# Il portale 3D-QSAR.com per lo Sviluppo di Modelli Ligand-Based e Structure-Based Robusti, Predittivi e Riproducibili. Applicazione ad una serie di ligandi allosterici delle chinasi TYK2


SAPIENZA
UNIVERSITÀ DI ROMA

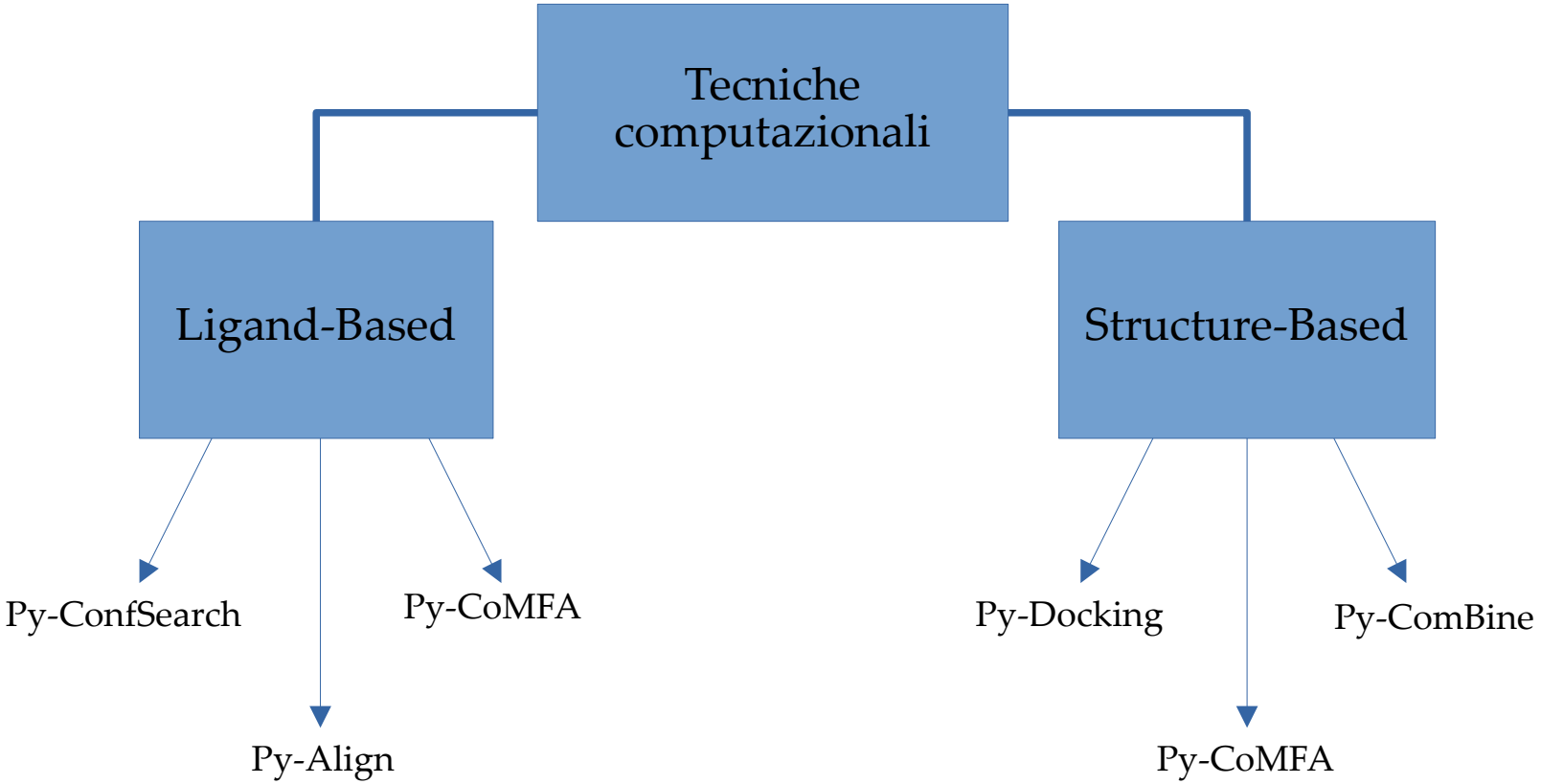
Facoltà di Farmacia e Medicina
Corso di Laurea in Chimica e Tecnologie Farmaceutiche
Tesi Sperimentale in Chimica Farmaceutica
a.a. 2021/2022

Laureando: Carlotta Coggio
Matricola: 1606333

Relatore: prof. Rino Ragno

# Obiettivo dello studio

Verificare e validare la riproducibilità di modelli LB e SB

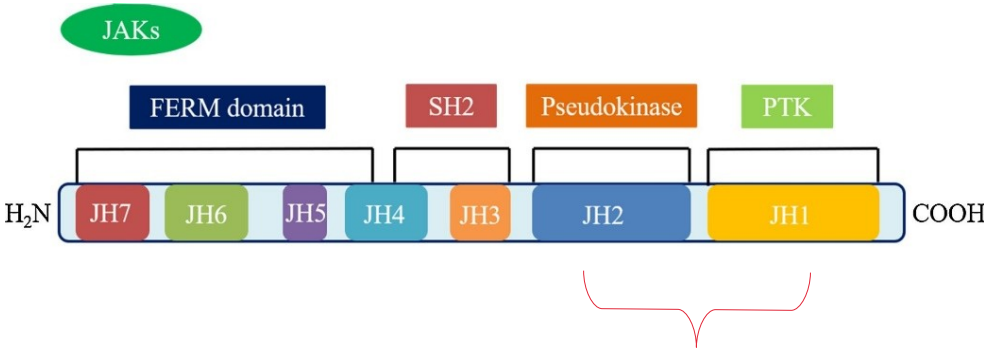


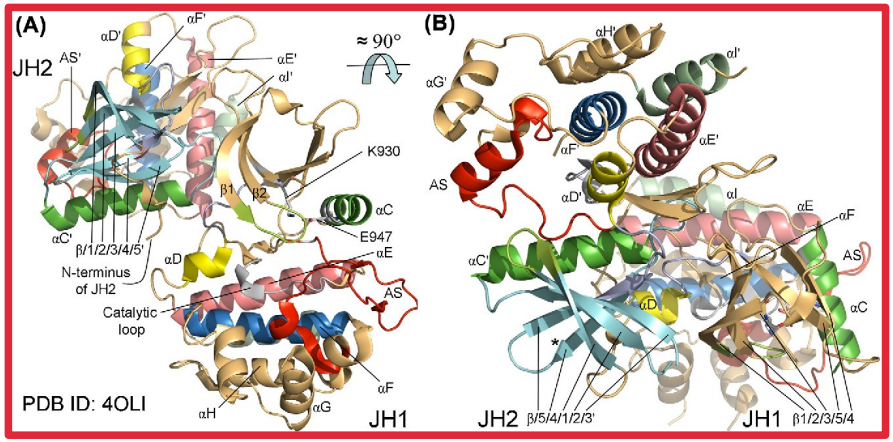
# Target: TYK2

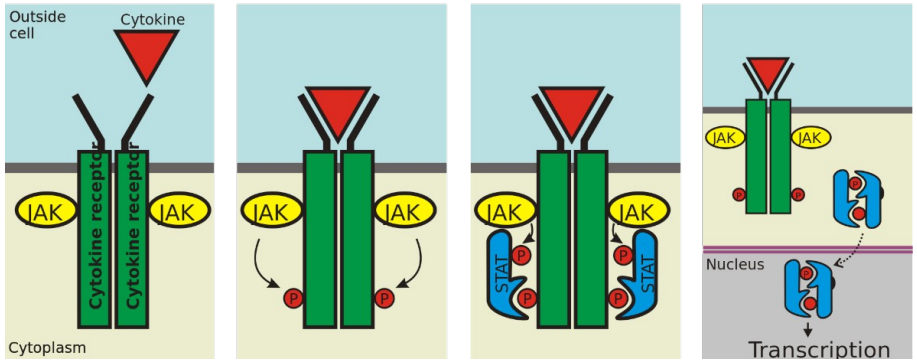
Le Tyk2 sono un membro della famiglia delle Janus Kinase (JAKs), ovvero una famiglia di tirosin chinasi non-recettoriale.

Strutturalmente presenta:

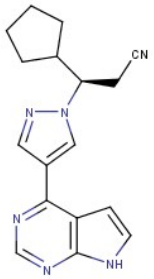
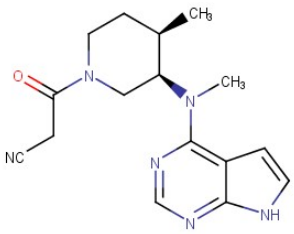
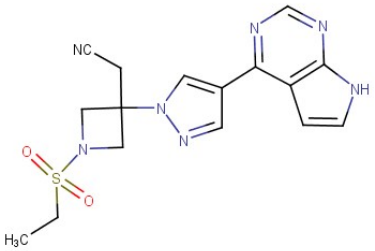
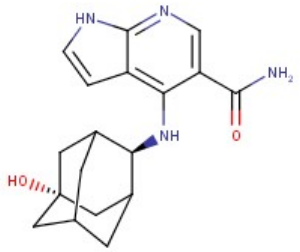
- sette regioni omologhe‘’JAK homology‘’(JH);
- la porzione carbossiterminale che comprende JH1 e JH2;
- i domini JH3, JH4 e JH5 che presentano una certa omologia con i domini SH2 (Src-homology-2);
- la porzione amino terminale, caratterizzata dai domini JH6 e JH7 e dalla zona FERM.



