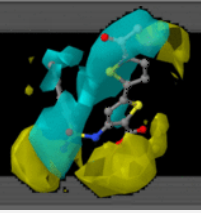
Sviluppo di un protocollo per la generazione automatica di modelli 3-D QSAR.

SAPIENZA
UNIVERSITÀ DI ROMA

Facoltà di Farmacia e Medicina
Corso di Laurea in Biotecnologie Farmaceutiche
Tesi Sperimentale in Chimica Farmaceutica
a.a. 2012/2013

Laureanda : Manuela Sabatino
Matricola: 1196251

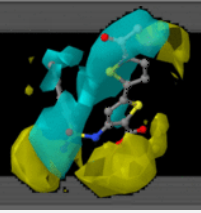
Relatore: prof. Rino Ragno



SCOPO DEL LAVORO

Focalizzazione sulle regole di allineamento

sviluppare un protocollo per la generazione automatica di modelli 3-D QSAR mediante un approccio *quasi*-sistematico



OBIETTIVI 3-D QSAR

(QSAR)

Quantitative **S**tructure-**A**ctivity **R**elationships

- Spiegazione dei dati Biologici su base tridimensionale
- Ottimizzazione molecole esistenti
- Predizione composti non ancora saggiati e/o sintetizzati

COSTRUZIONE DI UN MODELLO

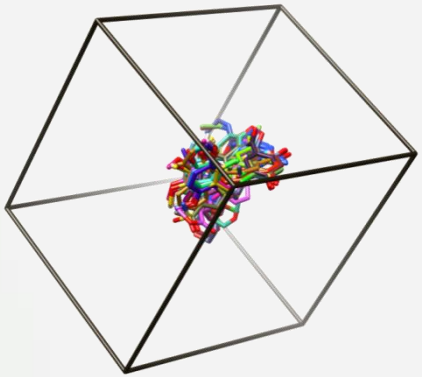
Scelta del training set

Allineamento delle molecole

Calcolo del campi di interazione molecolare (MIF)

Generazione del modello statistico

Validazione interna, esterna e y-scrambling



Analisi di regressione PLS (Partial Least Square)

- ✓Modello grezzo
- ✓Modello Pretrattato
- ✓Cross-Validazione (LOO, K-F5)
- ✓ Generazione contour maps CoMFA-like

OBIETTIVI 3-D QSAR

Scelta del training set

Allineamento delle molecole

Calcolo del campi di interazione molecolare (MIF)

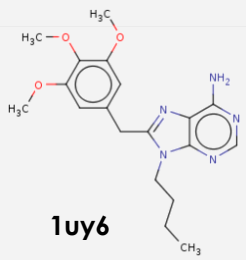
Generazione del modello statistico

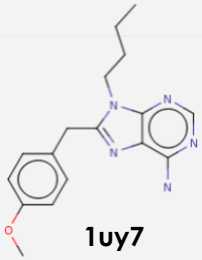
Validazione interna, esterna e y-scrambling

