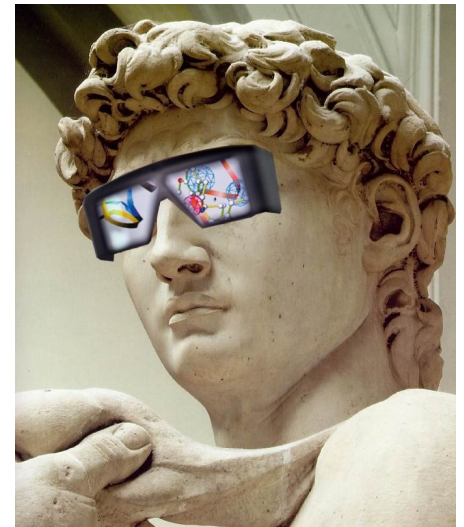




SAFAN-ISP: a new small molecule fragment informatics approach for in-silico profiling.

- ◆ **Drug Repositioning**
- ◆ **Side Effect prediction**

Luisa Pugliese - S.A.F.A.N. BIOINFORMATICS (Torino)
luisa.pugliese@safan-bioinformatics.it



S.A.F.A.N. BIOINFORMATICS is:

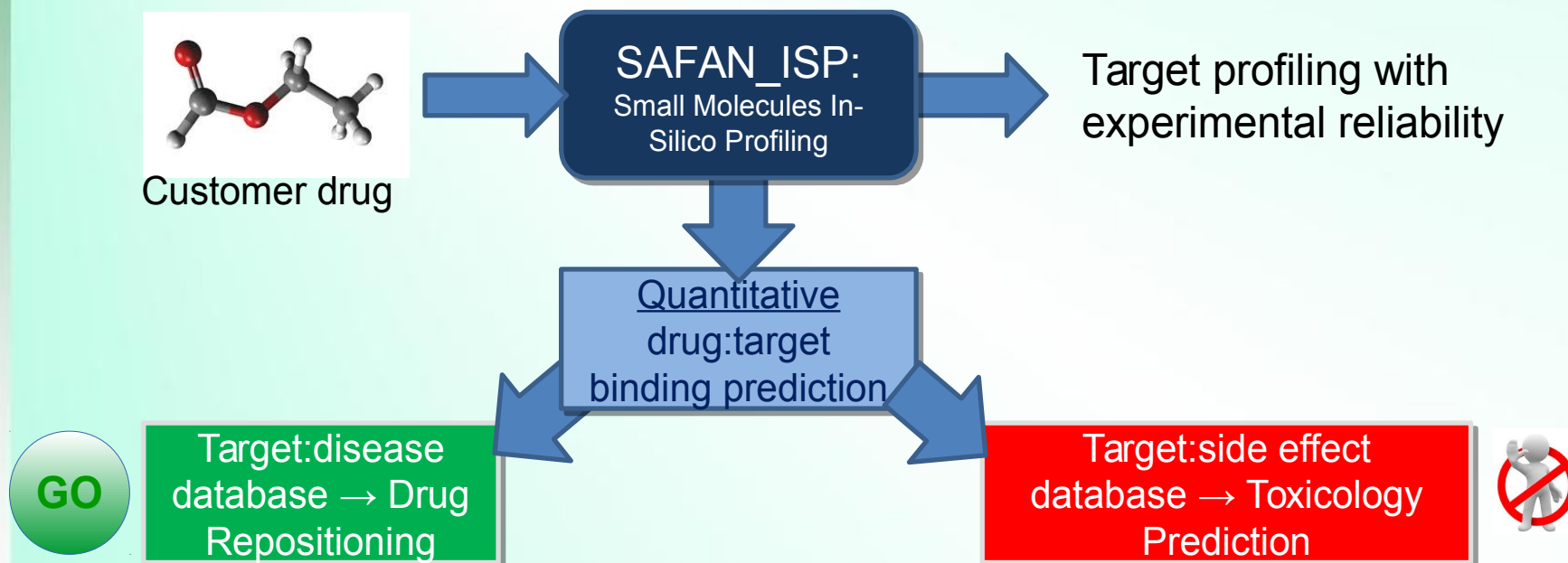
- A research bioinformatic company located in Turin,
- based on 18 years of experience in Italian and European universities,
- working for 11 years on applied research & development projects.
- Involved in the **NEWPROT** fp7 project than designs and implements a one-stop-shop portal for all in silico protein engineering work.
- Involved in the **CaSR Biomedicine** H2020 Marie Curie project.