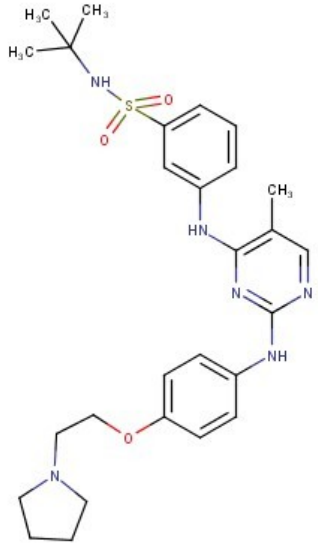
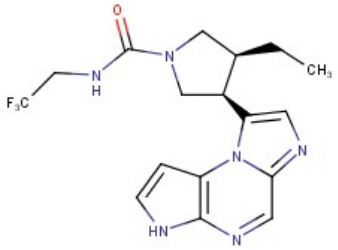



# Inibitori delle JAKs

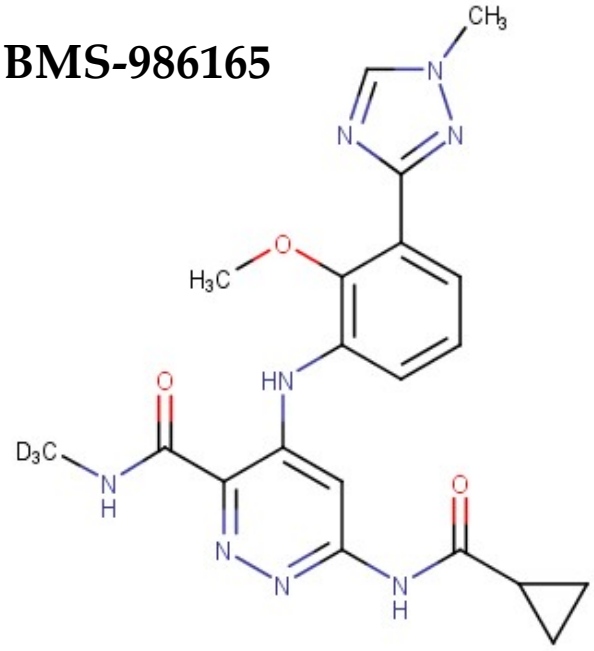





Inibitori NON SELETTIVI: Ruxolitinib, Tofacitinib, Baricitinib e Peficitinib

Inibitori SELETTIVI: Fedratinib, Upadacitinib

**BMS-986165**

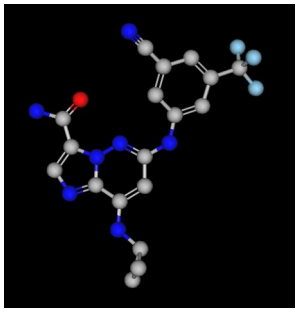


# Ligand-Based

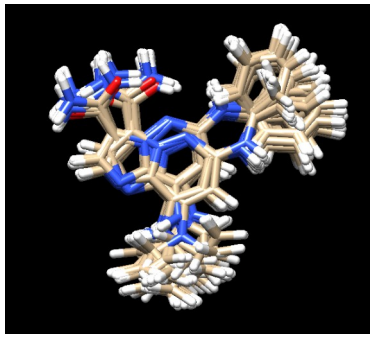
## Creazione del dataset

| SMILES | IC50(nM) |
|---|---|
| NC(=O)C1=CN=C2N1N=C(NC1=CC(=CC(=C1)C#N)C(F)(F)F)C=C2NC1CC1 | 480 |
| NC(=O)C1=CN=C2N1N=C(NC1=CC=CC=C1)C=C2NC1CC1 | 32 |
| CC1=CC(NC2=NN3C(=CN=C3C(NC3CC3)=C2)C(N)=O)=CC(C)=C1 | 22 |
| CC1=CC(F)=CC(NC2=NN3C(=CN=C3C(NC3CC3)=C2)C(N)=O)=C1 | 24 |
| NC(=O)C1=CN=C2N1N=C(NC1CCCCC1)C=C2NC1CC1 | 220 |
| CNC1=CC(NC2=CC(C)=CC(C)=C2)=NN2C(=CN=C12)C(N)=O | 7 |
| CCNC1=CC(NC2=CC(C)=CC(C)=C2)=NN2C(=CN=C12)C(N)=O | 17 |
| CC1=CC(NC2=NN3C(=CN=C3C(NC3CCC3)=C2)C(N)=O)=CC(C)=C1 | 110 |