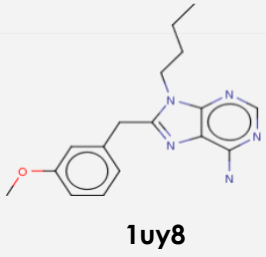
Ligand-based

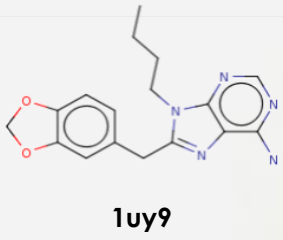
Structure-based

DATASET: HSP90

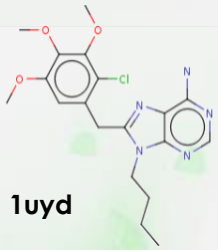

1uy6

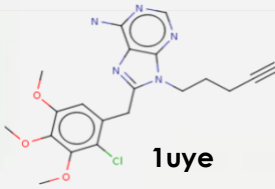

1uy7

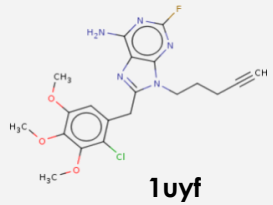

1uy8

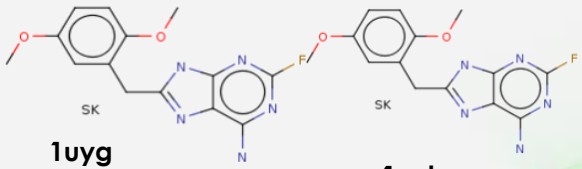

1uy9


1uyc

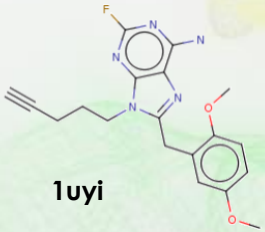

1uyd

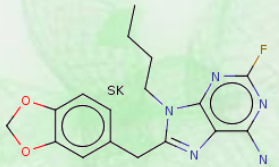

1uye


1uyf

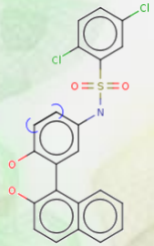


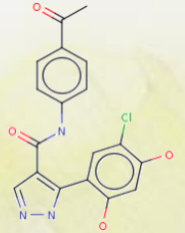
1uyg 1uyh

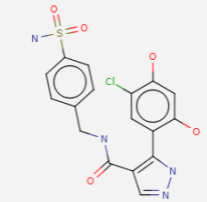

1uyi

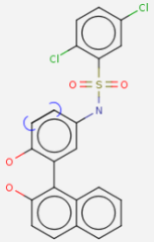


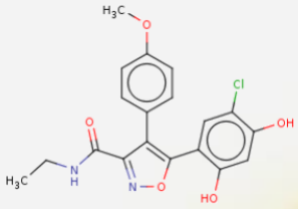
1uyk

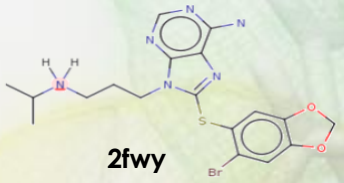

2bz5

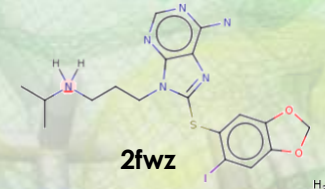

2byh

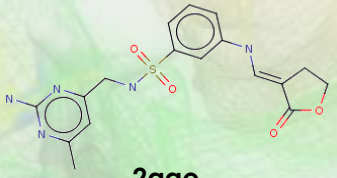

2byi

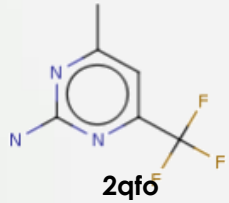

2bz5

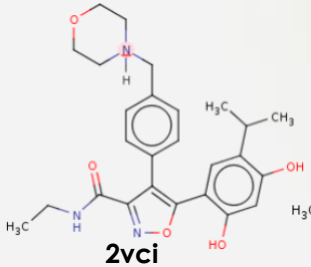

2uwz

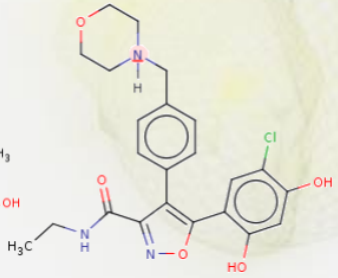

2fwy

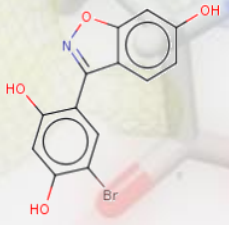

2fwz

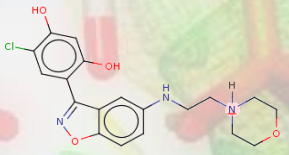

2qgo

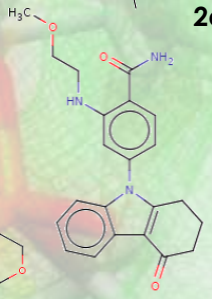

2qfo

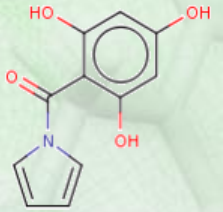

2vci


2vcj

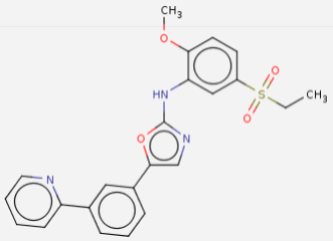

3bm9


3bmy

