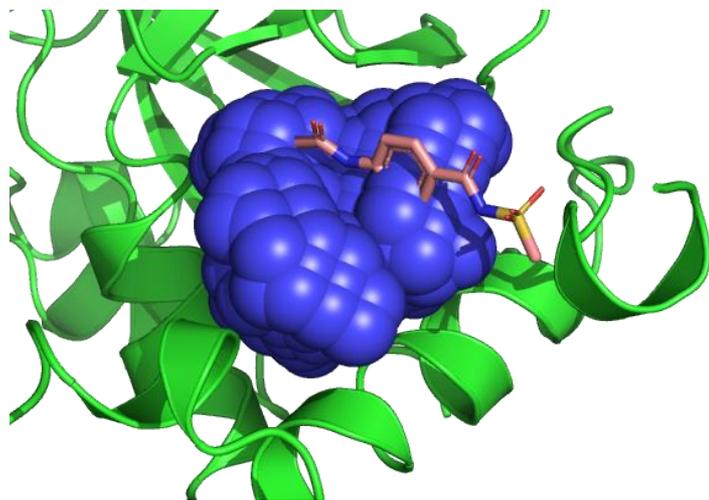
	AUTOSITE	EPOS	FPOCKET	P2RANK	SITE_GRID	SITE_MIFS
ASBench	53.88%	46.11%	33.88%	61.66%	42.22%	16.11%
ASTEX	79.22%	64.93%	59.74%	90.90%	67.53%	22.07%
CHEN11	62.42%	59.39%	48.48%	84.24%	41.21%	20%
COACH420	69.05%	57.23%	53.24%	86.18%	42.59%	21.13%
NS3_HCV	38.46%	38.46%	38.46%	55.38%	20%	30.76%
Dt-198	56.64%	53.84%	67.13%	74.12%	50.34%	18.88%