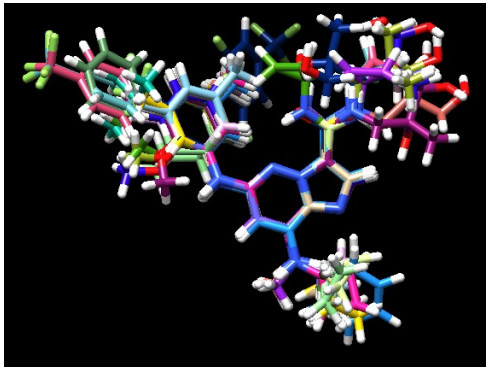
## Conversione delle SMILES in 3D



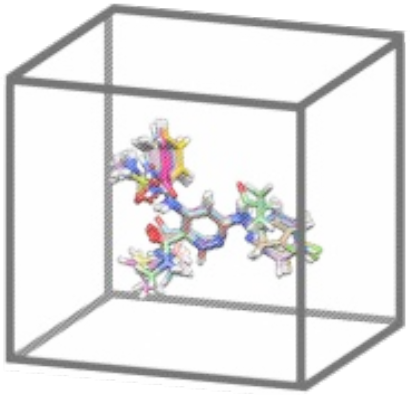
## Analisi conformazionale



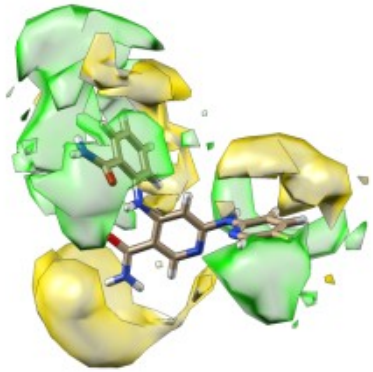
## Allineamenti



## Calcolo dei MIF



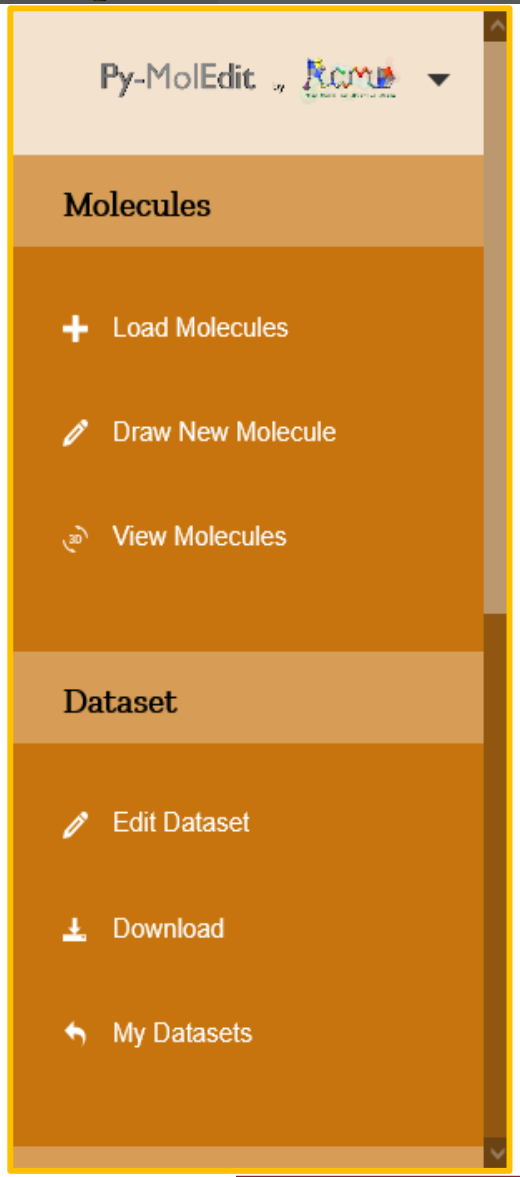
## Analisi grafica

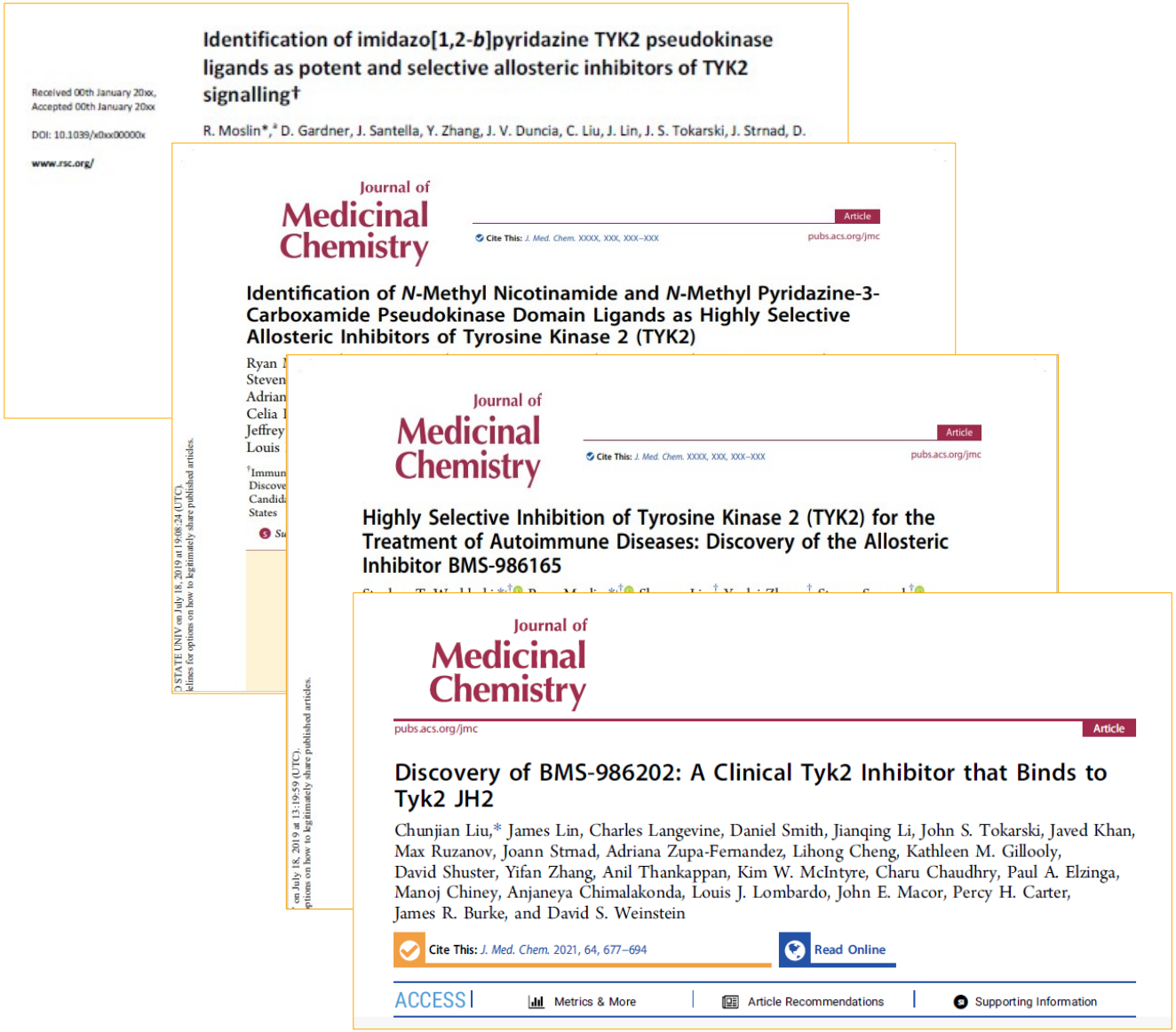



Per lo studio Ligand-Based utilizziamo i moduli:

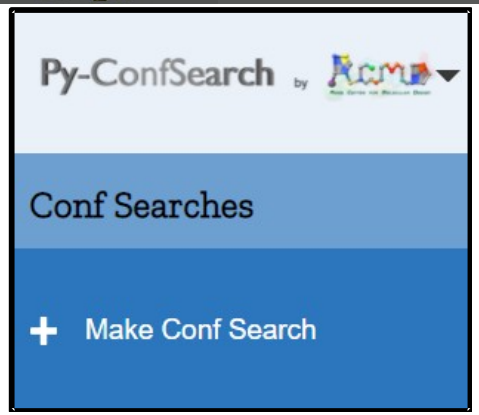
**Py-MolEdit**
**Py-ConfSearch**
**Py-Align**
**Py-CoMFA**

# Py-MolEdit


Py-MolEdit
Molecules
Load Molecules
Draw New Molecule
View Molecules
Dataset
Edit Dataset
Download
My Datasets


Identification of imidazo[1,2-b]pyridazine TYK2 pseudokinase ligands as potent and selective allosteric inhibitors of TYK2 signalling†
Received 00th January 20xx,
Accepted 00th January 20xx
DOI: 10.1039/x0xx00000x
www.rsc.org/
R. Moslin*,‡ D. Gardner, J. Santella, Y. Zhang, J. V. Duncia, C. Liu, J. Lin, J. S. Tokarski, J. Strnad, D.
Journal of Medicinal Chemistry
Article
Cite This: J. Med. Chem. XXXX, XXX, XXX−XXX
pubs.acs.org/jmc
Identification of N-Methyl Nicotinamide and N-Methyl Pyridazine-3-Carboxamide Pseudokinase Domain Ligands as Highly Selective Allosteric Inhibitors of Tyrosine Kinase 2 (TYK2)
Highly Selective Inhibition of Tyrosine Kinase 2 (TYK2) for the Treatment of Autoimmune Diseases: Discovery of the Allosteric Inhibitor BMS-986165
Discovery of BMS-986202: A Clinical Tyk2 Inhibitor that Binds to Tyk2 JH2
Chunjian Liu,* James Lin, Charles Langevine, Daniel Smith, Jianqing Li, John S. Tokarski, Javed Khan, Max Ruzanov, Joann Strnad, Adriana Zupa-Fernandez, Lihong Cheng, Kathleen M. Gillooly, David Shuster, Yifan Zhang, Anil Thankappan, Kim W. McIntyre, Charu Chaudhry, Paul A. Elzinga, Manoj Chiney, Anjaneya Chimalakonda, Louis J. Lombardo, John E. Macor, Percy H. Carter, James R. Burke, and David S. Weinstein
Cite This: J. Med. Chem. 2021, 64, 677−694
Read Online
ACCESS
Metrics & More
Article Recommendations
Supporting Information

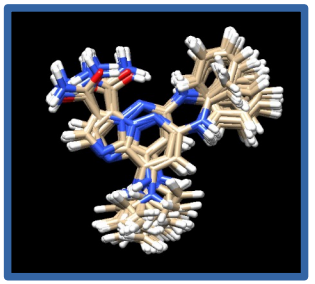
# Py-ConfSearch



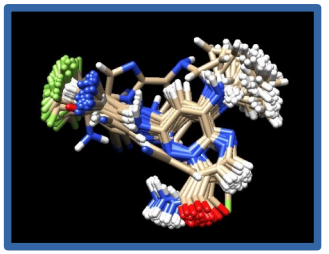

OpenBabel → EEM, MMFF94, SFKEEM

Balloon → UFF, MMFF94S, MMFF94, GAFF, GHEMICAL

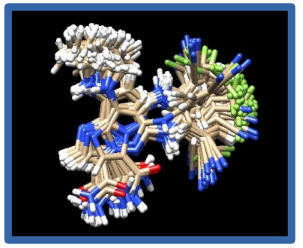
RDkit → UFF, MMFF94S, MMFF94



OpenBabel



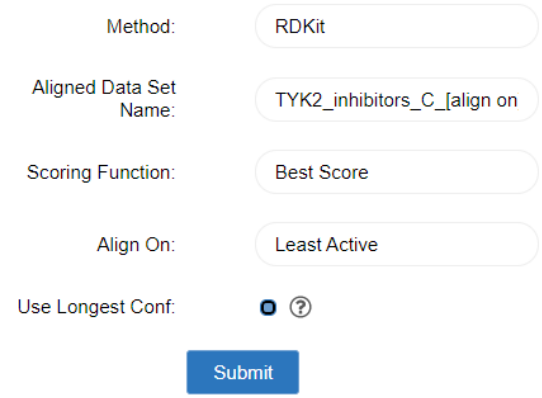
Baloon



RDkit

| Conf Search ID | Status | | Method | Max Confs | CG MaxIters | SD MaxIters | Force Field |
|---|---|---|---|---|---|---|---|
| 4751 | completed | Results | openbabel | 50 | 50 | 50 | UFF |
| 4752 | completed | Results | openbabel | 50 | 50 | 50 | MMFF94 |
| 4753 | completed | Results | openbabel | 50 | 50 | 50 | MMFF94s |
| 4754 | completed | Results | openbabel | 50 | 50 | 50 | GAFF |
| 4755 | completed | Results | openbabel | 50 | 50 | 50 | Ghemical |
| 4756 | completed | Results | balloon | 50 | 50 | 50 | EEM |
| 4757 | completed | Results | balloon | 50 | 50 | 50 | MMFF94 |
| 4758 | completed | Results | balloon | 50 | 50 | 50 | SFKEEM |
| 4759 | completed | Results | rdkit | 50 | 50 | | UFF |
| 4760 | completed | Results | rdkit | 50 | 50 | | MMFF94 |
| 4761 | completed | Results | rdkit | 50 | 50 | | MMFF94s |