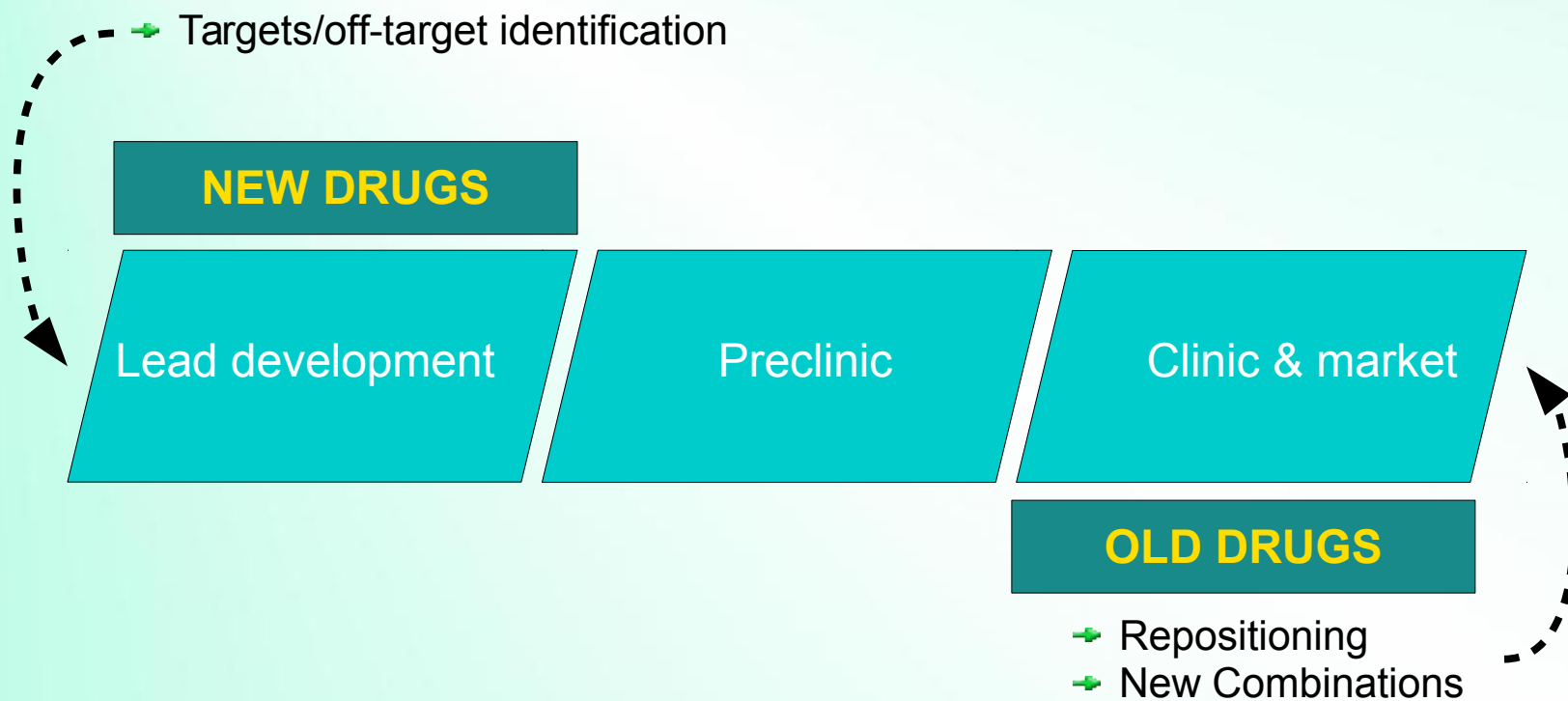
New platform for in silico profiling and drug repositioning: SAFAN-ISP

What we do:

We find the right target for each molecule, offering experimental precision with in-silico cost.



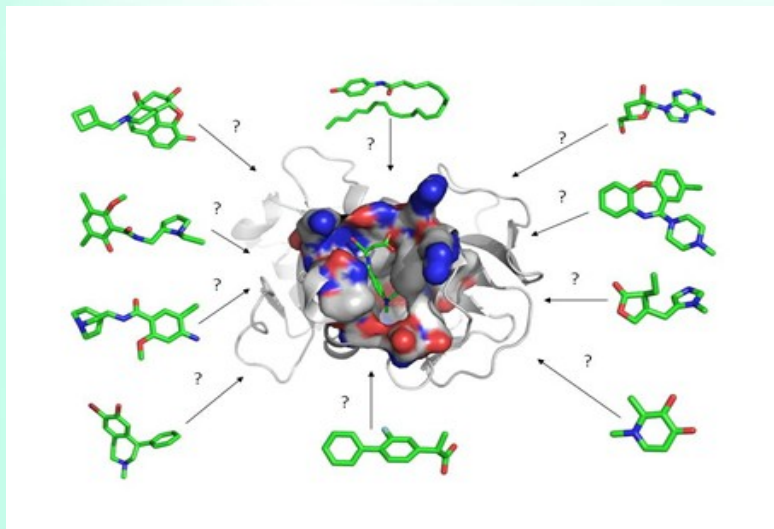
In silico Drug Profiling Applications



Technologies for in silico drug repositioning

COMPANY	TECHNOLOGY		
	DATABASE ANALYSIS: IMPORTANT INFORMATION NO NEW PATENTS	COMPUTATIONAL CHEMISTRY PREDICTIONS: NEW IP & NEW PATENTS	QUANTITATIVE SCORING: NEW IP, NEW PATENTS & THERAPEUTIC WINDOW
S.A.F.AN. BIOINFORMATICS	√	√	√
CHEMOTARGETS	√	√	√
GVK BIOSCIENCE	√		
THOMPSON REUTER	√	√	
NUMEDICUS Ldt	√		
EUROFINS BIOPRINT	√		
CRESSET		√	
SOM BIOTECH		√	

Virtual Screening or Profiling?

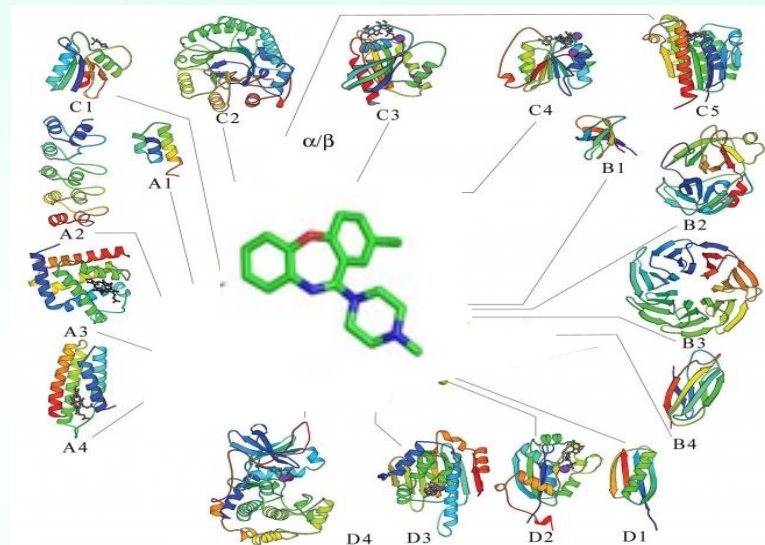


Virtual screening

One target



Thousands of small
molecules



SAFAN-ISP: Profiling

One small molecules



Thousands of targets

Small molecule profiling strategies

- **Ligand-based**

- 2D similarity

- SAFAN_ISP**

- 3D similarity

Database drug:target
Information. *The quality of the
database highly influences results.*