# Esempio di Overlap

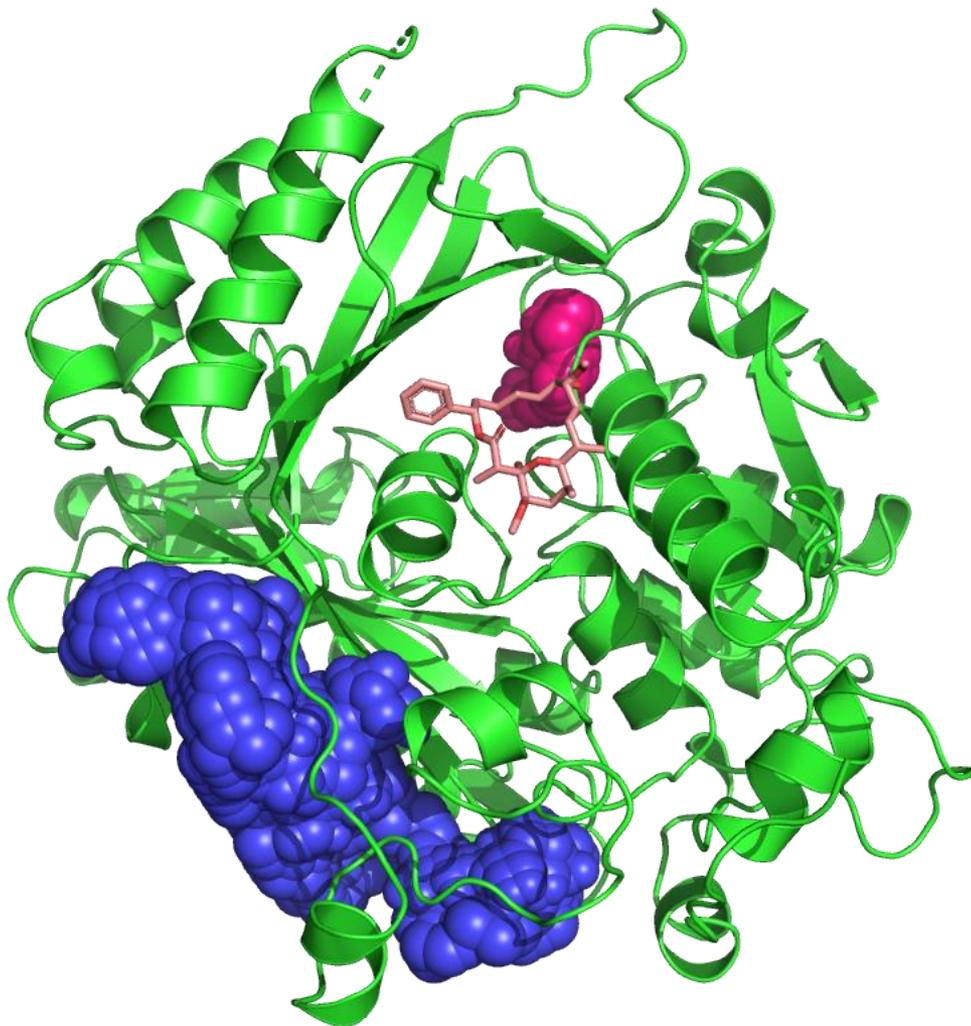


PDB 4p2t , dove in blu si ha prima pocket identificata da Autosite. Mentre in magenta la prima pocket individuata da P2Rank. In basso viene mostrato l'overlap tra le due tasche, nella zona dove si lega il ligando.





# Esempio di Mancato Overlap



PDB 1w96. In blu è mostrata la prima tasca identificata da Autosite e in magenta la prima pocket individuata da Epos. Si può notare come in questo caso non c'è overlap tra le due tasche.