# Py-Align

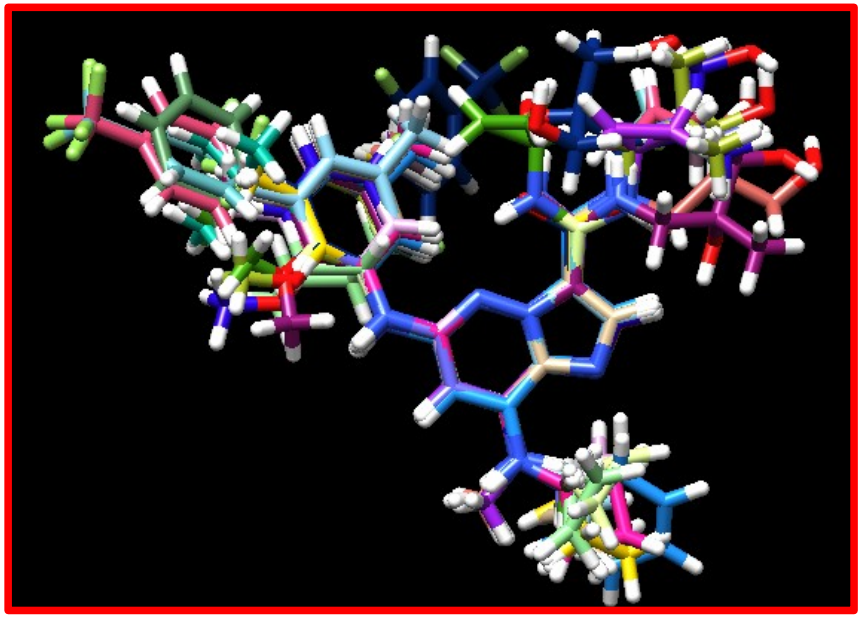




RDKit → BEST SCORE
LOWEST RMSD
TANIMOTO DISTANCE
PROTUDE DISTANCE

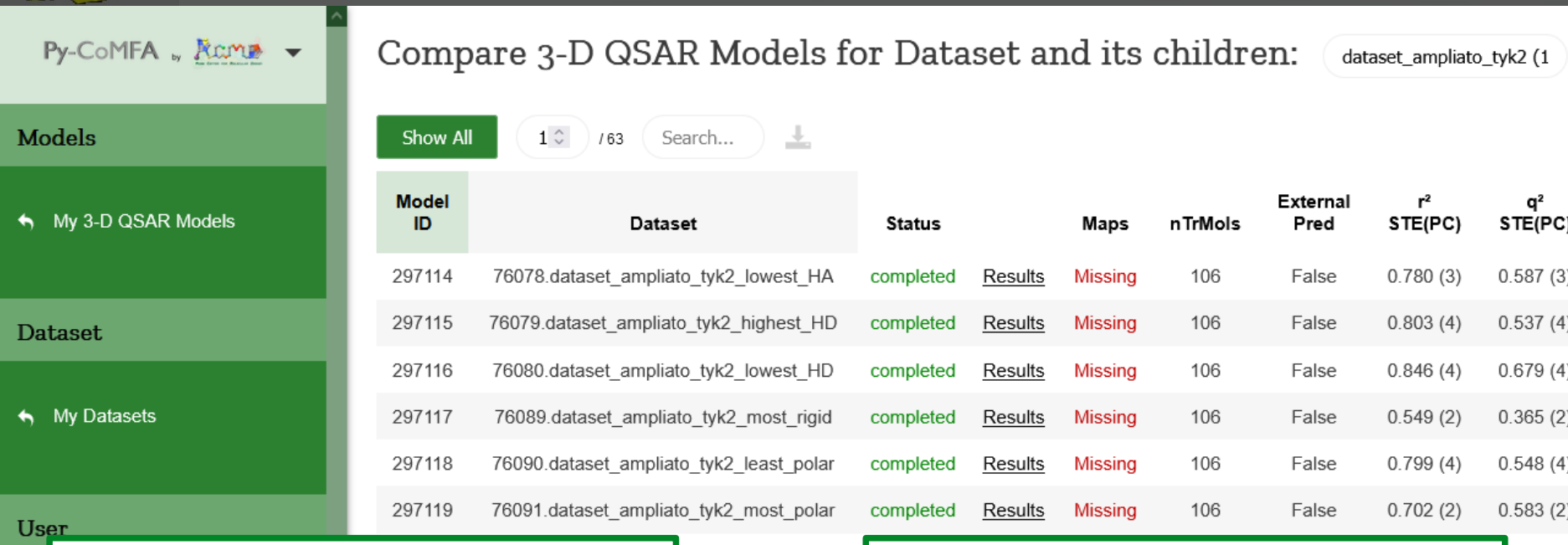
SHAEP → ONLY SHAEP SIMILARITY

1056 allineamenti totali (x3)



16 allineamenti ciascuno:
- least active;
- most active
-most flexible;
-most rigid;
-heaviest;
-longest;
-least polar;
-most polar;
-highest MR (Molecular Refractivity);
-lowest MR;
-highest HD (Hydrogen Donating groups);
-lowest HD;
-highest HA (Hydrogen Acceptor groups);
-lowest HA;
-highest LogP;
-lowest LogP;

# Py-CoMFA




| Model ID | Dataset | Status | | Maps | nTrMols | External Pred | $r^2$ STE(PC) | $q^2$ STE(PC) |
|---|---|---|---|---|---|---|---|---|
| 297114 | 76078.dataset_ampliato_tyk2_lowest_HA | completed | Results | Missing | 106 | False | 0.780 (3) | 0.587 (3) |
| 297115 | 76079.dataset_ampliato_tyk2_highest_HD | completed | Results | Missing | 106 | False | 0.803 (4) | 0.537 (4) |
| 297116 | 76080.dataset_ampliato_tyk2_lowest_HD | completed | Results | Missing | 106 | False | 0.846 (4) | 0.679 (4) |
| 297117 | 76089.dataset_ampliato_tyk2_most_rigid | completed | Results | Missing | 106 | False | 0.549 (2) | 0.365 (2) |
| 297118 | 76090.dataset_ampliato_tyk2_least_polar | completed | Results | Missing | 106 | False | 0.799 (4) | 0.548 (4) |
| 297119 | 76091.dataset_ampliato_tyk2_most_polar | completed | Results | Missing | 106 | False | 0.702 (2) | 0.583 (2) |
| | o_tyk2_highest_MR | | | | | | | .458 (4) |
| | o_tyk2_lowest_MR | | | | | | | .408 (3) |
| | o_tyk2_lowest_HA | | | | | | | .547 (4) |

**PARAMETRI DI DEFAULT**

**Probe**: C3

**Dieletric Constant**: 8

**Grid Spacing**: 2Å

**Cutoff Value**: 30 kcal/mol

**Minimum Sigma**: 0,05

**PARAMETRI VPO**

**Probe**: C3.H3; H.P; N.am; O.2

**Dieletric Constant**: (1; 81; 1)

**Grid Spacing**: (1; 3,6; 0,1)

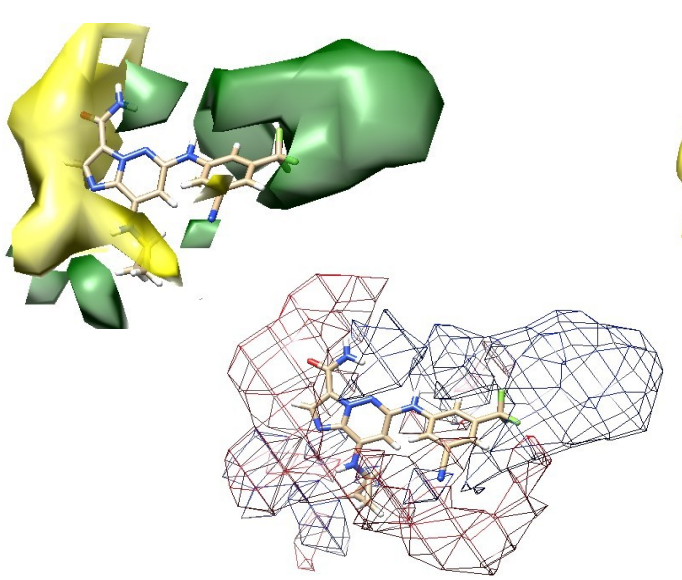
**Cutoff Value**: (5; 36; 1)

**Minimum Sigma**: (0,1; 1,1; 0,1)

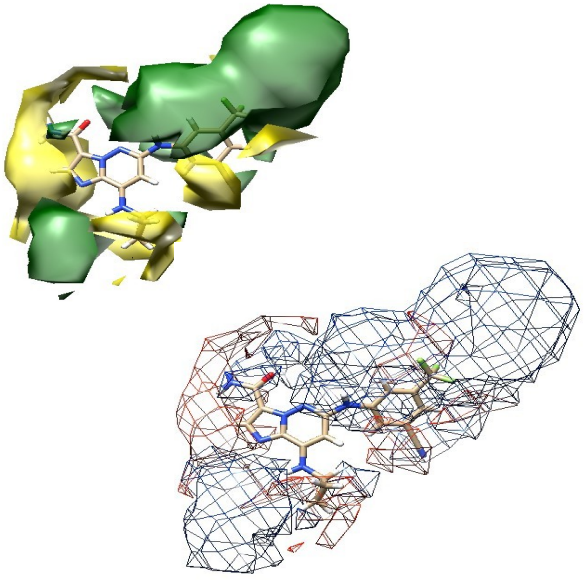
$$r^2 = 1 - \frac{\sum_{i=1}^{n} (y_{exp\_i} - y_{calc\_i})^2}{\sum_{i=1}^{n} (y_{exp\_i} - \bar{y})^2}$$

$$q^2 = 1 - \frac{\sum_{i=1}^{n} (y_{exp\_i} - y_{pred\_i})^2}{\sum_{i=1}^{n} (y_{exp\_i} - \bar{y})^2}$$