- **TARGET-based**

- Binding site similarity

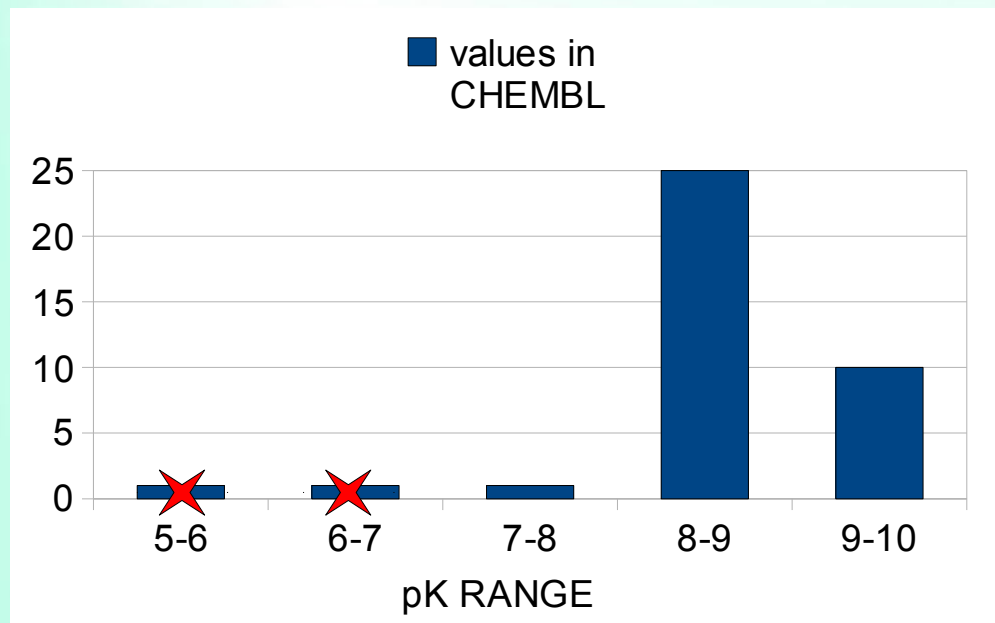
- Molecular Docking

Experimental Data Quality: the example of CHEMBL* DRD2:CHEMBL54 data

DRD2_HUMAN

CHEMBL54

SAFAN-DB



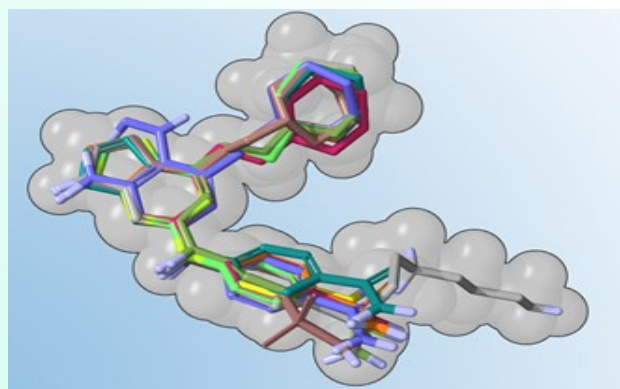
8.52

Data in the original database, coming from different experiments span 5 log units. **SAFAN_DB** retains the most probable.






















*<https://www.ebi.ac.uk/chembl/>

www.safan-bioinformatics.it

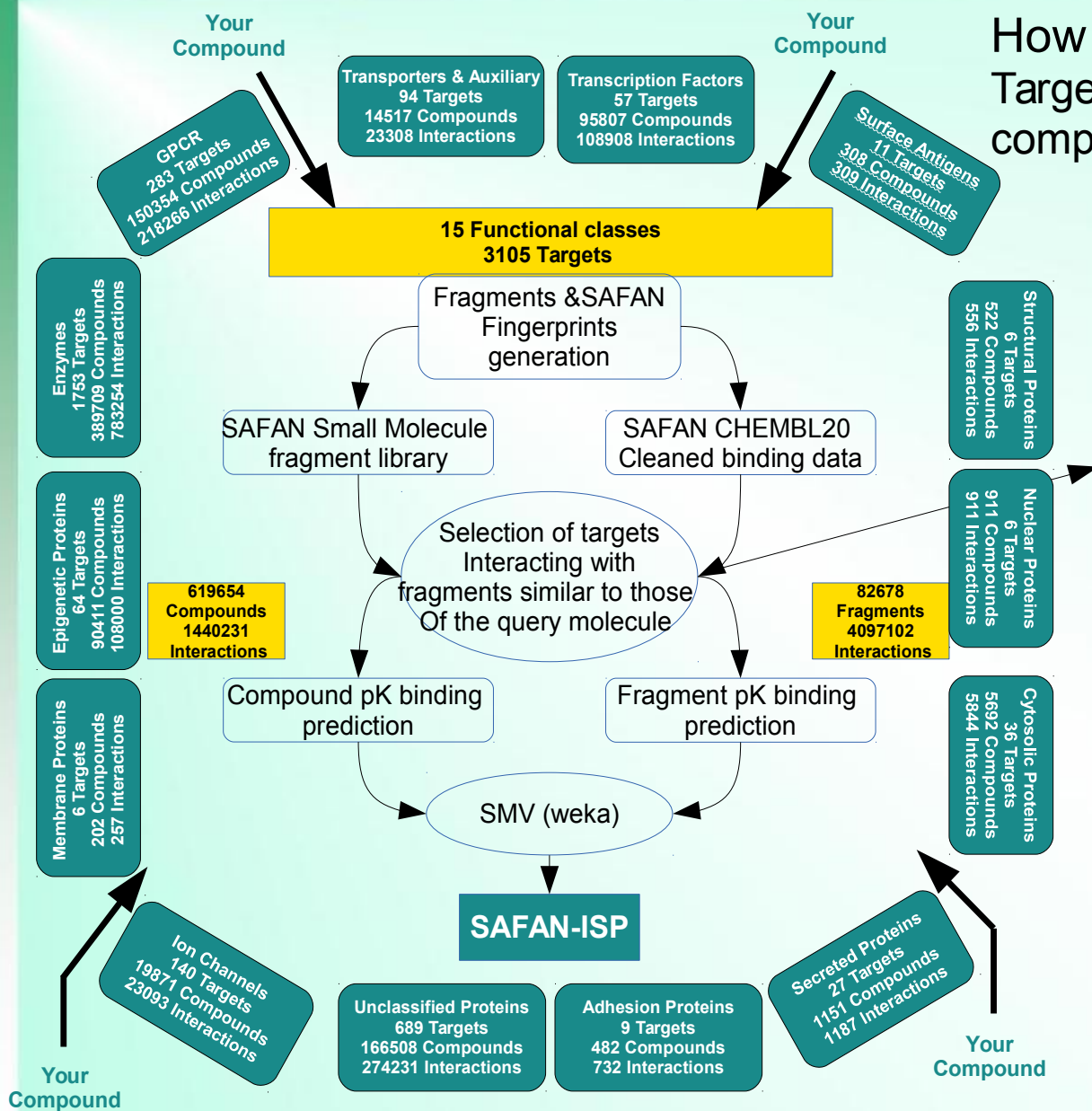
SAFAN_ISP uses 2D similarity searches
because conformations may induce in error