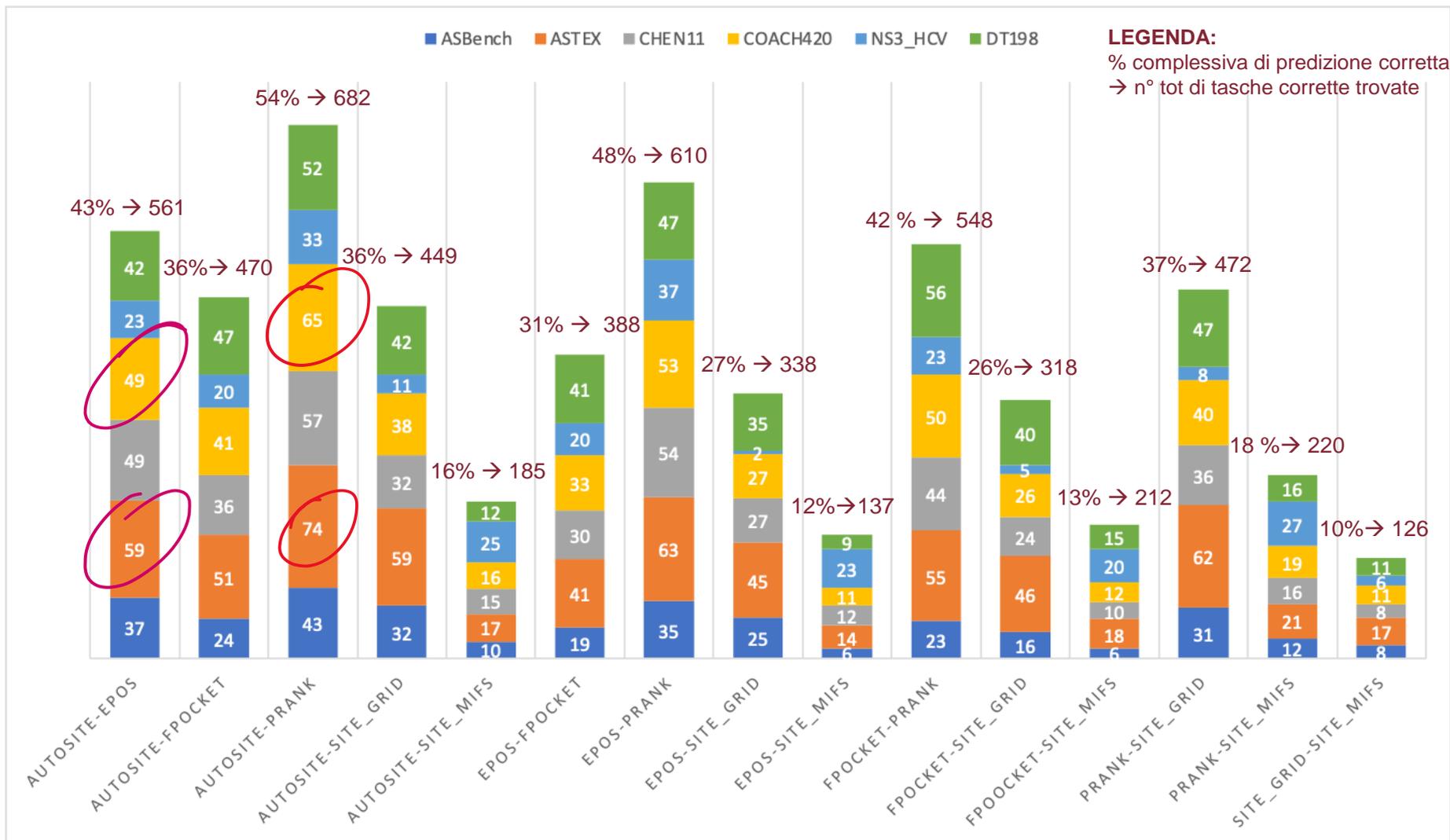


# Risultati OLP : Percentuali di successo nel predire la tasca corretta



➤ La combinazione di software con il più alto