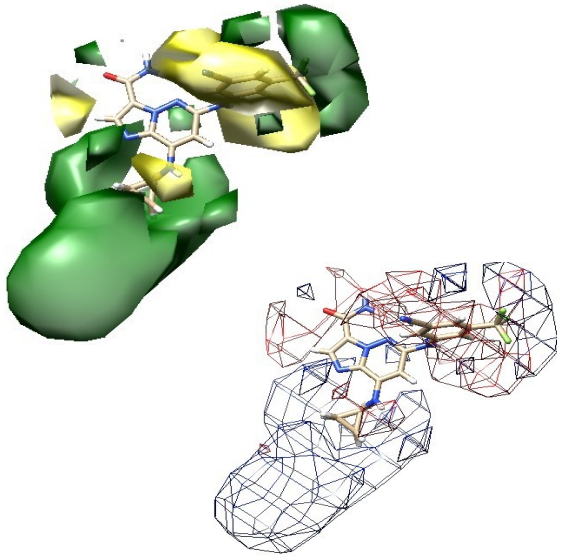
| Nome Dataset | ID | $r^2$ | $q^2$ | OPC |
|---|---|---|---|---|
| Tyk_2_inhibitors_M | 323310 | 0,800 | 0,663 | 4 |
| Tyk2_inhibitors_C | 318063 | 0,804 | 0,675 | 3 |
| Tyk2_inhibitors_S | 302986 | 0,709 | 0,507 | 3 |
| Media ± d.s. | | 0,771± 0,072 | 0,615± 0,077 | |



Tyk2_inhibitors_M



Tyk2_inhibitors_C

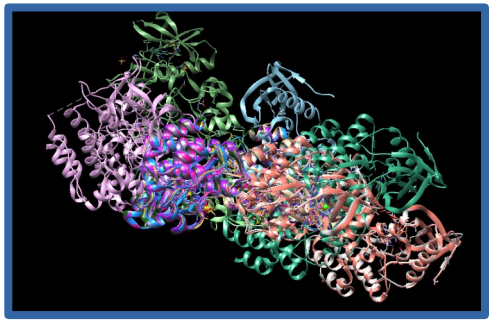


Tyk2_inhibitors_S

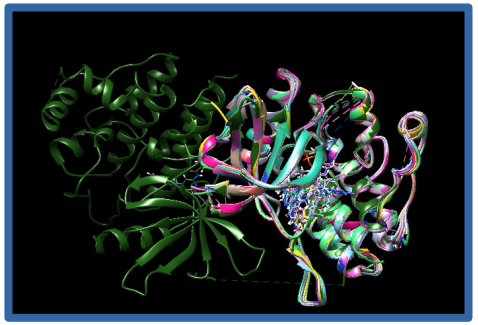
# Structure-Based

RCSB PDB PROTEIN DATA BANK

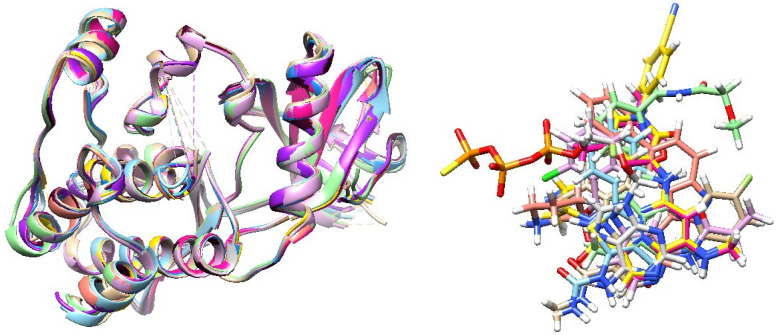
14 Cristalli



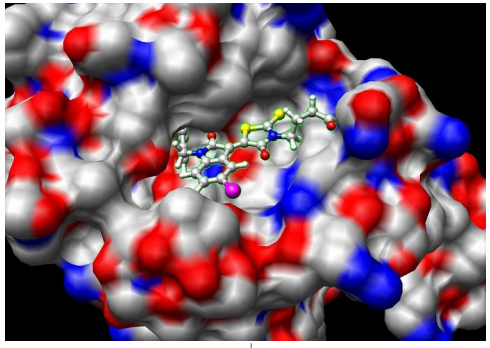
Allineamento



Separazione Key-Lock



Docking

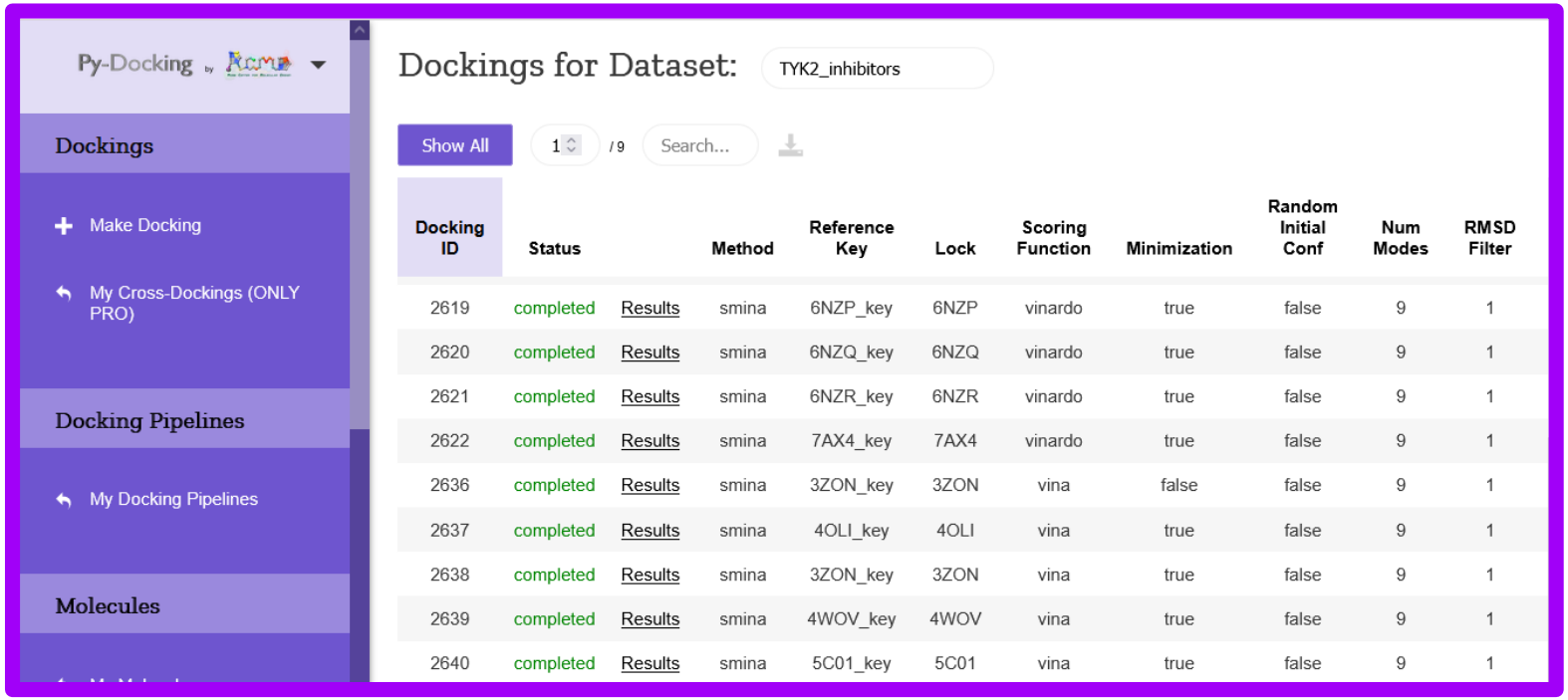


Risultati

COMBINE

CoMFA SB

# Py-Docking




| Docking ID | Status | | Method | Reference Key | Lock | Scoring Function | Minimization | Random Initial Conf | Num Modes | RMSD Filter |
|---|---|---|---|---|---|---|---|---|---|---|
| 2619 | completed | Results | smina | 6NZP_key | 6NZP | vinardo | true | false | 9 | 1 |
| 2620 | completed | Results | smina | 6NZQ_key | 6NZQ | vinardo | true | false | 9 | 1 |
| 2621 | completed | Results | smina | 6NZR_key | 6NZR | vinardo | true | false | 9 | 1 |
| 2622 | completed | Results | smina | 7AX4_key | 7AX4 | vinardo | true | false | 9 | 1 |
| 2636 | completed | Results | smina | 3ZON_key | 3ZON | vina | false | false | 9 | 1 |
| 2637 | completed | Results | smina | 4OLI_key | 4OLI | vina | true | false | 9 | 1 |
| 2638 | completed | Results | smina | 3ZON_key | 3ZON | vina | true | false | 9 | 1 |
| 2639 | completed | Results | smina | 4WOV_key | 4WOV | vina | true | false | 9 | 1 |
| 2640 | completed | Results | smina | 5C01_key | 5C01 | vina | true | false | 9 | 1 |

| Programma | Scoring function |
|---|---|
| SMINA | Vinardo<br>Vina<br>Ad4_scoring |
| PLANTS | Plp<br>plp95<br>Chemplp |