Prediction methods comparison

Method	2-D similarity search	3-D similarity search	Binding site similarity search	Docking
Speed				
Conformation dependency				
Automation				
Target coverage				
Stereochemistry				
Straightforward scoring				
Ligand dependency				
Orphan targets				
Binding mode				

How SAFAN_ISP works: Target classes and compounds included



If:

- The compound contains a selective fragment for that target
- Or
- All the compound fragments are included in SAFAN_fragments DB
- Or
- Similarity computed from fragments is above 0.6

Validation on a benchmarking dataset

DUD: A Directory of Useful Decoys

DUD is a directory of useful decoys for benchmarking *virtual screening methods*.

It contains:

A total of 2,950 active compounds against a total of 40 targets

For each active, 36 "decoys" with similar physical properties (e.g. molecular weight, calculated LogP) but dissimilar topology.

<http://dud.docking.org/r2/>

DUD LIB VS 1.0 is a DUD subsets for ligand-based virtual screening clustered and selected following *J. Comput.-Aided Mol. Des.* 2008, 22, 169–178. indications

How to evaluate results?

ROC AUC: **The expected proportion of positives ranked before a uniformly drawn random negative.** It becomes more reliable if, the performance of each chemotype is averaged for all the molecules sharing it

$$ROC\ AUC_{CA} = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^{k_i} \frac{1 - f_{ij}}{k_i}$$

k_i : number of active in each chemotype
 m : number of chemotypes
 f_{ij} : fraction of decoys ranked better than the i th active in the j th chemotype

Reference values:

- 1: all positives top scoring
- 0: all negatives top scoring
- 0.5 random distribution

However ROC AUC is not weighted toward measuring early retrieval.

For this was introduced the BAROC metric, high if the ranking is below α and dropping quickly to 0 if it is higher.

$$BAROC(\alpha) = \sum_{i=1}^{n_a} e^{-\alpha f_i}$$

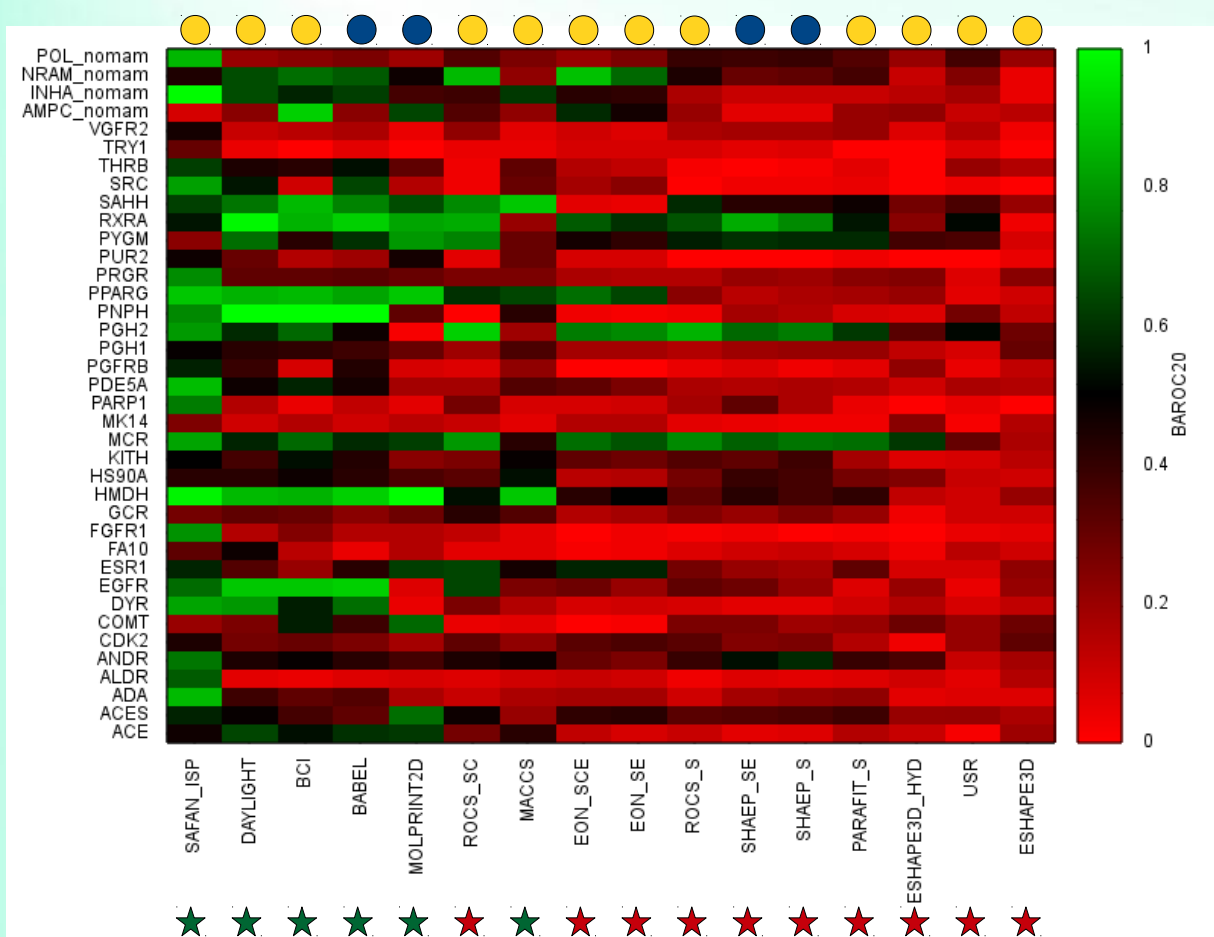
f_i : fraction of decoys ranked better than the i th active.
 n_a : number of actives