tasso di successo in tutti i dataset esaminati è P2Rank-Autosite

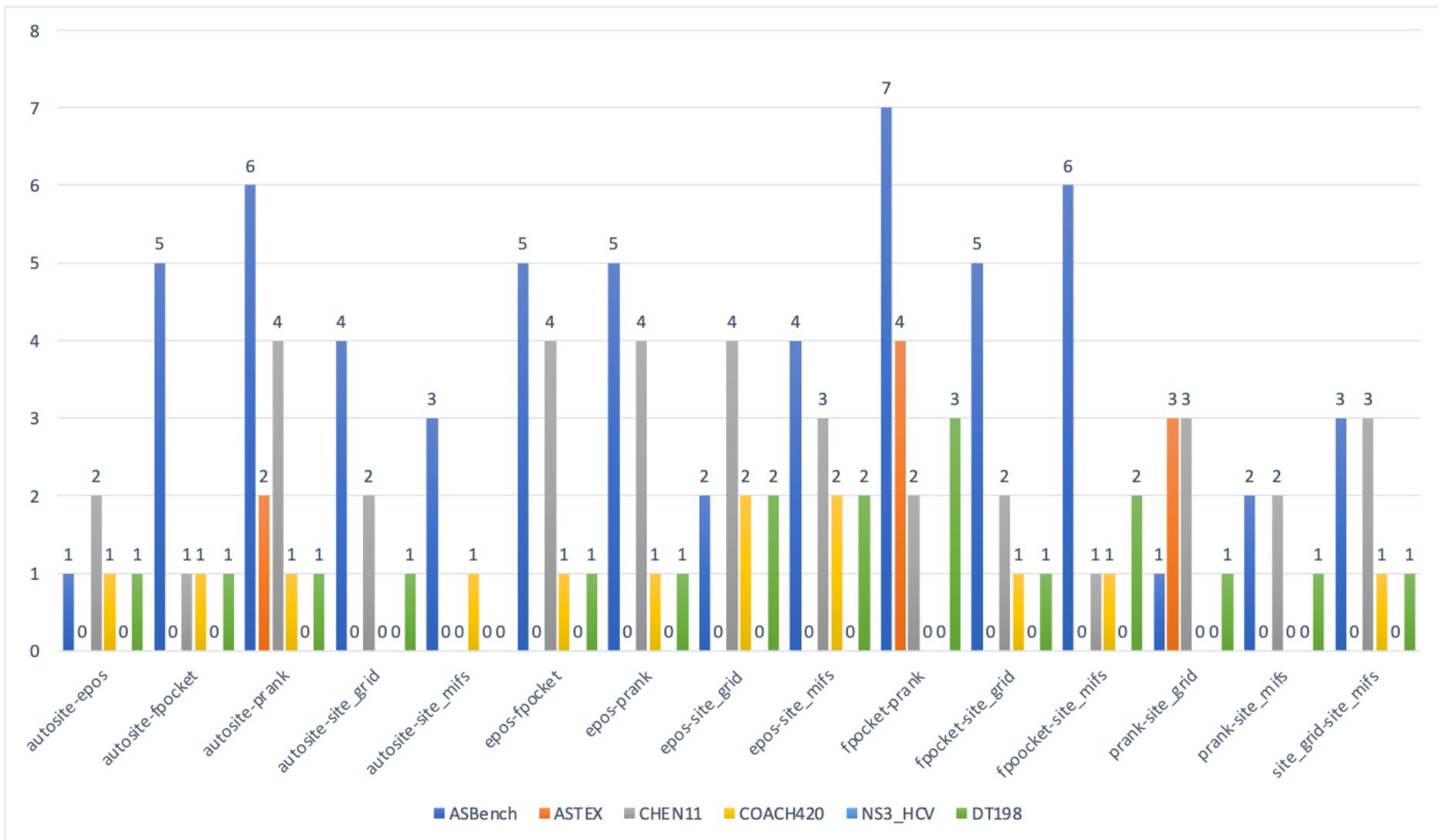
$$OLP = \frac{|P \cup P'| - |P \cap P'|}{|P \cup P'|}$$