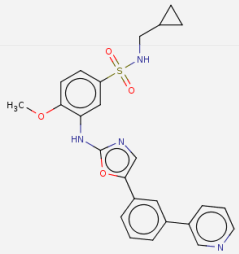

3d0b


3eko

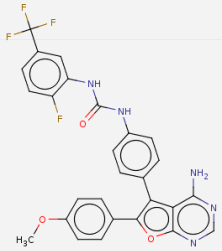
DATASET: VEGFR2



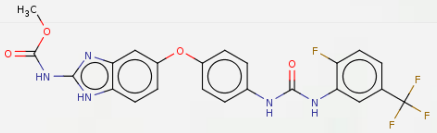
1y6a



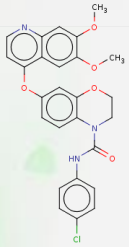
1y6b



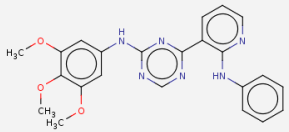
1ywn



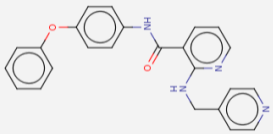
2 OH4



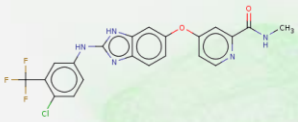
2rl5



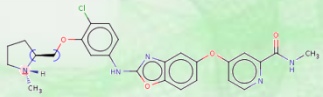
2p2h



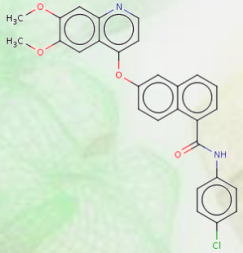
2p2i



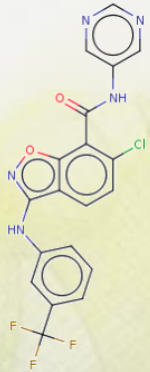
2qu5



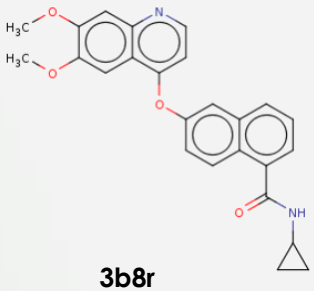
2qu6



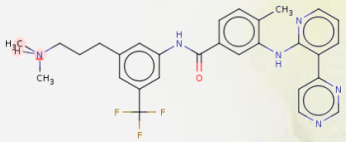
3b8q



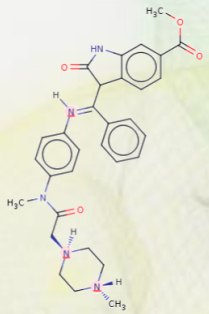
3dtw



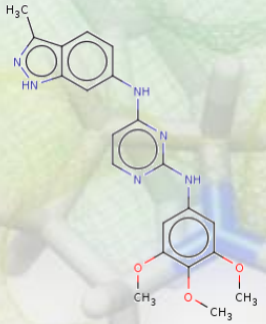
3b8r



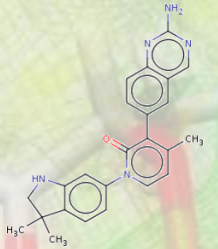
3be2



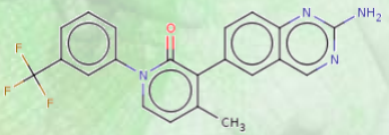
3c7q



3cjf

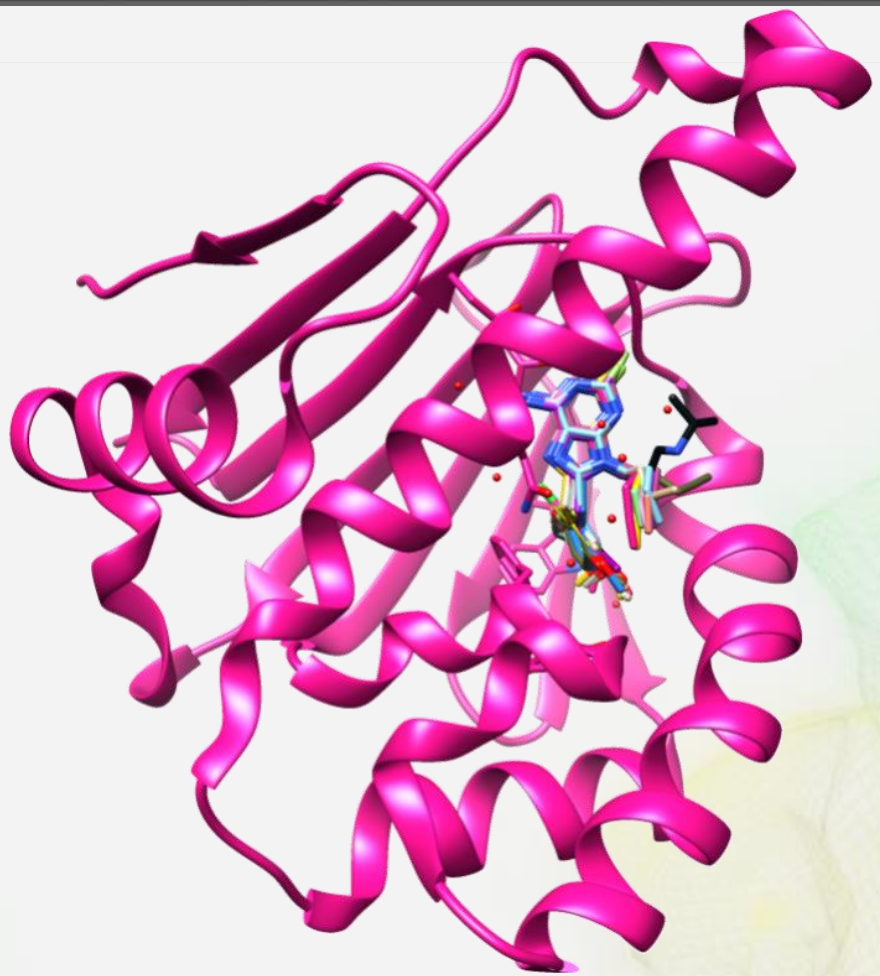


3cpb

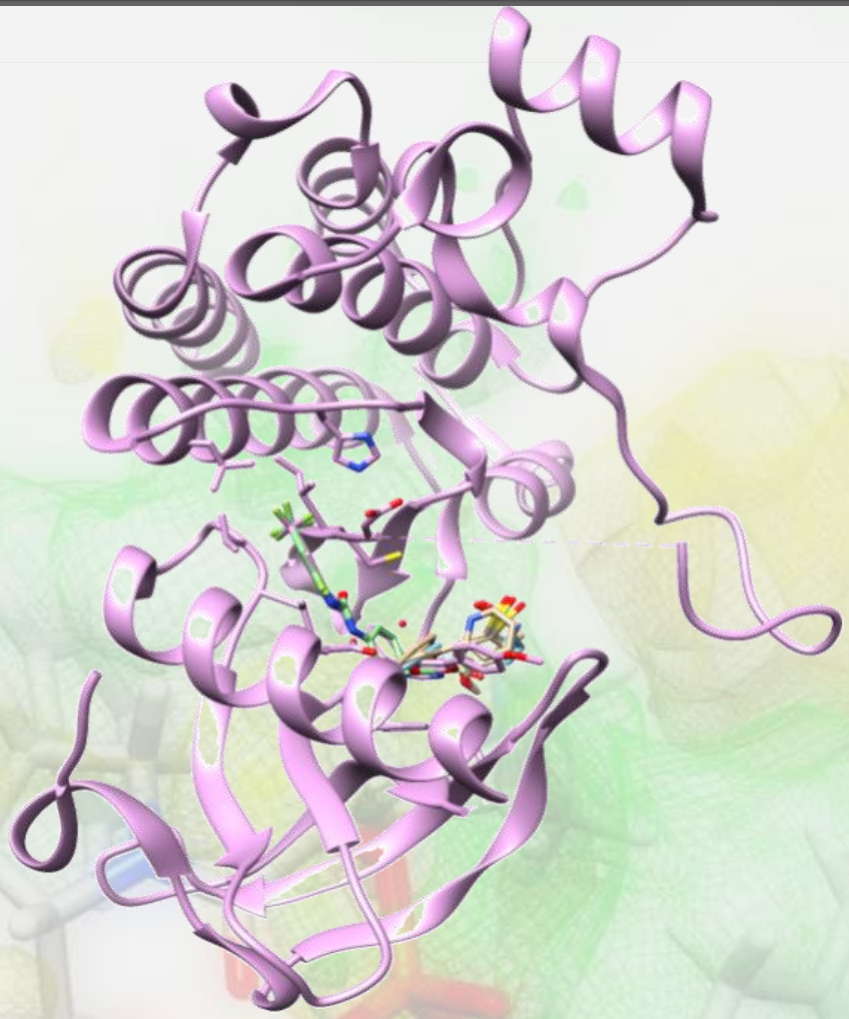


3cpc

COMPLESSI CRISTALLIZZATI

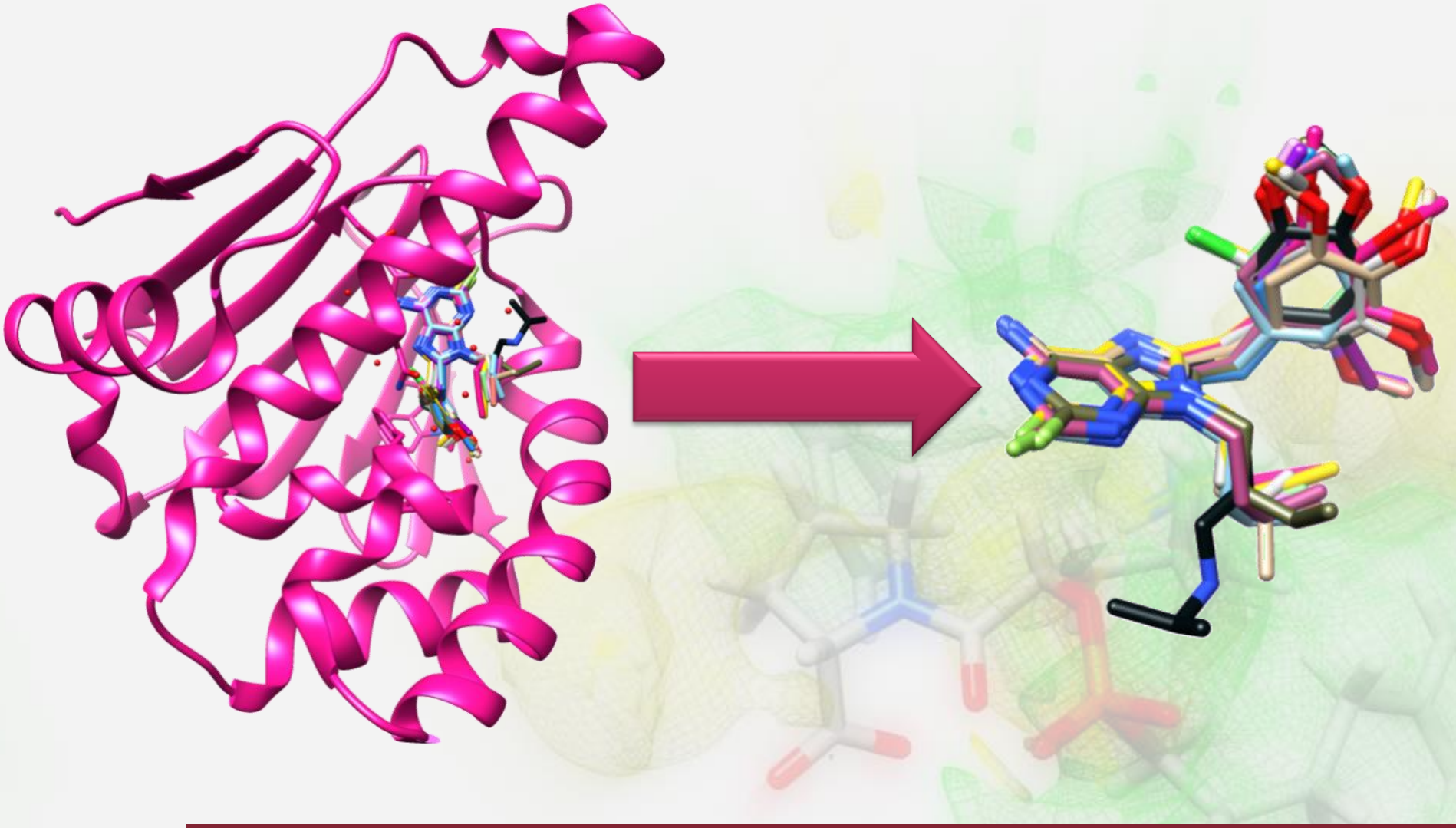


HSP90



VEGFR2

ALLINEAMENTO STRUCTURE-BASED



COSTRUZIONE DI UN MODELLO DI RIFERIMENTO

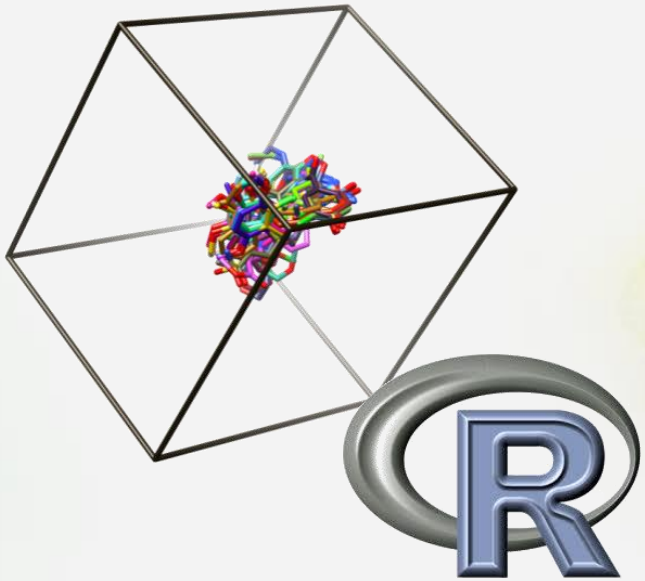
Scelta del training set

Allineamento delle molecole

Calcolo del campi di interazione molecolare (MIF)