Reference values:

- 1: all positives scoring higher than α
- 0: all positives scoring lower than α
- $1/\alpha$: random distribution

SAFAN_ISP performs better on DUD dataset than several scoring functions availables

Venkatraman et al. J. Chem. Inf. Model., 2010, 50 (12), pp 2079–2093



Methods are sorted by average BAROC20

● Freely available

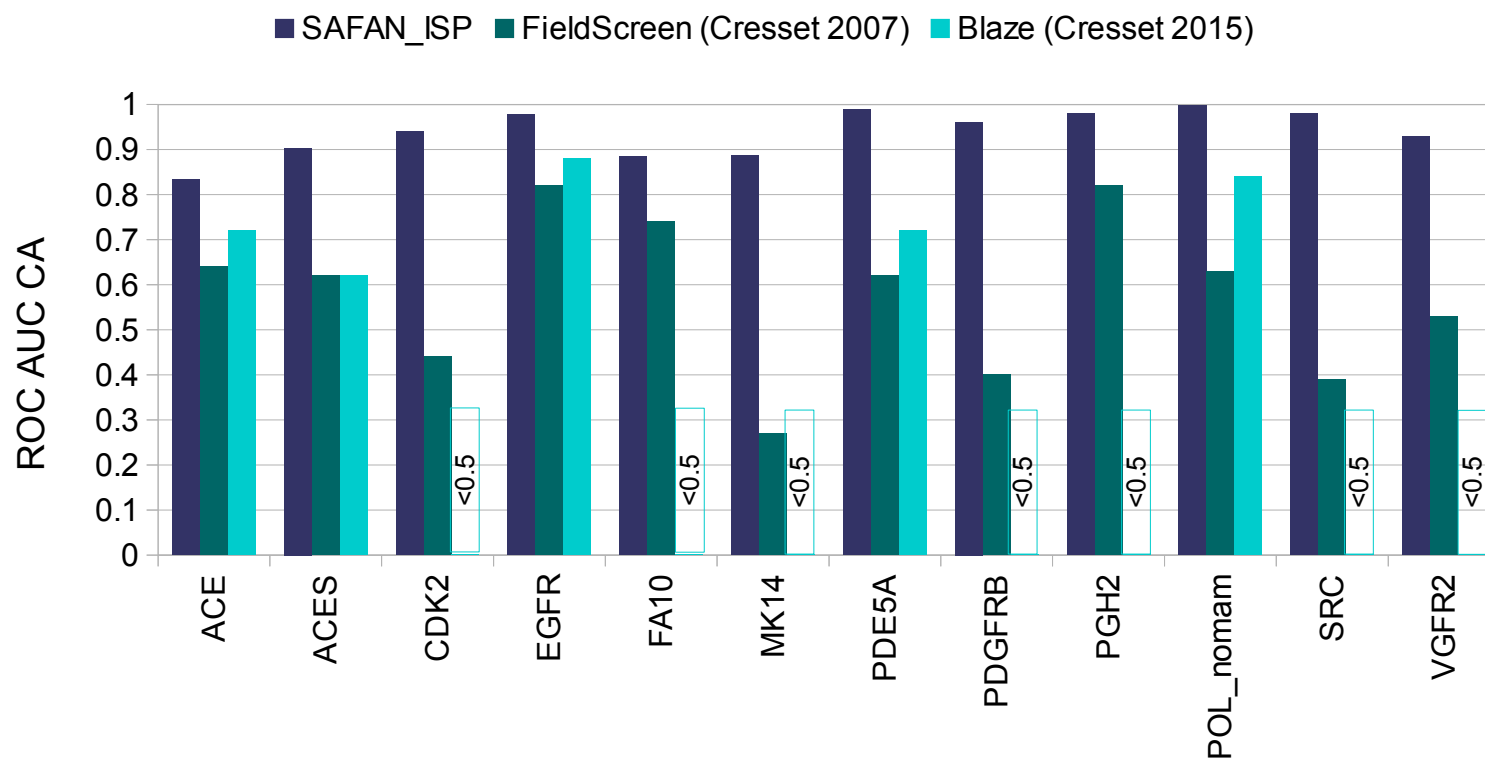
● commercial

★ 2D method

★ 3D method

For 20 targets SAFAN_ISP performs better than all the other methods analyzed.

And that's true also for very recent softwares!



SAFAN-ISP: safan validation set & predictivity

ONE LEAVE OUT methods on:

- 4800 compounds
- 1455 protein targets (35 classes) and
- 19588 interactions

In binary classification, i.e. after setting a cutoff, all values above the cutoff are positives and all values below are negatives:

- **Accuracy** is the proportion of true results (positive and negatives) among the total number of cases .
- **Precision** is the proportion of true positives among all trues.
- **Recall** is the proportion of true positive retrieved.
- **Specificity** is the proportion of negatives that are correctly identified.