# Risultati OLP : Percentuale di pocket corrette che non si sovrappongono



➤ Il Dataset che ha le percentuali più alte è ASBench, mentre il dataset che non presenta questa situazione è HCV\_NS3

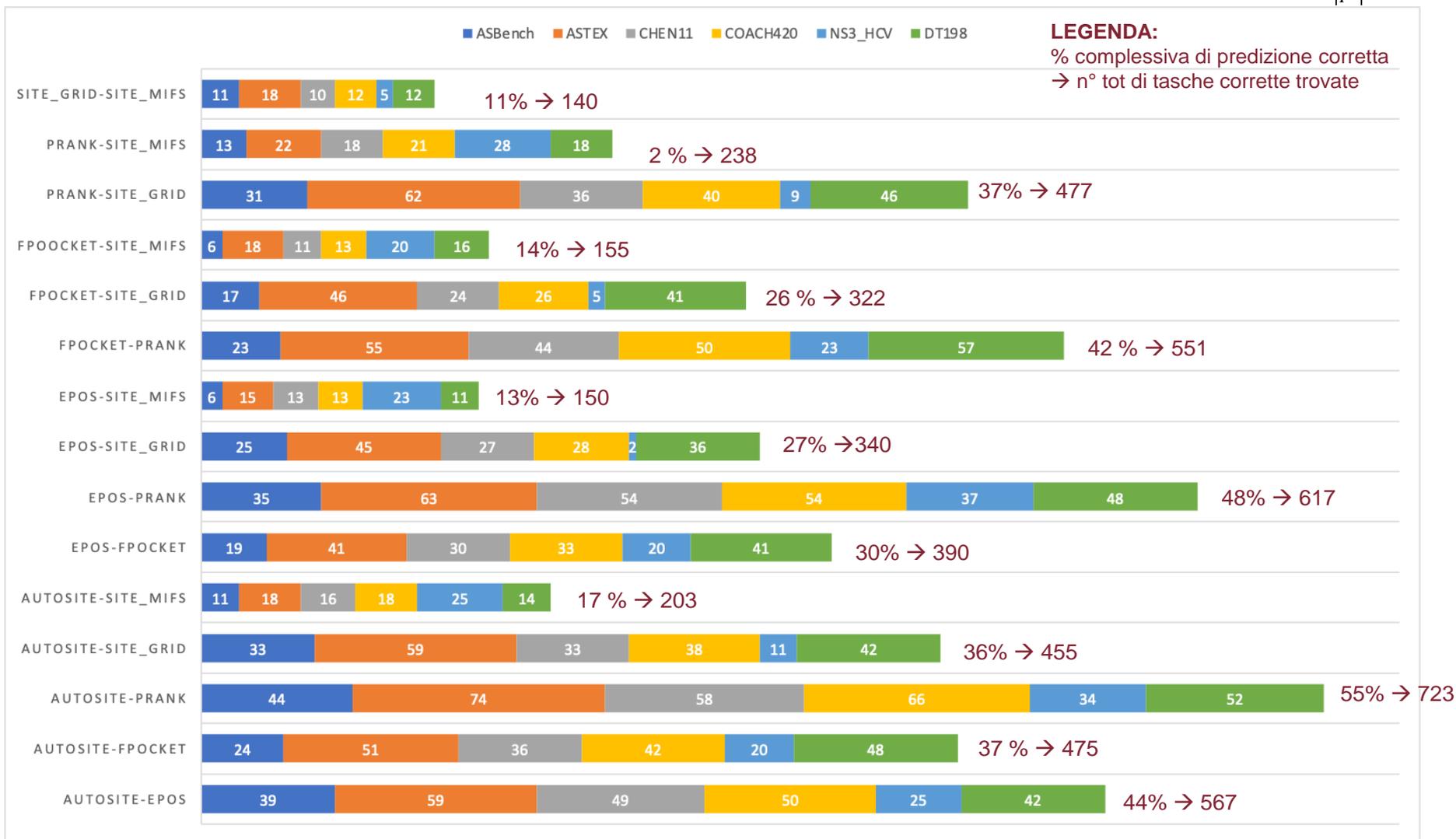


# Risultati LCR : Percentuali di successo nel predire la tasca corretta



➤ I risultati ottenuti per questa metrica sono coincidenti con quelli di OLP