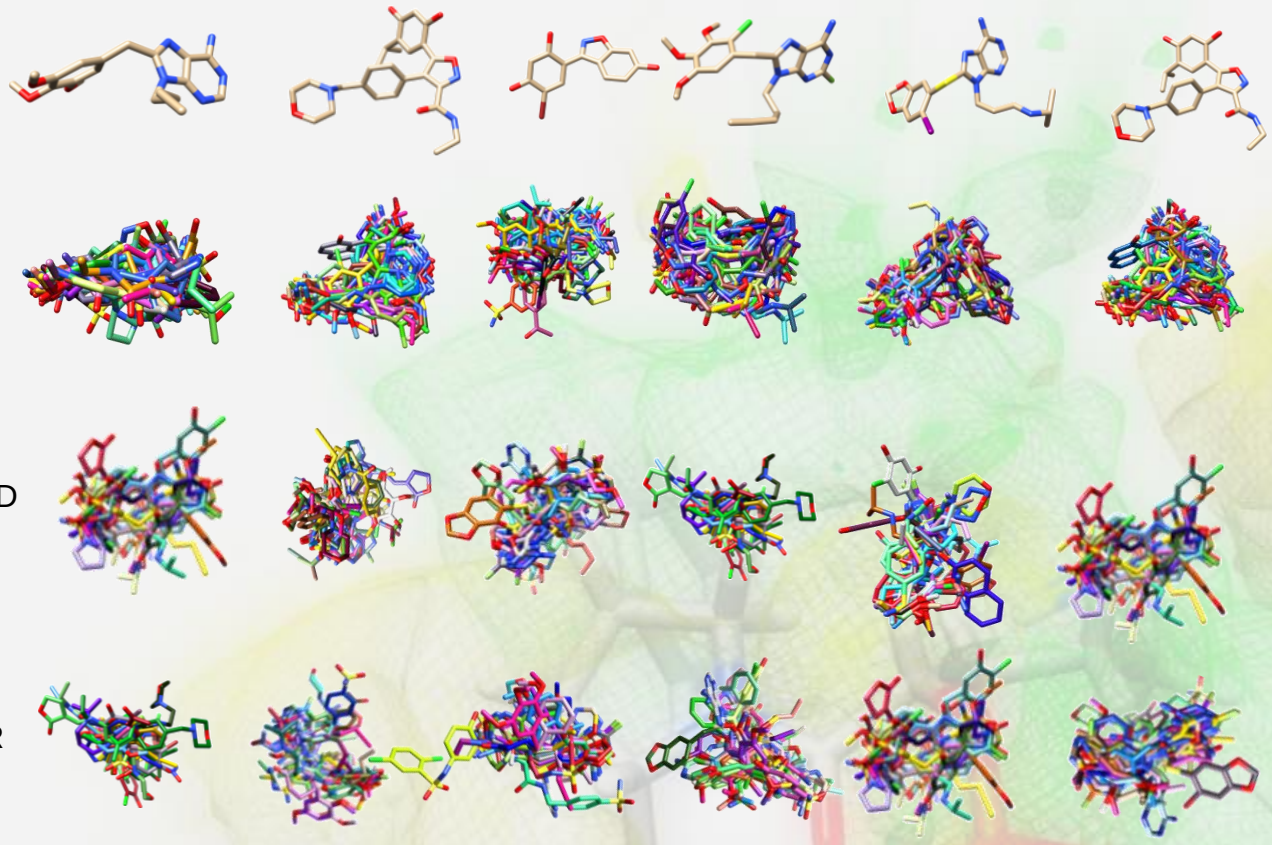
Generazione del modello statistico

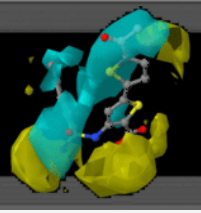
Validazione interna, esterna e y-scrambling



PROBE	PC0	PC1	PC2	PC3	PC4	PC5
A_DAT	-0,078	0,426	**0,552**	0,572	0,519	0,484
C_DAT	-0,078	0,427	**0,544**	0,569	0,519	0,486
OA_DAT	-0,078	0,408	**0,533**	0,583	0,547	0,533
N_DAT	-0,078	0,409	**0,516**	0,551	0,505	0,481
NA_DAT	-0,078	0,410	**0,540**	0,577	0,535	0,516
HD_DAT	-0,078	0,402	**0,550**	0,596	0,557	0,546
e_DAT	-0,078	0,740	**0,821**	0,832	0,821	0,818
d_DAT	-0,078	0,416	**0,384**	0,324	0,125	0,253

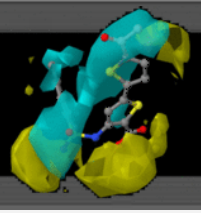
GLI ALLINEAMENTI HSP90





UN PO' DI RISULTATI...(HSP90)

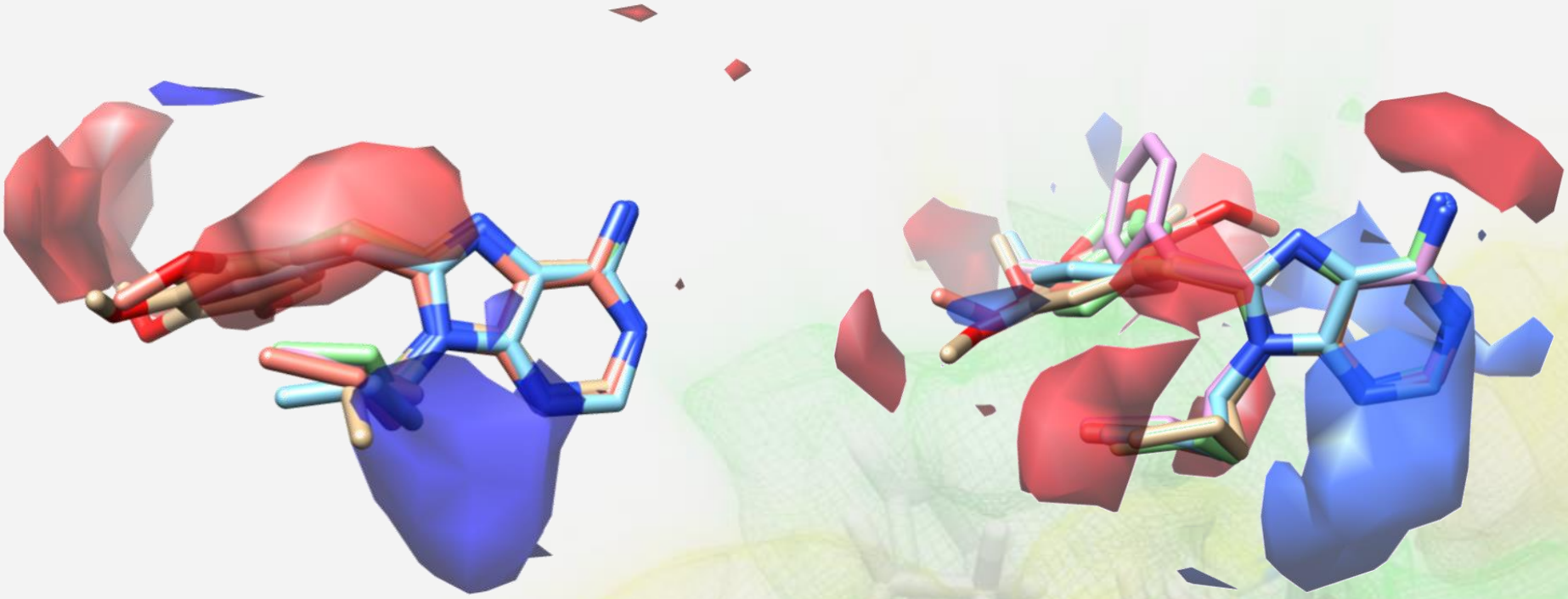
Probe	Struttura di Riferimento	Rriferimento	Surflex-sim	open3Dalign/GBSA			open3Dalign/VACUUM		
				Atom	Pharm	Mixed	Atom	Pharm	Mixed
A_DAT	1uy6_meno_attiva	0,552	0,412	0,202	0,583	0,049	0,386	0,543	0,433
C_DAT	1uy6_meno_attiva	0,544	0,405	1856	0,574	0,056	0,376	0,539	0,430
OA_DAT	1uy6_meno_attiva	0,533	0,404	0,21	0,540	0,020	0,360	0,515	0,397
N_DAT	1uy6_meno_attiva	0,516	0,407	0,197	0,563	0,028	0,373	0,526	0,424
NA_DAT	1uy6_meno_attiva	0,54	0,41	0,198	0,539	0,040	0,357	0,532	0,417
HD_DAT	1uy6_meno_attiva	0,55	0,451	0,312	0,542	0,117	0,314	0,465	0,395
e_DAT	1uy6_meno_attiva	0,821	0,764	0,57	0,609	0,551	0,592	0,545	0,677
d_DAT	1uy6_meno_attiva	0,384	0,55	-0,243	0,149	-0,738	-0,333	-0,115	-0,189