*A measurement is considered valid if it is both **accurate and precise**.*

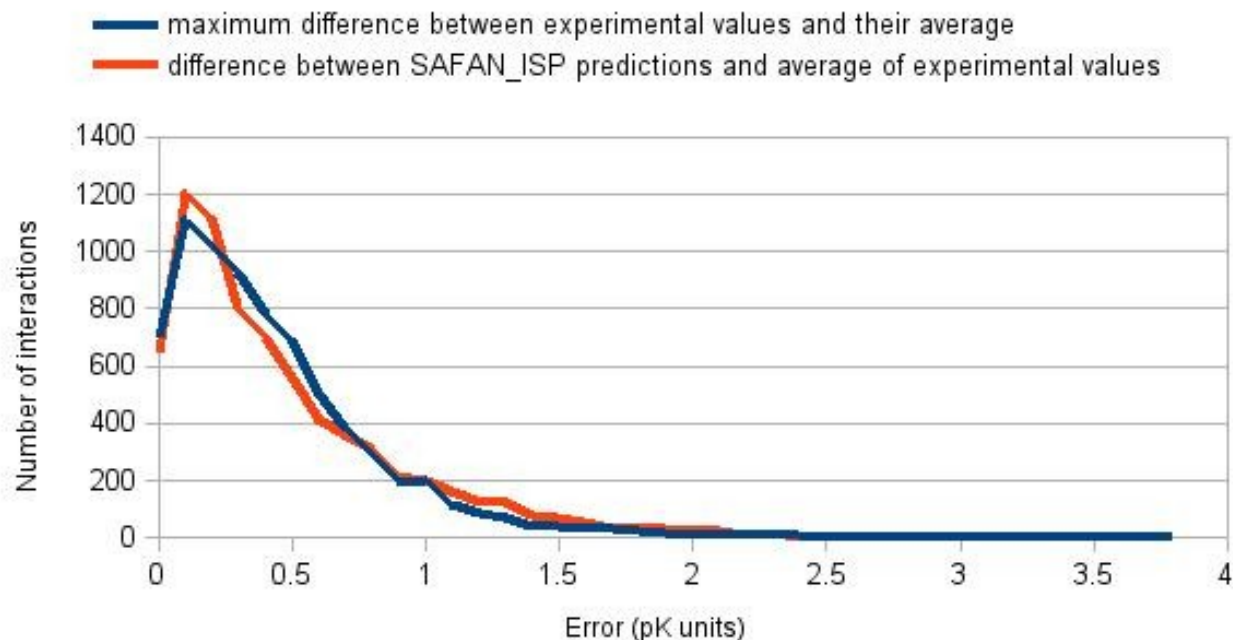
Interactions	4205
Pearson	0.86
Accuracy (50nM)	0.83
Specificity (50nM)	0.96
Precision (50nM)	0.87
Recall (50nM)	0.58
% of True Positives	89

SAFAN_ISP predictions are not only accurate and precise but also quantitatively correlated with the experimental data.

SAFAN_ISP can replace experiments

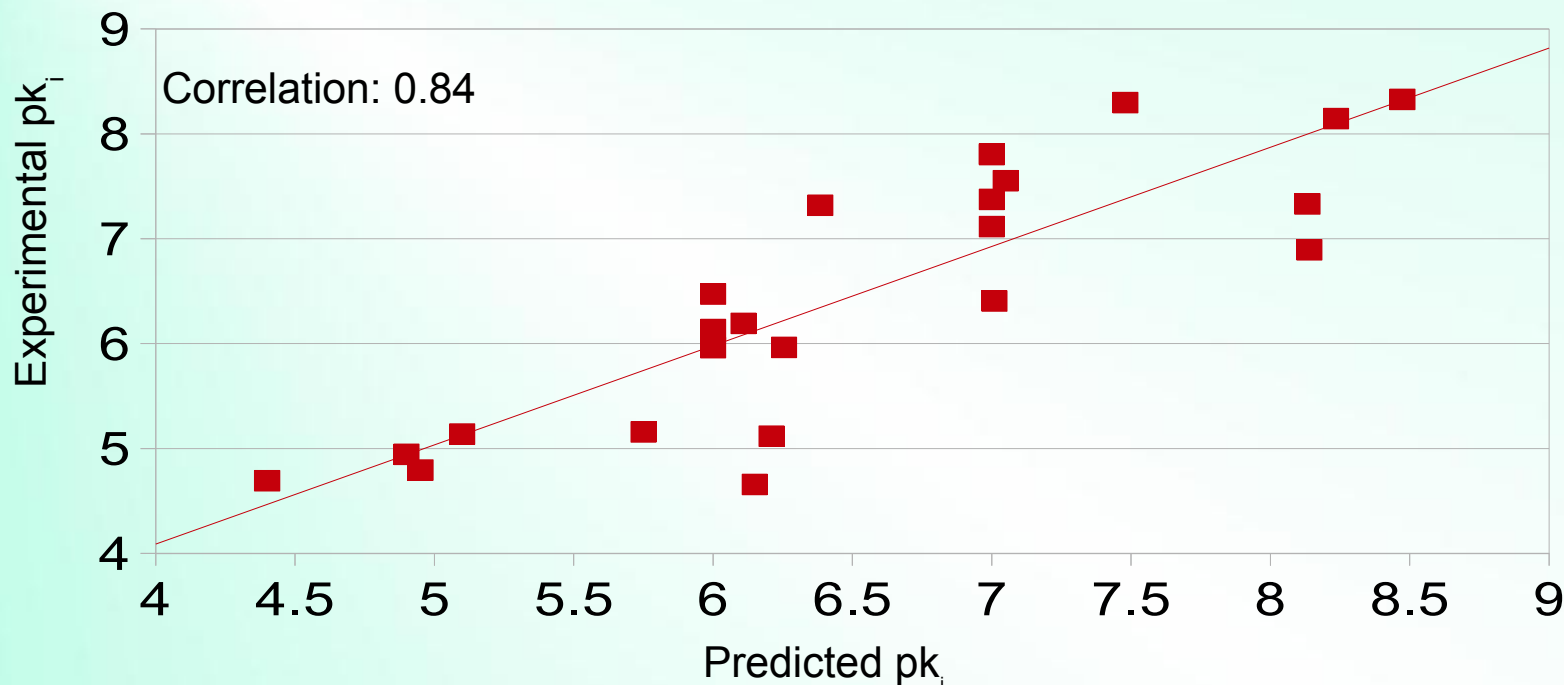
DATA SET:

- 5943 compounds
- 7351 interactions
- 4.5 average number of experiments replication



SAFAN_ISP predictions are quantitatively correlated with experimental data.