$$LCR = \frac{|P \cap P'|}{|P'|}$$



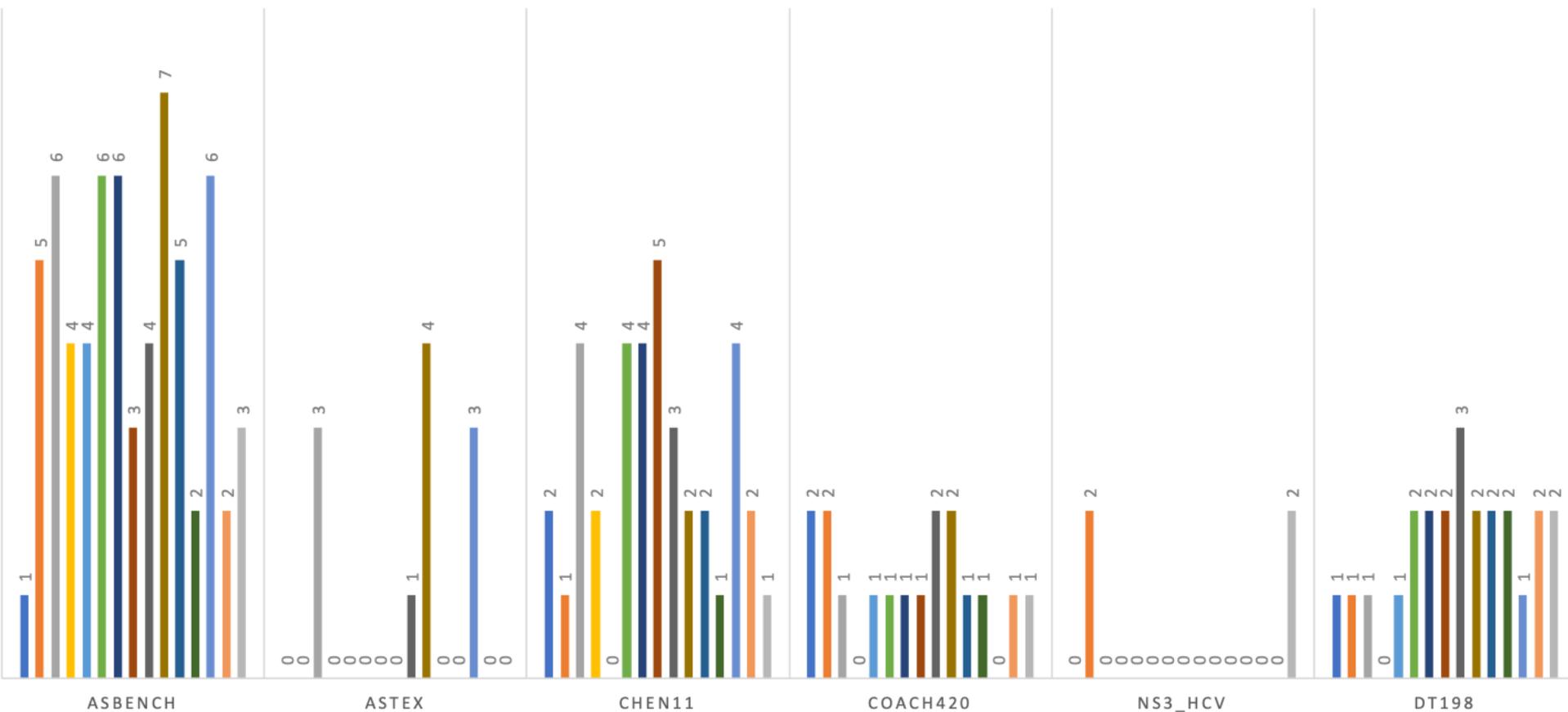


# Risultati LCR : Percentuali di pocket corrette che non si sovrappongono



- Il dataset che presenta il numero più alto di pocket, con un OLP riferito al ligando minore di 1, ma che non si sovrappongono è ASBench