ECRD
RCRD
ECCD
RCCD

# Docking Accuracy

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2}$$

*DA = frmsd ≤ a + 0.5 (frmsd ≤ b – frmsd ≤ a)*

**Punteggi RMSD:**

- **tra 0 e 2 → 1**
- **tra 2 e 3 → 0,5**
- **≥ 3 → 0**

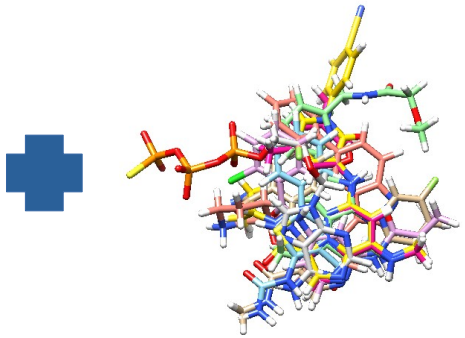
**PLANTS plp è il miglior metodo**

| METHOD | | ERCD | RCRD | ECCD | RCCD |
|---|---|---|---|---|---|
| **Plants plp** | **Media (%)** | **58,85** | **69,05** | **44,04** | **46,43** |
| | **d.s.** | **±7,73** | **±10,31** | **±7,44** | **±6,18** |
| Plants Chemplp | Media | 75,42 | 78,57 | 44,04 | 42,86 |
| | d.s. | ± 11,47 | ± 6,18 | ± 4,13 | ±7,15 |
| Plants plp95 | Media | 71,42 | 72,62 | 46,76 | 46,42 |
| | d.s. | ±0,01 | ±2,06 | ±5,92 | ±0,01 |
| Smina vina | Media | 35,71 | 39,28 | 35,71 | 34,52 |
| | d.s. | ±0,00 | ±3,57 | ±0,00 | ±4,12 |
| Smina vina minimizzata | Media | 9,52 | 8,33 | 0,00 | 4,76 |
| | d.s. | ±2,06 | ±2,06 | ±0,00 | ±4,12 |
| Smina vinardo | Media | 31,76 | 34,52 | 29,76 | 36,90 |
| | d.s. | ±14,92 | ±7,43 | ±8,99 | ±8,99 |
| Smina vinardo minimizzata | Media | 2,38 | 9,52 | 7,14 | 2,38 |
| | d.s. | ±4,12 | ±4,12 | ±7,14 | ±4,12 |
| Smina ad4_scoring | Media | 51,59 | 54,76 | 46,42 | 45,23 |
| | d.s. | ±5,46 | ±8,99 | ±0,01 | ±4,13 |
| Smina ad4_scoring minimizzata | Media | 8,33 | 16,66 | 1,19 | 2,38 |
| | d.s. | ±5,46 | ±5,45 | ±2,06 | ±2,06 |

# Ampliamento dello studio

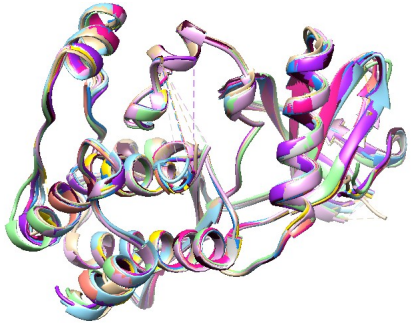
## Creazione dei dataset

| SMILES | IC50(nM) |
|---|---|
| NC(=O)C1=CN=C2N1N=C(NC1=CC(=CC(=C1)C#N)C(F)(F)F)C=C2NC1CC1 | 480 |
| NC(=O)C1=CN=C2N1N=C(NC1=CC=CC=C1)C=C2NC1CC1 | 32 |
| CC1=CC(NC2=NN3C(=CN=C3C(NC3CC3)=C2)C(N)=O)=CC(C)=C1 | 22 |
| CC1=CC(F)=CC(NC2=NN3C(=CN=C3C(NC3CC3)=C2)C(N)=O)=C1 | 24 |
| NC(=O)C1=CN=C2N1N=C(NC1CCCCC1)C=C2NC1CC1 | 220 |
| CNC1=CC(NC2=CC(C)=CC(C)=C2)=NN2C(=CN=C12)C(N)=O | 7 |
| CCNC1=CC(NC2=CC(C)=CC(C)=C2)=NN2C(=CN=C12)C(N)=O | 17 |
| CC1=CC(NC2=NN3C(=CN=C3C(NC3CCC3)=C2)C(N)=O)=CC(C)=C1 | 110 |



14 dataset

Docking: PLANTS plp



Utilizzando la key corrispondente come riferimento

14 Best Docked Dataset

**COMBINE**

Risultati

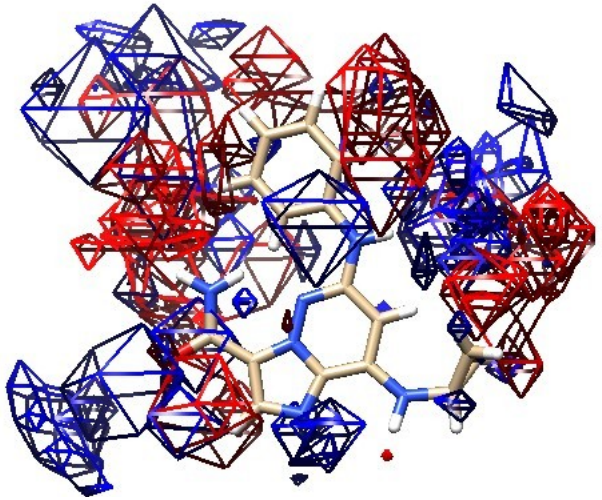
**CoMFA SB**

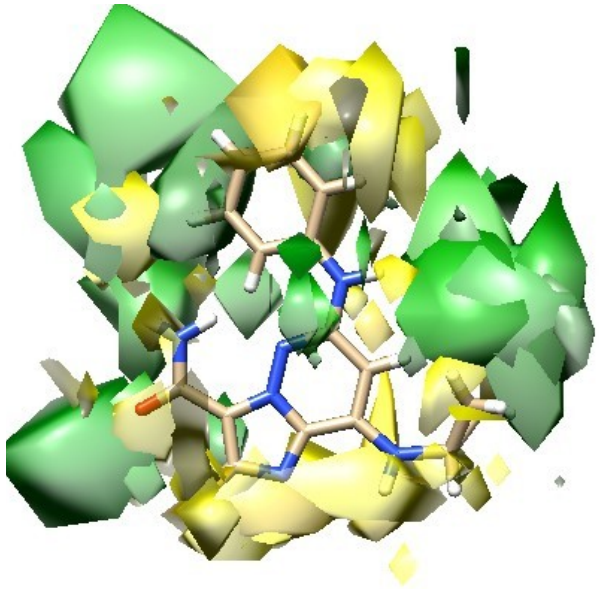
Risultati

# Modelli SB CoMFA

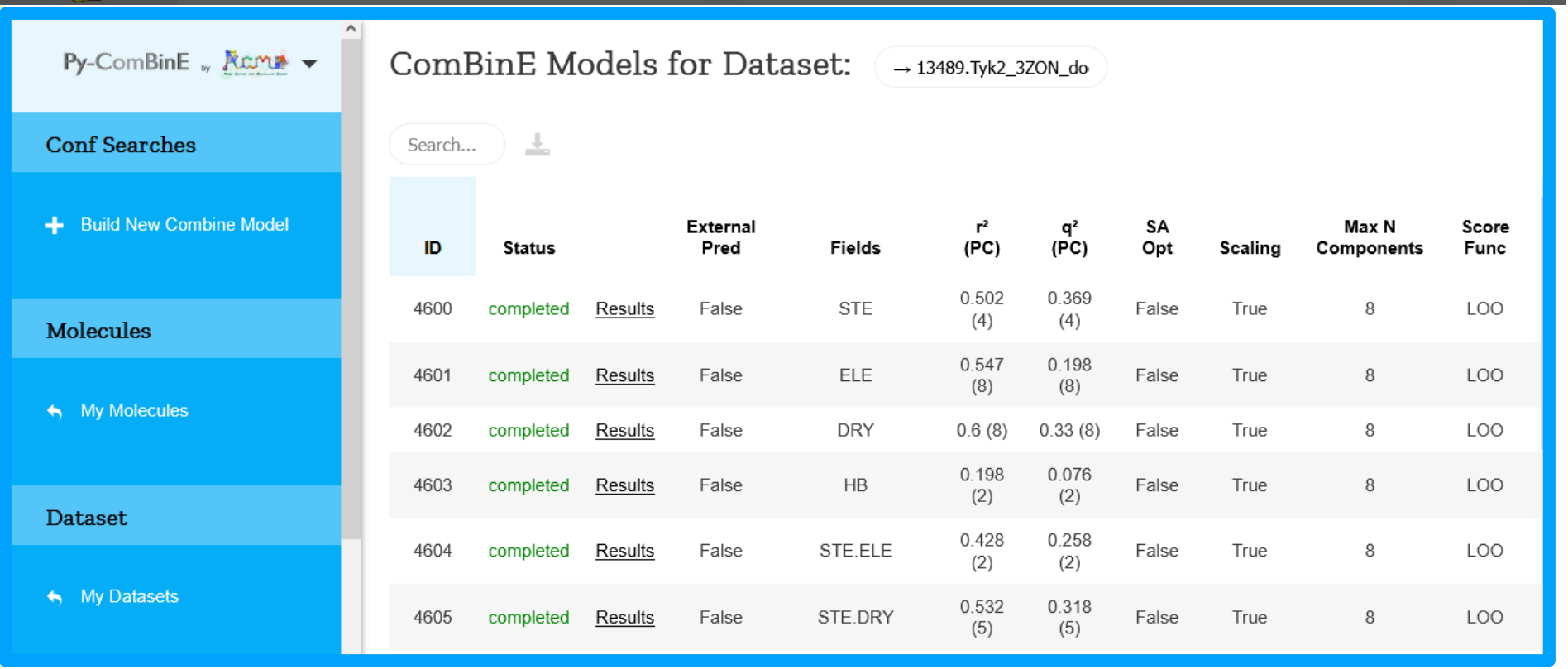
Py-CoMFA → 14 modelli con i parametri di default → 100 VPO per ciascun modello