UN PO' DI RISULTATI...(HSP90)

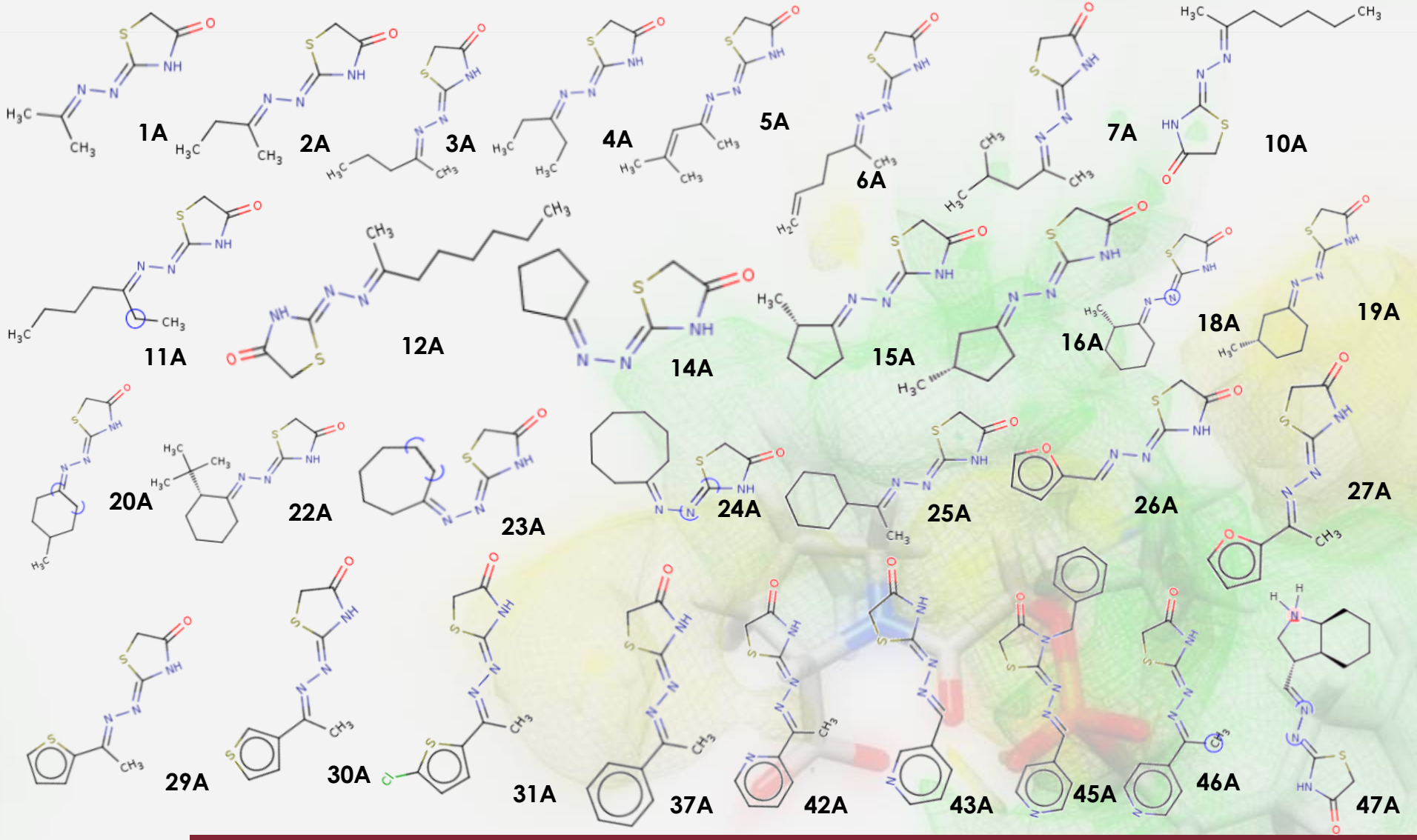
Probe	Struttura di Riferimento	Rriferimento	Surflex-sim	open3Dalign/GBSA			open3Dalign/VACUUM		
				Atom	Pharm	Mixed	Atom	Pharm	Mixed
A_DAT	1uy7_piu_flessibile	0,552	0,486	0,414	0,152	0,423	0,413	0,4372	0,395
C_DAT	1uy7_piu_flessibile	0,544	0,485	0,41	0,157	0,424	0,421	0,4243	0,402
OA_DAT	1uy7_piu_flessibile	0,533	0,492	0,353	0,138	0,371	0,408	0,4240	0,407
N_DAT	1uy7_piu_flessibile	0,516	0,508	0,394	0,120	0,403	0,424	0,4301	0,413
NA_DAT	1uy7_piu_flessibile	0,54	0,507	0,37	0,164	0,393	0,405	0,4477	0,402
HD_DAT	1uy7_piu_flessibile	0,55	0,522	0,373	0,147	0,378	0,386	0,4111	0,368
e_DAT	1uy7_piu_flessibile	0,821	0,592	0,602	0,473	0,663	0,599	0,5538	0,684
d_DAT	1uy7_piu_flessibile	0,384	-0,853	-0,467	-0,291	-0,383	-0,279	0,0132	-0,345

CONFRONTO CONTOUR MAPS

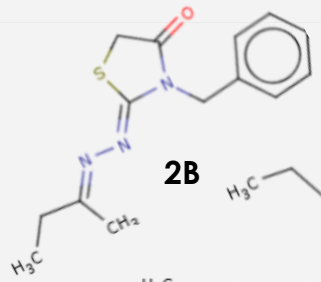


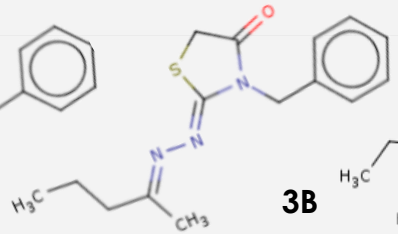
PLS Coefficient Plot - Probe N: a sinistra il plot ottenuto dall'allineamento sperimentale, a destra dall'allineamento con surflex