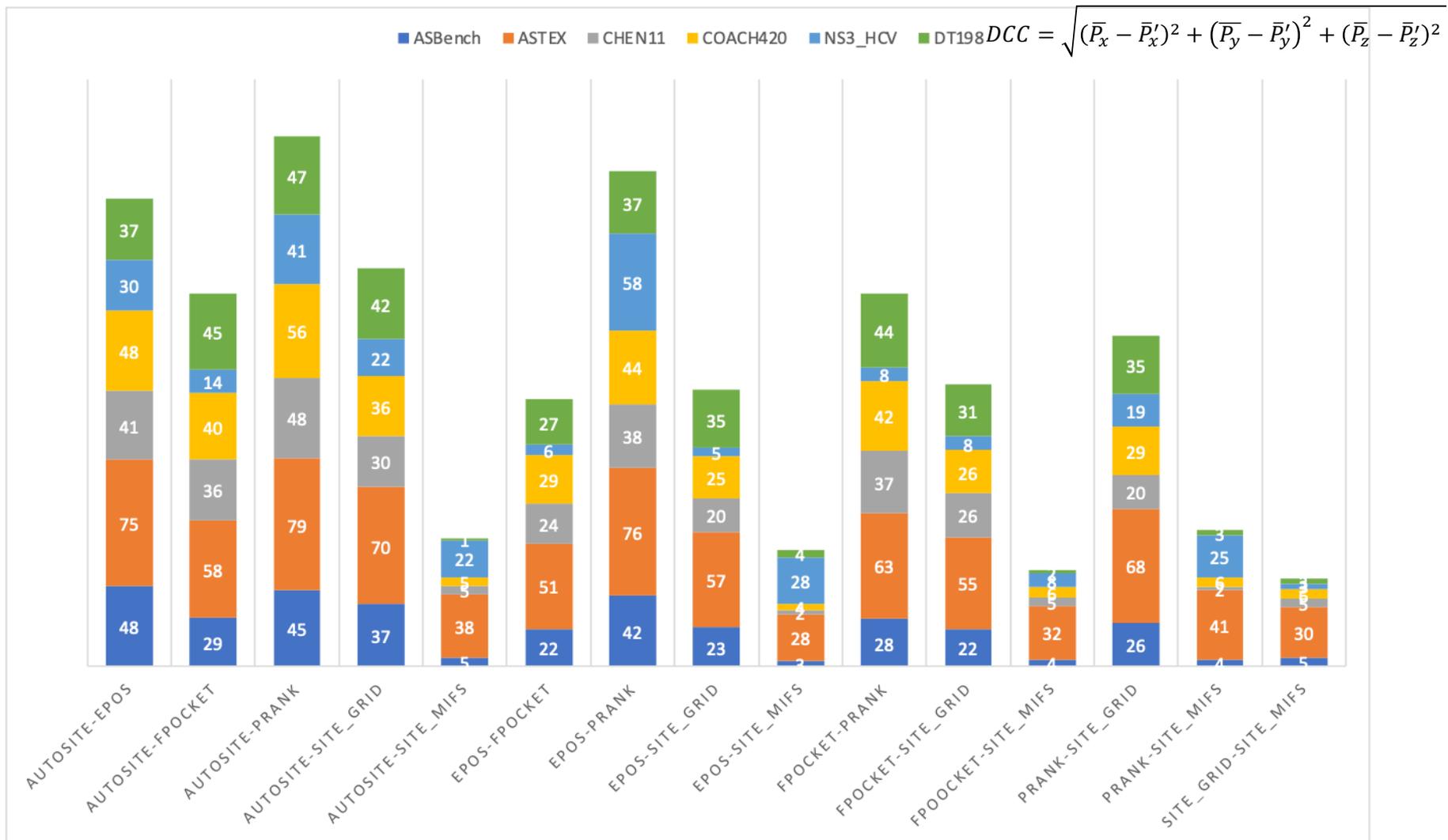
■ autosite-epos    ■ autosite-fpocket    ■ autosite-prank    ■ autosite-site\_grid    ■ autosite-site\_mifs    ■ epos-fpocket    ■ epos-prank    ■ epos-site\_grid  
■ epos-site\_mifs    ■ fpocket-prank    ■ fpocket-site\_grid    ■ fpocket-site\_mifs    ■ prank-site\_grid    ■ prank-site\_mifs    ■ site\_grid-site\_mifs