| Modello | $r^2$ | $q^2$ | PC |
|---|---|---|---|
| Tyk2_3ZON | 0,850 | 0,500 | 4 |
| Tyk2_4OLI | 0,738 | 0,422 | 3 |
| Tyk2_4WOV | 0,681 | 0,441 | 3 |
| Tyk2_5C01 | 0,768 | 0,497 | 3 |
| Tyk2_5C03 | 0,685 | 0,382 | 3 |
| Tyk2_5TKD | 0,703 | 0,504 | 3 |
| Tyk2_6NSL | 0,747 | 0,521 | 3 |
| Tyk2_6NZE | 0,711 | 0,445 | 3 |
| Tyk2_6NZF | 0,761 | 0,443 | 3 |
| Tyk2_6NZH | 0,680 | 0,472 | 3 |
| Tyk2_6NZP | 0,620 | 0,467 | 2 |
| Tyk2_6NZQ | 0,667 | 0,512 | 3 |
| Tyk2_6NZR | 0,795 | 0,512 | 4 |
| Tyk2_7AX4 | 0,783 | 0,533 | 4 |





# Py-ComBine



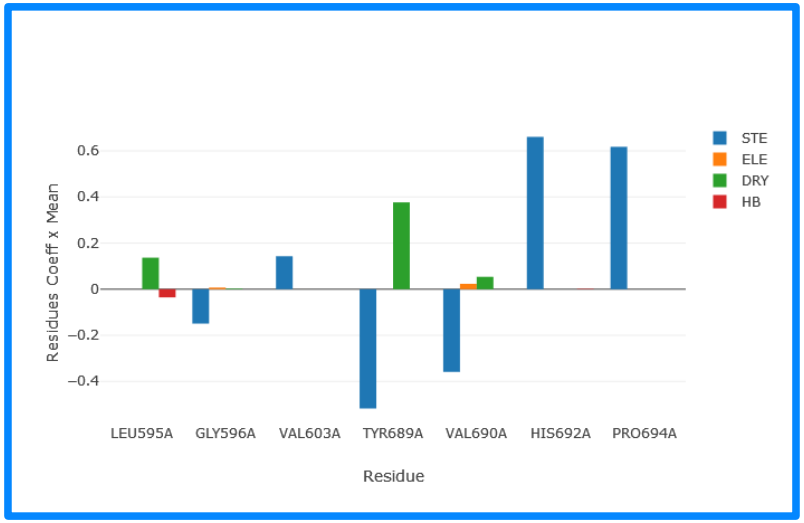

| ID | Status | | External Pred | Fields | $r^2$ (PC) | $q^2$ (PC) | SA Opt | Scaling | Max N Components | Score Func |
|---|---|---|---|---|---|---|---|---|---|---|
| 4600 | completed | Results | False | STE | 0.502 (4) | 0.369 (4) | False | True | 8 | LOO |
| 4601 | completed | Results | False | ELE | 0.547 (8) | 0.198 (8) | False | True | 8 | LOO |
| 4602 | completed | Results | False | DRY | 0.6 (8) | 0.33 (8) | False | True | 8 | LOO |
| 4603 | completed | Results | False | HB | 0.198 (2) | 0.076 (2) | False | True | 8 | LOO |
| 4604 | completed | Results | False | STE.ELE | 0.428 (2) | 0.258 (2) | False | True | 8 | LOO |
| 4605 | completed | Results | False | STE.DRY | 0.532 (5) | 0.318 (5) | False | True | 8 | LOO |

**15 modelli ComBine ottenuti dalle combinazioni di interazione ligando-proteina calcolate:**
**STE**
**ELE**
**DRY**
**HB**

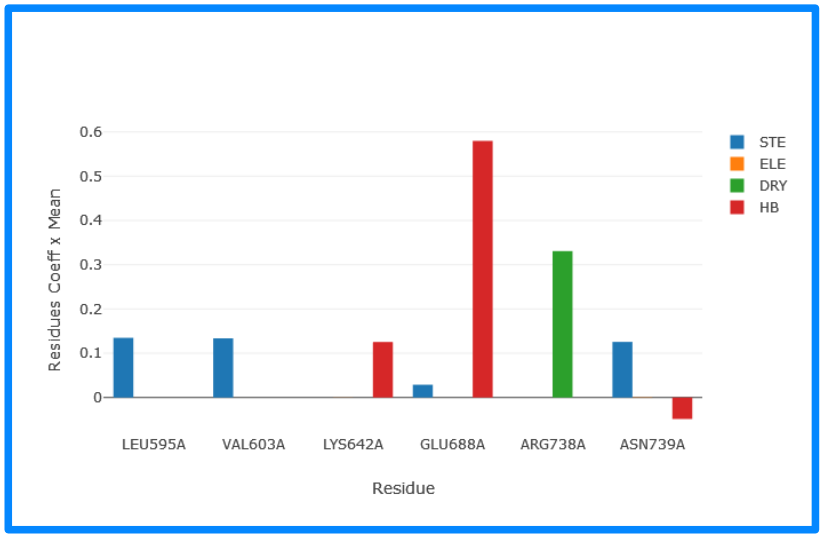
# Risultati Modelli ComBine

| Modello | $r^2$ | $q^2$ | PC |
|---|---|---|---|
| Tyk2_3ZON | 0,439 | 0,304 | 2 |
| Tyk2_4OLI | 0,508 | 0,338 | 4 |
| Tyk2_4WOV | 0,630 | 0,44 | 5 |
| Tyk2_5C01 | 0,613 | 0,429 | 5 |
| Tyk2_5C03 | 0,485 | 0,339 | 3 |
| Tyk2_5TKD | 0,672 | 0,436 | 5 |
| Tyk2_6NSL | 0,627 | 0,391 | 5 |
| Tyk2_6NZE | 0,603 | 0,507 | 3 |
| Tyk2_6NZF | 0,577 | 0,475 | 3 |
| Tyk2_6NZH | 0,657 | 0,567 | 5 |
| Tyk2_6NZP | 0,625 | 0,536 | 3 |
| Tyk2_6NZQ | 0,694 | 0,510 | 5 |
| Tyk2_6NZR | 0,650 | 0,540 | 4 |
| Tyk2_7AX4 | 0,586 | 0,410 | 4 |