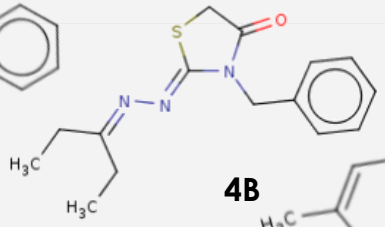
NUOVO DATASET: ANTIFUNGINI

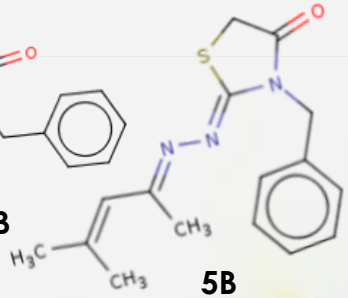


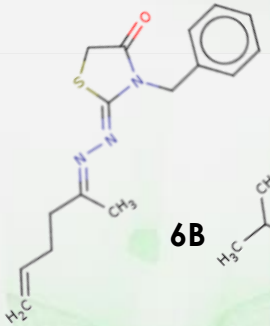
NUOVO DATASET: ANTIFUNGINI

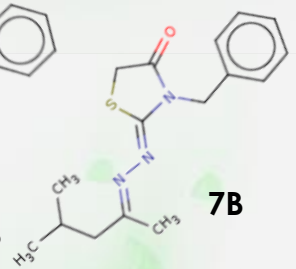


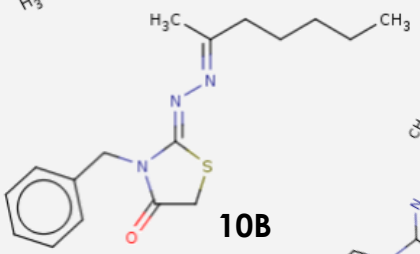


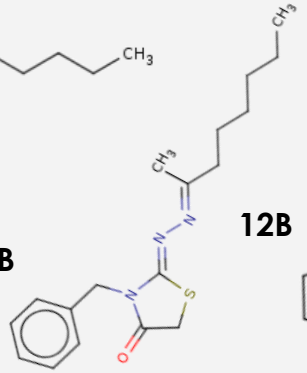


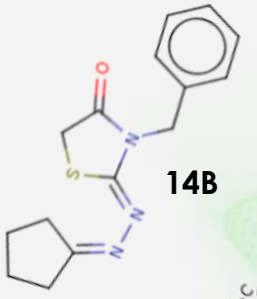


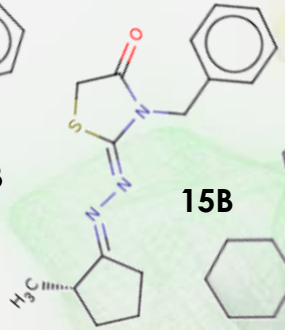


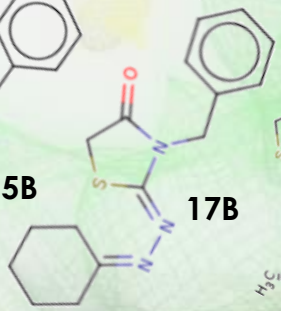


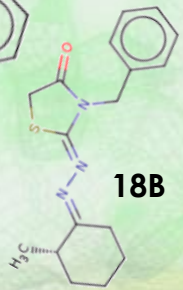


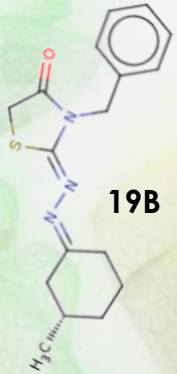


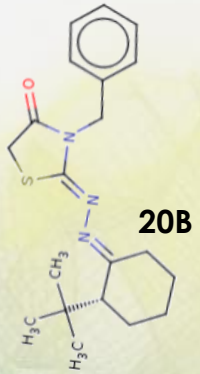


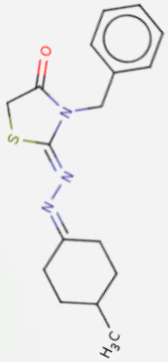


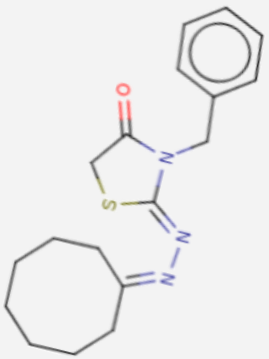


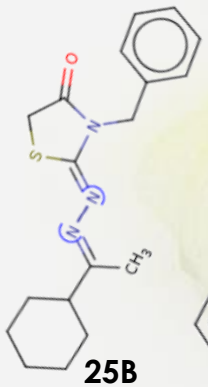


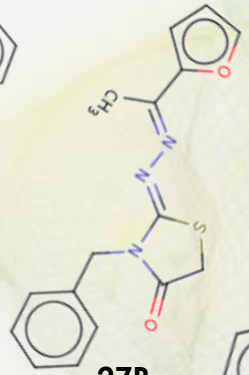


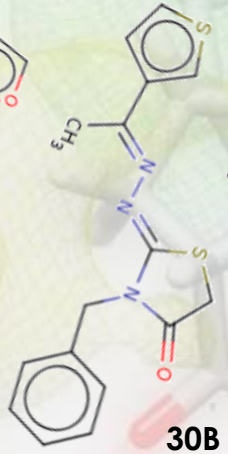


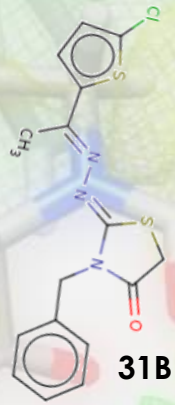


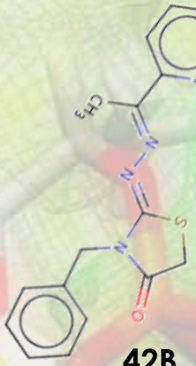


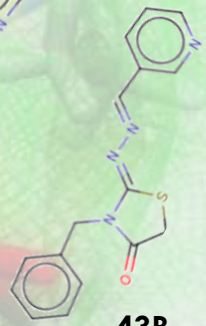