# Risultati DCC - Cutoff 5 Å

Con l'utilizzo del cutoff proposto dalla letteratura, i risultati ottenuti sono inconsistenti e non possono essere utilizzati





# Risultati DCC - Nuovo Cutoff



- DCC Medio è stato calcolato attraverso la media dei valori di DCC appartenenti solamente di quei PDB che presentavano un valori di OLP inferiore a 1
- In questo caso è stato difficile trovare un cutoff univoco da applicare a tutti i dataset in esame

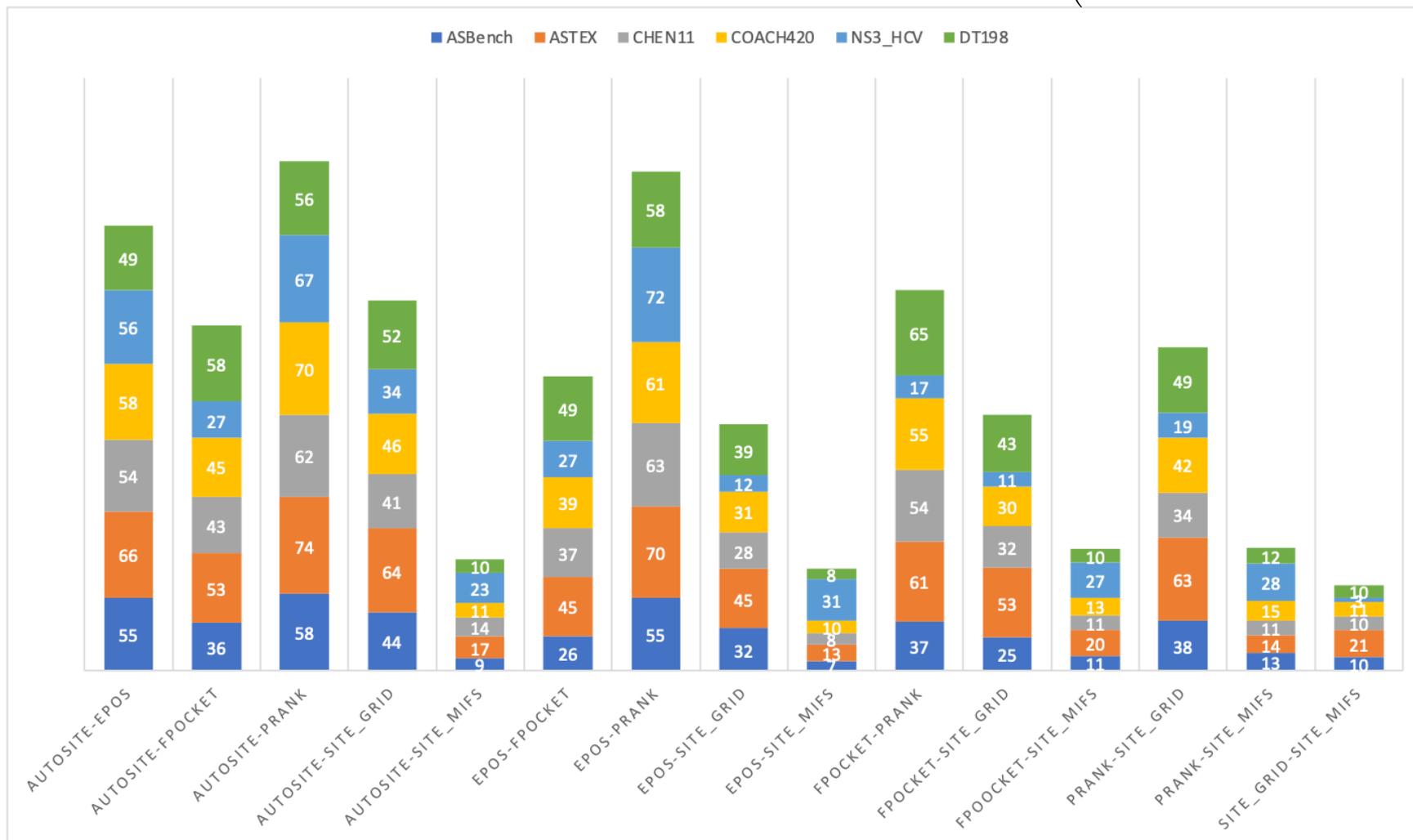
DATASET	DCC Medio	Std
ASBench	9.7	6.8
ASTEX	4.6	3.9
CHEN11	5.6	3.99
COACH420	6.9	4.4
NS3_HCV	6.1	2.8
Dt-198	5.8	4.4



# Risultati DCA - Cutoff 5 Å

Con l'utilizzo del cutoff proposto dalla letteratura, i risultati ottenuti sono inconsistenti e non possono essere utilizzati.

$$DCA = \min \left( \sqrt{(\bar{P}_x - P'_x)^2 + (\bar{P}_y - P'_y)^2 + (\bar{P}_z - P'_z)^2} \right)$$





# Risultati DCA- Nuovo Cutoff

DATASET	DCA Medio	Std
ASBench	2.7	2.3
ASTEX	2.4	2.2
CHEN11	2.5	2.0
COACH420	2.8	2.3
NS3_HCV	2.2	2.2
Dt-198	3.0	2.5

- DCA Medio è stato calcolato attraverso la media dei valori di DCC appartenenti solamente di quei PDB che presentavano un valori di OLP inferiore a 1
- A differenza di DCC, qui si è riusciti a trovare un valore da applicare pari a 2,5 Å . Ma applicando questo cutoff le percentuali di successo sono scese più basse rispetto a quelle ottenute con il cutoff standard



# Conclusioni



- ◆ **OLP**, la metrica più rilevante, ha permesso di definire le 3 combinazioni di software da considerare per avere un buon tasso di successo nella previsione della tasca corretta:

- 1) *P2Rank- Autosite*
- 2) *Autosite-Epos*
- 3) *Epos-Fpocket*

- ◆ **LCR**, metrica che indica la percentuale di volume di una pocket coperta dall'altra, ha confermato i risultati ottenuti con OLP.

- ◆ **DCC** e **DCA** non sono metriche adatte alla valutazione della performance dei software

## Prospettive future

- Ricalcolare tutte le metriche utilizzate in questo studio per il Dataset ASBench considerando solamente i ligandi allosterici
- Condurre altre analisi sui dati ottenuti in questo studio



# Grazie per l'attenzione!