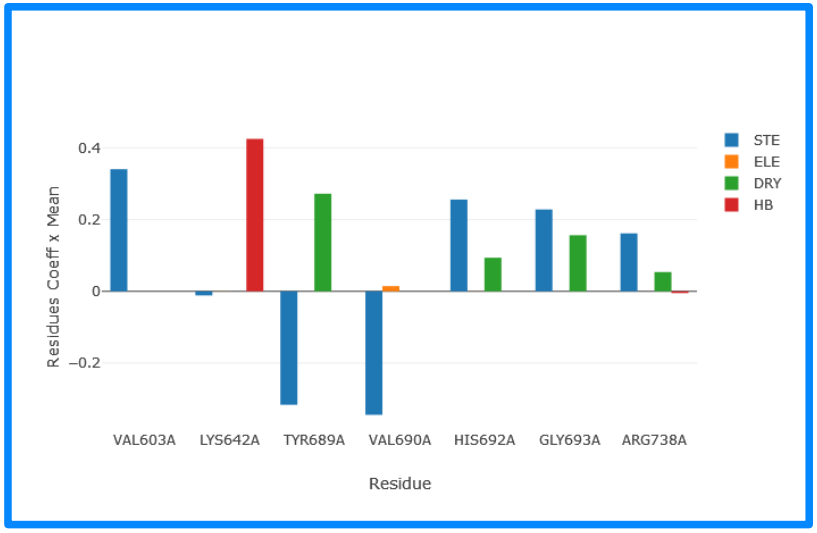
6NZH

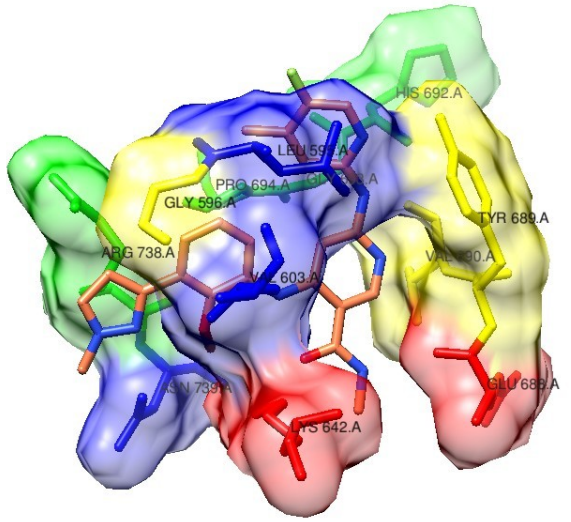



6NZP



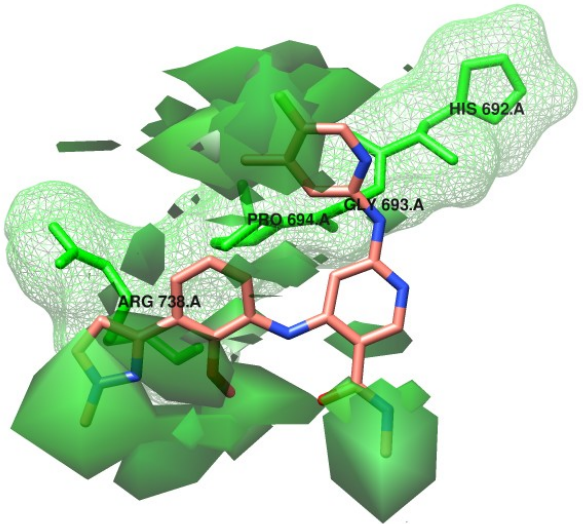

6NZR

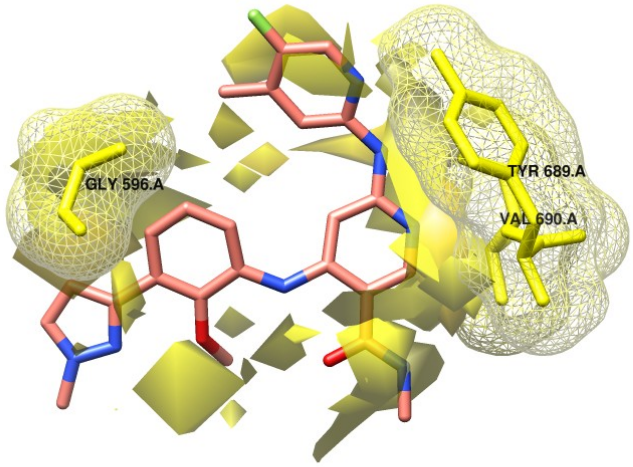


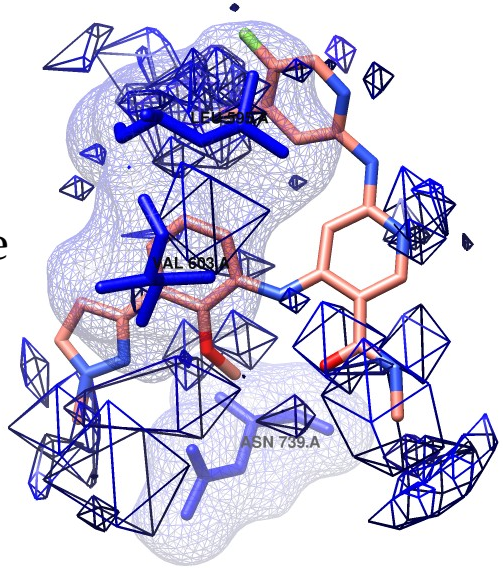
# COMFA e COMBINE: Consensus

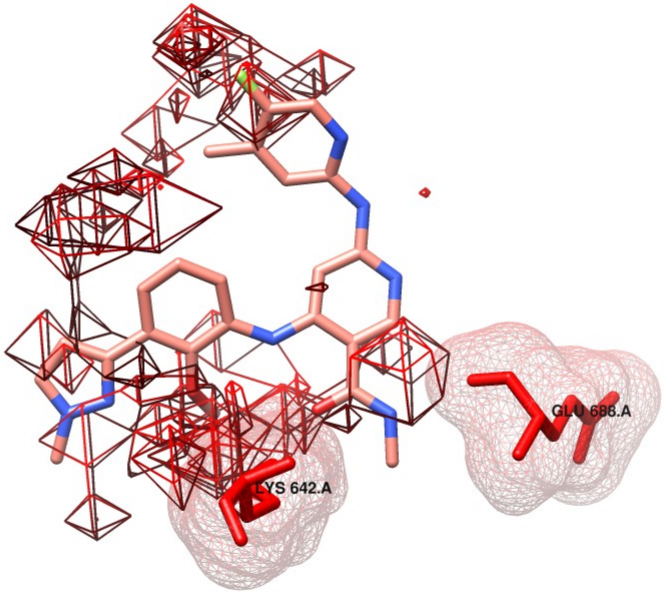
Interazioni steriche



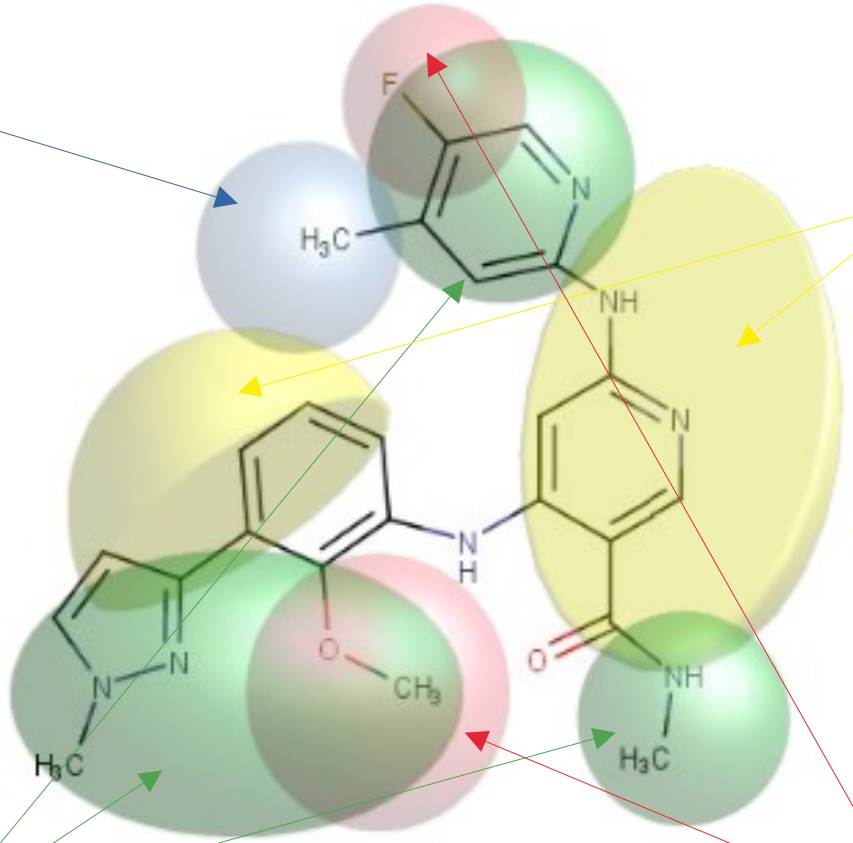

Interazioni elettrostatiche



Un aumento dell'ingombro elettrostatico modula positivamente l'attività biologica

Un ingombro sterico diminuirebbe notevolmente l'attività biologica




Un ingombro sterico aumenterebbe l'attività biologica

Un aumento dell'ingombro elettrostatico modula negativamente l'attività biologica

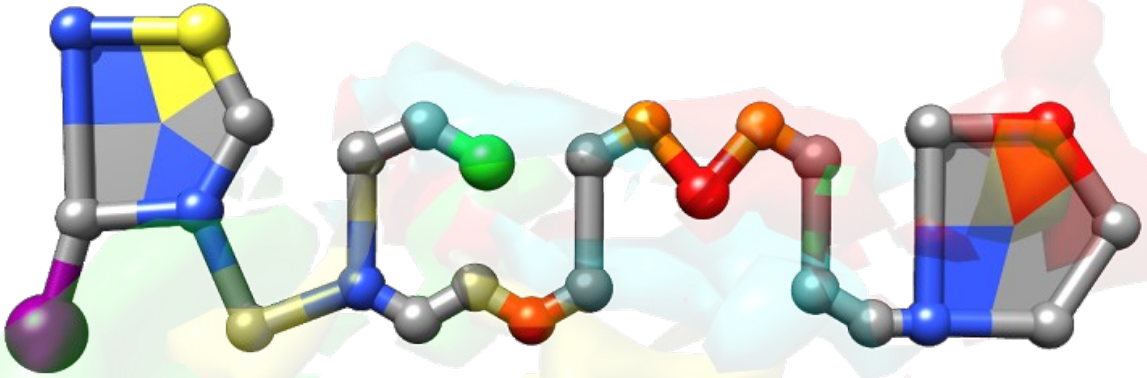
# Conclusioni

- Nonostante l'alta ed intrinseca casualità di calcolo, i modelli LB COMFA e il docking molecolare sono riproducibili;

- I modelli SB COMFA e COMBINE sono robusti, pertanto verranno utilizzati per effettuare un Virtual Screening;

- Esiste un consensus tra COMBINE e SB COMFA che ha permesso di derivare SAR tridimendionali utili per progettare nuovi inibitori allosterici Tyk2.

# Grazie a tutti per l'attenzione !




*Prof. Rino Ragno*
*Manuela Sabatino, PhD*
*Lorenzo Antonini, PhD*
*Filippo Sapienza*
*Eleonora Proia*

*Matteo Santucci*
*Sveva Nigro*