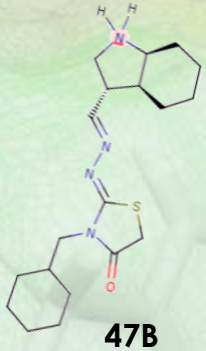




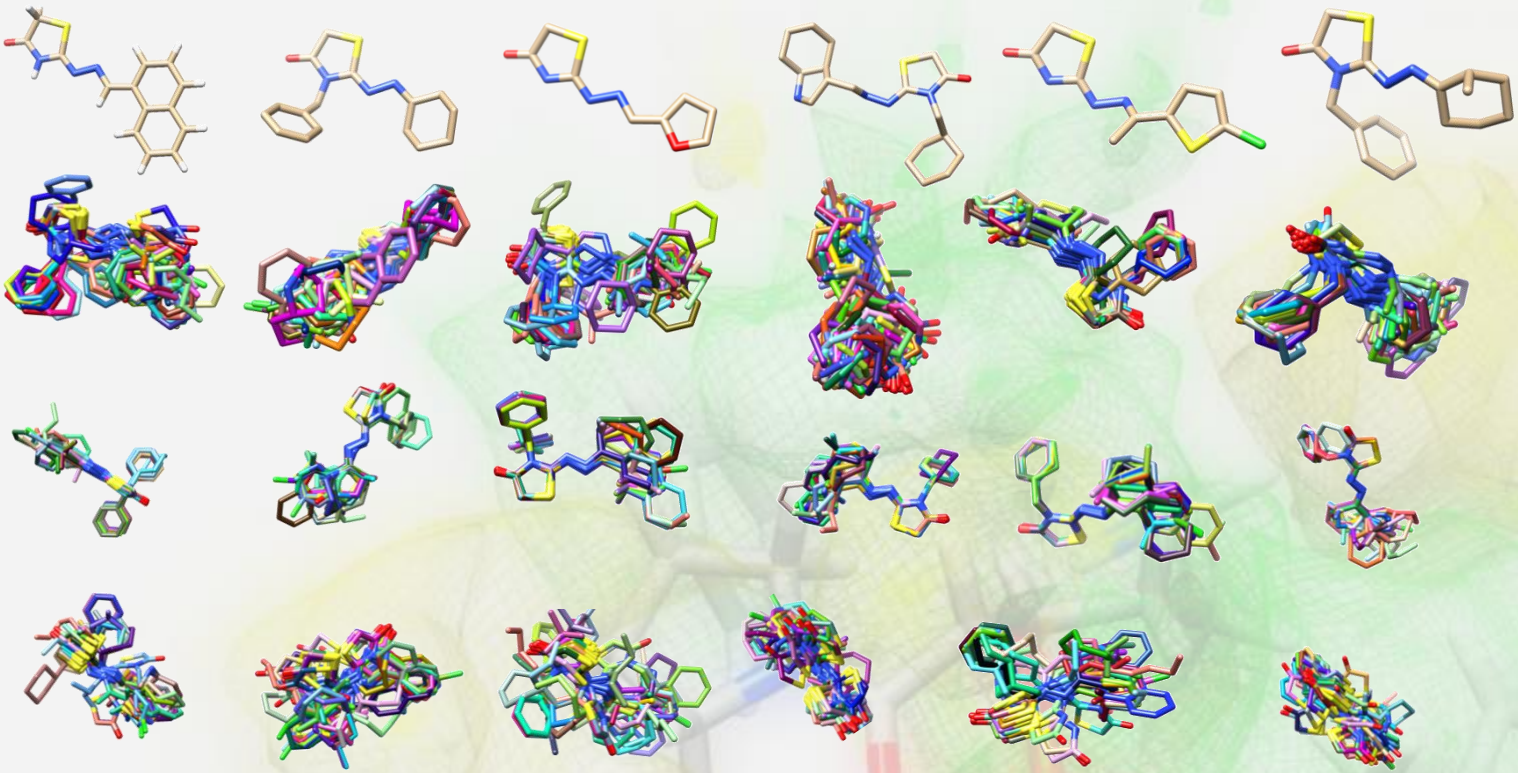


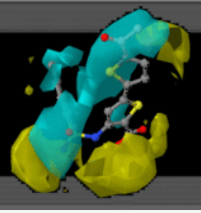






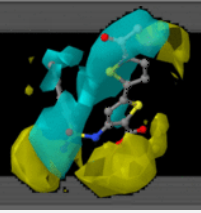
ALLINEAMENTI ANTIFUNIGNI





RISULTATI…(ANTIFUNGO)

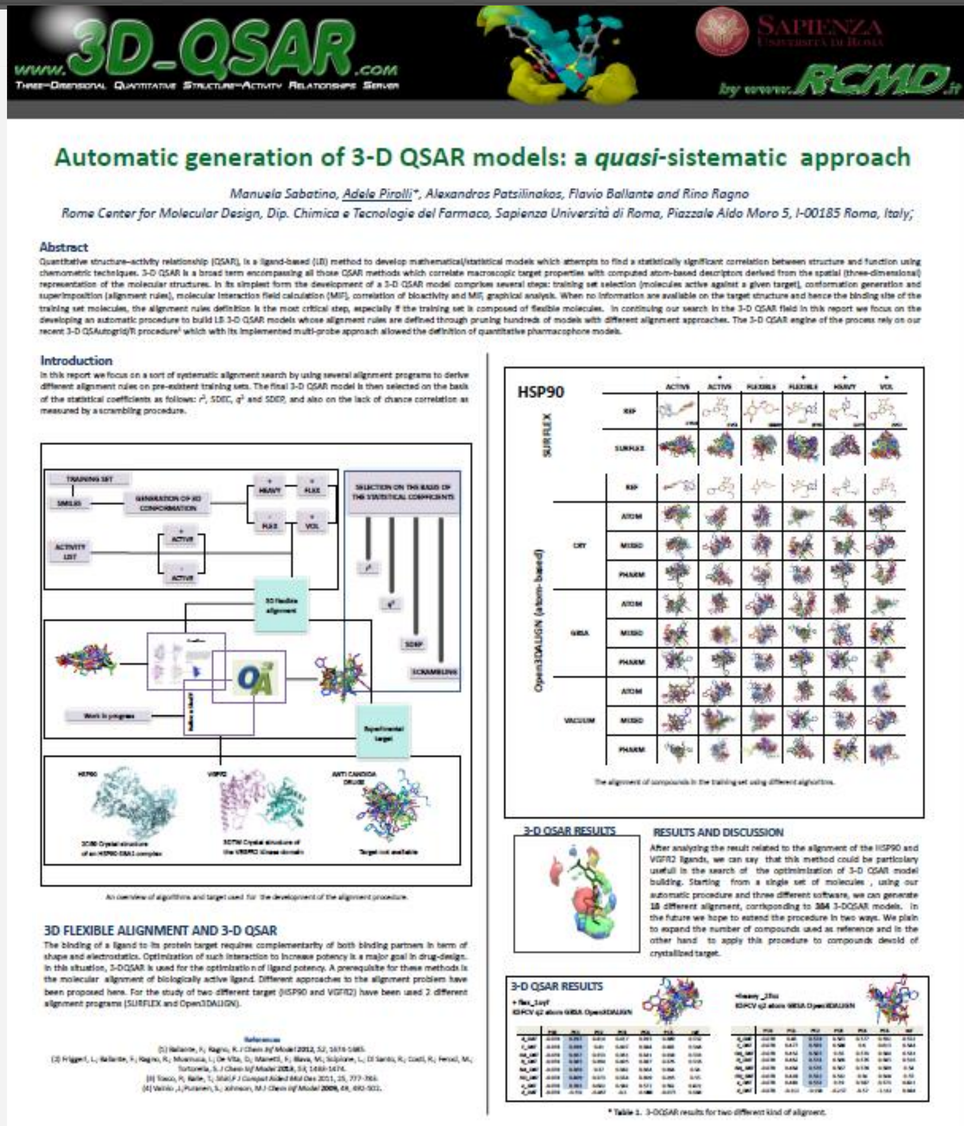
Probe	Struttura di Riferimento	Riferimento	Surflex-sim	open3Dalign/GBSA			open3Dalign/VACUUM		
				Atom	Pharm	Mixed	Atom	Pharm	Mixed
A_DAT	17B_più_attiva	/	0,231	0,313	0,358	0,298	0,354	0,274	0,351
C_DAT	17B_più_attiva	/	0,207	0,324	0,345	0,297	0,349	0,263	0,351
OA_DAT	17B_più_attiva	/	0,204	0,322	0,371	0,296	0,346	0,290	0,344
N_DAT	17B_più_attiva	/	0,213	0,325	0,374	0,296	0,348	0,279	0,349
NA_DAT	17B_più_attiva	/	0,197	0,318	0,372	0,288	0,349	0,292	0,346
HD_DAT	17B_più_attiva	/	0,223	0,327	0,371	0,296	0,337	0,291	0,333
e_DAT	17B_più_attiva	/	0,024	0,168	-0,175	0,031	0,068	-0,395	0,081
d_DAT	17B_più_attiva	/	0,058	0,222	0,164	0,201	0,278	0,181	0,276