Test Case: Aripiprazole



Target predicted:

5HT1A	5HT1B	5HT1D	5HT2A	5HT2C	5HT3A	5HT6R	5HT7R
ADA1A	ADA1B	DRD1	DRD2	DRD3	DRD4	HRH1	IMPA1
KCNH2	LYAG	MDR1	NPSR1	RORG	SC6A4	TAU	TYDP1

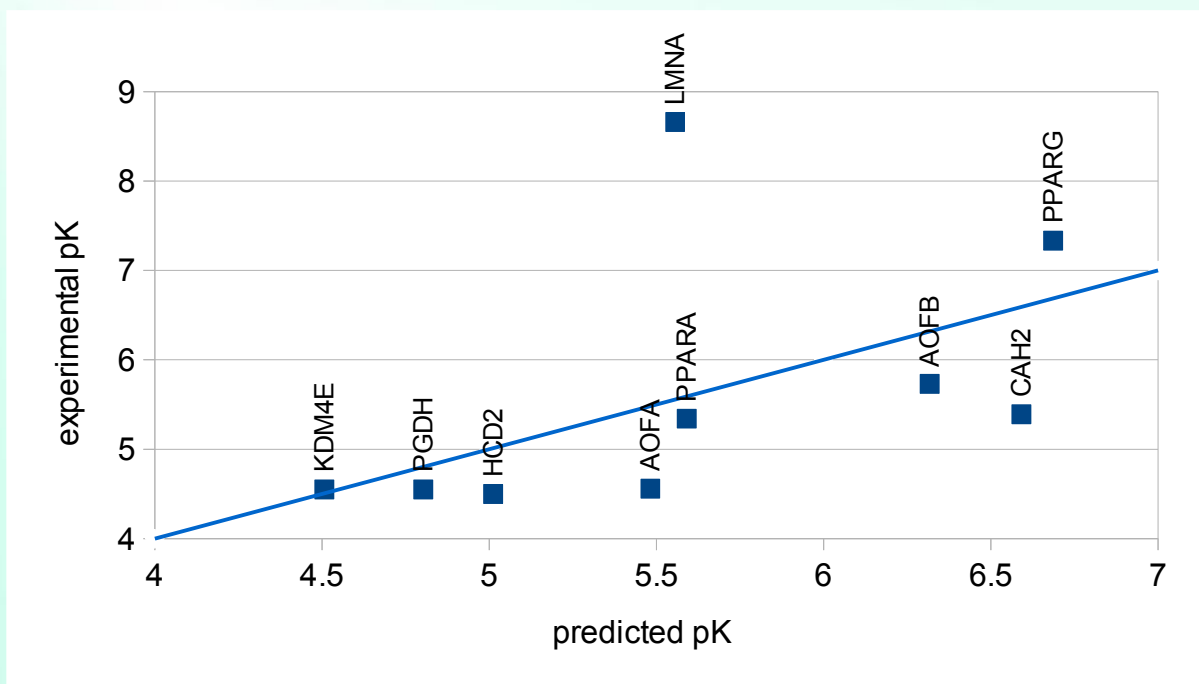
Real Life examples (1)

Rosiglitazone is an anti-diabetic drug in the thiazolidinedione class of drug. Because it increases the risk of heart attack it was withdrawn in Europe in 2010. This drug continues to be available in the US. Could **SAFAN_ISP** forecast it?

SAFAN_ISP newly predicted targets

Target: Mnemonic Name	Target: Gene Name	pK	Average Similarity with SAFAN-ISP fragments	Average Similarity with SAFAN-ISP compounds	Expected Correlation with in Vitro Data	Indications	Side effects
BRAF	BRAF	7.338	0.761	0.569	0.725	cancers, nervous system disorders, other congenital disorders, skin and subcutaneous tissue disorders	
KIT	KIT	7.148	0.786	0.55	0.725	blood and lymphatic system disorders, cancers, congenital disorders of metabolism, gastrointestinal disorders, nervous system disorders reproductive system and breast, disorders skin and subcutaneous tissue disorders	Lung disorder
PPARD	PPARD	7.145	0.807	0.565	0.772	metabolism and nutrition disorders (Diabetes mellitus)	Neuropathy peripheral
ESR2	ESR2	6.953	0.596	0.572		cancers	
GSK3B	GSK3B	6.877	0.775	0.696	0.725	nervous system disorders, metabolism and nutrition disorders (Diabetes mellitus)	Cardiac failure, Acute coronary syndrome, Cardiomegaly
ESR1	ESR1	6.438	0.576	0.618		cancers, musculoskeletal and connective tissue disorders, reproductive system and breast disorders	Fracture
ADRB3	ADRB3	6.344	0.781	0.654	0.78	metabolism and nutrition disorders (Diabetes mellitus), vascular disorders	Cardiovascular disorder, Cardiac failure, Hyperlipidaemia, Colitis, Hypertension
THRB	F2	6.248	0.73	0.65	0.857	blood and lymphatic system disorders cancers cardiovascular diseases congenital, familial and genetic disorders hepatobiliary disorders immune system disorders infections and infestations metabolism and nutrition disorders musculoskeletal and connective tissue disorders neoplasms benign, malignant and unspecified pregnancy, puerperium and perinatal conditions respiratory, thoracic and mediastinal disorders vascular disorders	Acute coronary syndrome. Cardiovascular disorder, Hypertension, Cerebral infarction, Transient ischaemic attack, Contusion, Coagulopathy, Post procedural complication, Ischaemia, Breast cancer, Thrombosis, Haemorrhage, Diabetes mellitus, Disseminated intravascular coagulation, Peripheral vascular disorder, Angiopath
DHB2	HSD17B2	6.237	0.363	0.496	0.861	cancers, reproductive system and breast disorders	
LKHA4	LTA4H	6.126	0.597	0.599	0.87		Asthma
VGFR2	KDR	6.059	0.747	0.602	0.725	cancers, infections and infestations	Erectile dysfunction, Cystitis, Colitis, Cerebral haemorrhage, Myocardial ischaemia, Hypertension

Rosiglitazone correlation for known targets



The line indicates a perfect match. With the exception of prelaminin all known targets well predicted.