RISULTATI...(ANTIFUNGO)

Probe	Struttura di Riferimento	Riferimento	Surflex-sim	open3Dalign/GBSA			open3Dalign/VACUUM		
				Atom	Pharm	Mixed	Atom	Pharm	Mixed
A_DAT	26A_meno_flessibile	/	-0,11	0,277	0,395	0,280	0,203	0,389	0,200
C_DAT	26A_meno_flessibile	/	-0,12	0,283	0,403	0,269	0,209	0,374	0,203
OA_DAT	26A_meno_flessibile	/	-0,114	0,287	0,385	0,281	0,239	0,391	0,234
N_DAT	26A_meno_flessibile	/	-0,105	0,282	0,394	0,272	0,229	0,396	0,216
NA_DAT	26A_meno_flessibile	/	-0,114	0,258	0,391	0,271	0,219	0,406	0,221
HD_DAT	26A_meno_flessibile	/	0,163	0,276	0,387	0,272	0,231	0,383	0,233
e_DAT	26A_meno_flessibile	/	0,073*	-0,062	0,024	-0,051	0,074	-0,012	0,079
d_DAT	26A_meno_flessibile	/	0,035*	0,069	0,286	0,083	0,036	0,349	0,040

UN PICCOLO TRAGUARDO



www.3D_QSAR.com
Three-Dimensional Quantitative Structure-Activity Relationships Server
Sapienza Università di Roma
by www.RCMD.it

Automatic generation of 3-D QSAR models: a *quasi*-sistematic approach

Manuela Sabatino, Adele Pirolli*, Alexandros Patsilinakos, Flavio Ballante and Rino Ragno
Rome Center for Molecular Design, Dip. Chimica e Tecnologie del Farmaco, Sapienza Università di Roma, Piazzale Aldo Moro 5, I-00185 Roma, Italy;

Abstract
Quantitative structure–activity relationship (QSAR), is a ligand-based (LB) method to develop mathematical/statistical models which attempts to find a statistically significant correlation between structure and function using chemometric techniques. 3-D QSAR is a broad term encompassing all those QSAR methods which correlate macroscopic target properties with computed atom-based descriptors derived from the spatial (three-dimensional) representation of the molecular structures. In its simplest form the development of a 3-D QSAR model comprises several steps: training set selection (molecules active against a given target), conformation generation and superimposition (alignment rules), molecular interaction field calculation (MIF), correlation of bioactivity and MIF, graphical analysis. When no information are available on the target structure and hence the binding site of the training set molecules, the alignment rules definition is the most critical step, especially if the training set is composed of flexible molecules. In continuing our search in the 3-D QSAR field in this report we focus on the developing an automatic procedure to build LB 3-D QSAR models whose alignment rules are defined through pruning hundreds of models with different alignment approaches. The 3-D QSAR engine of the process rely on our recent 3-D QSAutogrid/R procedure[1] which with its implemented multi-probe approach allowed the definition of quantitative pharmacophore models.

Introduction
In this report we focus on a sort of systematic alignment search by using several alignment programs to derive different alignment rules on pre-existent training sets. The final 3-D QSAR model is then selected on the basis of the statistical coefficients as follows: r^2, SDEC, q^2 and SDEP, and also on the lack of chance correlation as measured by a scrambling procedure.

An overview of algorithms and target used for the development of the alignment procedure.

HSP90

The alignment of compounds in the training set using different alghoritms.

3D FLEXIBLE ALIGNMENT AND 3-D QSAR
The binding of a ligand to its protein target requires complementarity of both binding partners in term of shape and electrostatics. Optimization of such interaction to increase potency is a major goal in drug-design. In this situation, 3-DQSAR is used for the optimizatio n of ligand potency. A prerequisite for these methods is the molecular alignment of biologically active ligand. Different approaches to the alignment problem have been proposed here. For the study of two different target (HSP90 and VGFR2) have been used 2 different alignment programs (SURFLEX and Open3DALIGN).

3-D QSAR RESULTS

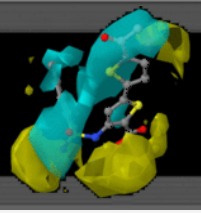
RESULTS AND DISCUSSION
After analyzing the result related to the alignment of the HSP90 and VGFR2 ligands, we can say that this method could be particolary usefull in the search of the optimimization of 3-D QSAR model building. Starting from a single set of molecules , using our automatic procedure and three different software, we can generate 18 different alignment, corrisponding to 384 3-DQSAR models. In the future we hope to extend the procedure in two ways. We plain to expand the number of compounds used as reference and in the other hand to apply this procedure to compounds devoid of crystallized target.

3-D QSAR RESULTS

* Table 1. 3-DQSAR results for two different kind of alignment.

References

Manuela Sabatino, Adele Pirolli, Alexandros Patsilinakos, Flavio Ballante, Rino Ragno, Automatic generation of 3-D QSAR models: a *quasi*- sistematic approach .



CONCLUSIONI

Sono stati creati **2016** modelli 3-D QSAR con il quale:

- Si è validata la robustezza del metodo 3-D QSAutogrid/R per la generazione di modelli predittivi.
- Creato un approccio automatico per la creazione di allineamenti e dei modelli.
- Allestimento di un database di modelli 3-D QSAR per predire l'attività biologica di molecole incognite attraverso un portale web

CONCLUSIONI

www.3d-qsar.com

www.3d-qsar.com/predictor/model/ache_a/

www.3D-QSAR.com
THREE-DIMENSIONAL QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIPS SERVER

Home | Launch Predictor | Datasets | About | Change Log

You are inspecting model AchE_A

Model Description

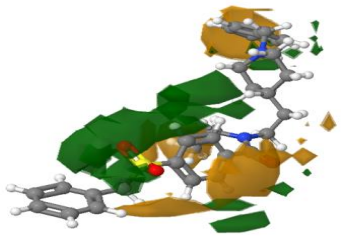
The AchE dataset was taken from J. Med. Chem. 2004, 47, 5541-5554. Paper abstract: A large number of methods are available for modeling quantitative structure-activity relationships (QSAR). We examine the predictive accuracy of several methods applied to data sets of inhibitors for angiotensin converting enzyme, acetylcholinesterase, benzodiazepine receptor, cyclooxygenase-2, dihydrofolate reductase, glycogen phosphorylase b, thermolysin, and thrombin. Descriptors calculated with CoMFA, CoMSIA, EVA, HQSAR, and traditional 2D and 2.5D descriptors were used for developing models with partial least squares (PLS). In addition, the genetic function approximation algorithm, genetic PLS, and back-propagation neural networks were used for deriving models from 2.5D descriptors (i.e., 2D descriptors and 3D descriptors calculated from CORINA structures and Gasteiger-Marsili charges). Predictive accuracy was assessed using designed test sets. It was found that HQSAR generally performs as well as CoMFA and CoMSIA; other descriptor sets performed less well. When 2.5D descriptors were used, only neural network ensembles were found to be similarly or more predictive than PLS models. In addition, we show that many cross-validation procedures yield similar estimates of the interpolative accuracy of methods. However, the lack of correspondence between cross-validated and test set predictive accuracy for four sets underscores the benefit of using designed test sets.

AchE Models

AchE_e
AchE_d
AchE_C
AchE_A
AchE_HD
AchE_N

Other Datasets

AC
ACE
AchE
AI
ARB
ATA



Training Set

Show	Mol id	Exp p[act]
c	94	9.520
c	93	9.220
c	141	8.920
c	128	8.700
c	140	8.550
c	77	8.380
c	81	8.270
c	117	8.240
c	118	8.190

Grazie per l'attenzione