Real Life examples (2)

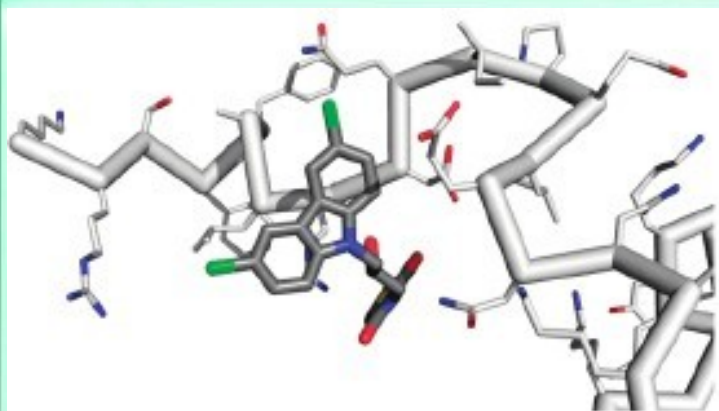
Could SAFAN_ISP help avoiding the BIA 10-2474 clinical disaster?

Target: Mnemonic Name	Target: Gene Name	pK	Average Similarity with SAFAN-ISP fragments	Average Similarity with SAFAN-ISP compounds	Expected Correlation with in Vitro Data	Indications	Side effects
V1AR	AVPR1A	8.871	0.734	0.238	0.828	psychiatric disorders	Cardiomegaly, Acute coronary syndrome, Schizophrenia, Haemorrhage, Hypertension
CSF1R	CSF1R	8.155	0.992	0.395	0.766	cancers, musculoskeletal and connective tissue disorders,	Neoplasm
SC6A9	SLC6A9	7.935	0.796	0.519		nervous system diseases	
CBPB2	CPB2	7.771	0.4	0.443	0.87	vascular disorders	
HRH3	HRH3	7.569	1	0.647	0.78	psychiatric disorders	Hyperaemia, Amnesia, Convulsion, Hypertension, Hyperglycaemia
SHH	SHH	7.529	0.779	0.348		cancers, musculoskeletal and connective tissue disorders, digestive diseases, eye disorders, nervous system disorders, skin and subcutaneous disorders	Hypothyroidism
ADA	ADA	7.401	0.8	0.141	0.87	blood and lymphatic system disorders, cancer, immune system disorders, infections, nervous system disorders, respiratory disorders	Hypertension, Hyperglycaemia, Peritonitis
CATS	CTSS	7.131	0.784	0.396	0.87	cancers, musculoskeletal and connective tissue disorders, skin and subcutaneous disorders	
ACHA7	CHRNA7	7.1	0.518	0.473		psychiatric disorders	Convulsion, Withdrawal syndrome
PTAFR	PTAFR	6.937	0.798	0.295	0.828		
MK08	MAPK8	6.858	0.985	0.333	0.766	cancers, metabolism and nutrition disorders, nervous system disorders	Dermatitis atopic, Hypertension, Cerebrovascular accident
QPCT	QPCT	6.789	0.4	0.402	0.766		
GPBAR	GPBAR1	6.767	0.6	0.204	0.828		
THAS	TBXAS1	6.733	0.799	0.367		musculoskeletal and skin diseases, vascular disorders	Hypertension
CNR2	CNR2	6.674	0.793	0.54	0.613		Breast neoplasm, Pain, Arthritis, Immune system disorder, Osteoporosis
JAK2	JAK2	6.606	0.726	0.308	0.766	blood and lymphatic system disorders, cancer, nervous system disorders, vascular disorders	Myocardial ischaemia, Cerebrovascular accident , Transient ischaemic attack
HPGDS	HPGDS	6.454	0.4	0.179	0.766	cancers	Asthma
TSSK2	TSSK2	6.386	0.668	0.193	0.766	reproductive system and breast disorders	
IRAK4	IRAK4	6.33	0.742	0.42	0.766	immune system diseases	
JAK3	JAK3	6.262	0.734	0.41	0.766	cancers, immune system disorders	
IMPA1	IMPA1	6.209	0.662	0.522	0.87	psychiatric disorders, vascular disorders	
HS90A	HSP90AA1	6.156	0.692	0.423		cancers, renal and urinary disorders	
AAPK1	PRKAA1	6.154	0.583	0.291	0.766		
ALR	GFER	6.127	0.721	0.353	0.861		
GSK3A	GSK3A	6.066	0.689	0.382	0.766		

yes

Why Choose SAFAN_ISP

- Predictions are quantitative and average errors are comparable to experimental ones.
- It performs in average better than other methods available.
- We can apply classical structural bioinformatics methods to analyze in more details drug:target interactions outlined by SAFAN_ISP
- Our experience and our know-how allow us to help you in many different projects.
- SAFAN_ISP gives you true predictions allowing you to patent your results.



Luisa Pugliese
S.A.F.AN. BIOINFORMATICS
Via Fulvio Croce 23/B Torino
+390112767365
luisa.pugliese@safan-bioinformatics.it

“Dr. Luisa Pugliese is a leading scientist for 'in silico' methods for discovery of new lead compounds for drugs.

I strongly recommend Luisa as an outstanding scientist with a keen sense of doing both basic and applied research in the area of structural biology and structure based drug discovery.”

- Joel L. Sussman

Professor of Structural Biology Director of the Israel Structural Proteomics Center