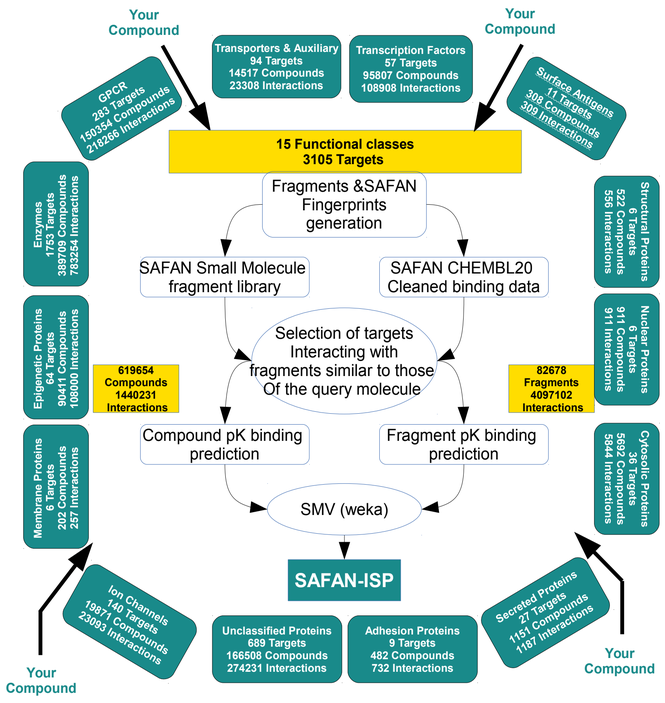
SAFAN-ISP: A NEW SMALL MOLECULE FRAGMENT INFORMATICS APPROACH FOR IN-SILICO PROFILING,DRUG REPOSITIONING AND SIDE EFFECT PREDICTION

Luisa Pugliese*, Marco Botta** & Frank Dullweber ***
*S.A.F.AN. BIOINFORMATICS via Fulvio Croce 10136 Torino Italy, ** Dipartimento di Informatica Corso Svizzera 185 10149 Torino Italy, *** Boehringer Ingelheim Pharma GmbH&Co. KG Birkendorfer Str. D - 88400 Biberach(Riss)
Web:www.safan-bioinformatics.it e-mail:luisa.pugliese@safan-bioinformatics.it

ABSTRACT: In recent years the interest in using computational bioactivity profiling (CBP) in order to predict the most likely macromolecolar targets of bioactive compounds in early drug discovery has grown for many reasons:
- Phenotypic screens are now the primary source of first-in-class low molecular weight drugs. Since phenotype to target identification is successful in only ca. 40% of cases, CBP is a fast and cheap alternative to experimental target identification approaches.
- The early identification of possible off-targets is an important preclinic safety asset that helps to avoid potential side effects and severe adverse reactions.
- The competitive advantage in targeting several a-priori unrelated targets to other therapeutic indications promotes the design of polypharmacological ligands with controlled selectivity profiles.
- Because the easiest route to a novel drug is to start from an old one, pressure towards repurposing clinical candidates or marketed drugs to novel indications and targets is increasing, sparing the long and costly initial clinical trials.

Fragment chemogenomics
Using powerful concepts in modern chemistry and biology and linking combinatorial chemistry with genomics and proteomics, chemogenomics includes systematic relationships between targets and ligands. By adding information about molecular fragments SAFAN-ISP samples chemical space more effectively than what could be obtained using drug-sized molecules. Using SAFAN structural fingerprints to score similarity of fragments and compounds and applying Support Vector Machine algorithm SAFAN-ISP searches for and **quantitatively** predicts new drug:target interactions.

How it works



Method	2-D similarity search	3-D similarity search	Binding site similarity search	Docking	SAFAN-ISP
Speed	👍	👍	👍	👍	👍
Conformation dependency	👍	👎	👎	👎	👍
Automation	👍	👍	👍	👍	👍
Target coverage	👍	👍	👎	👎	👍
Stereochemistry	👎	👍	👍	👍	👎
Straightforward scoring	👍	👍	👎	👎	👍
DATA dependency	👎	👎	👍	👍	👎
Orphan targets	👎	👎	👍	👍	👎
Binding mode	👎	👎	👍	👍	👎

Validation strategy

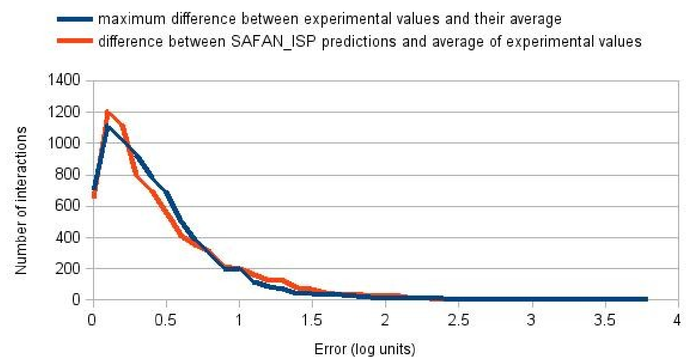
SAFAN-ISP performances were compared to the experiemntal data errors by, using the LEAVE ONE OUT validation method, on 5943 compounds for which on 7351 interactions with different targets, an average of 4 replications are available within CHEMBL database.

Validation Set

Compounds	5943
Targets	2878
Target classes	38
Compound:Target Interactions	7351

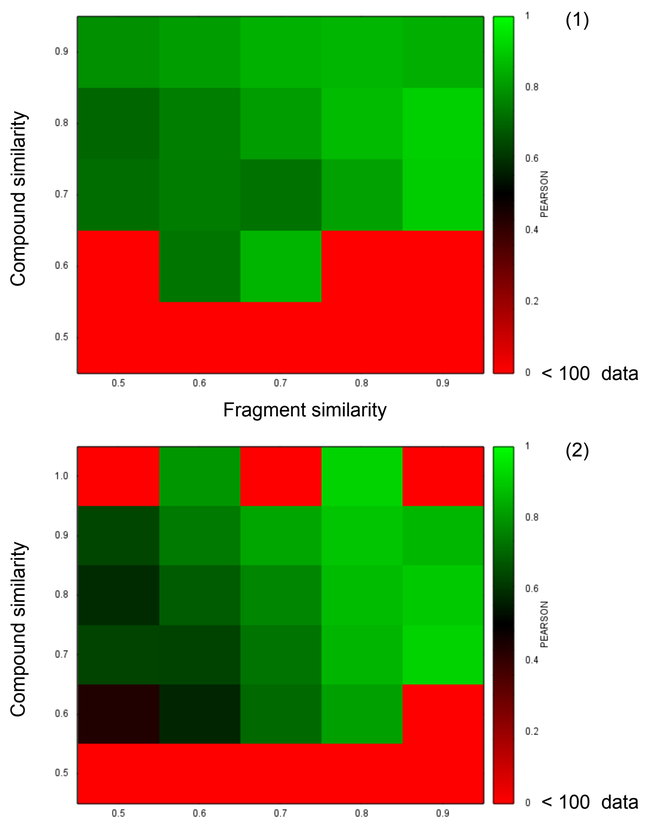
Predictions quality

Comparison with published in vitro data



WHAT ARE SPECIFIC FRAGMENTS?

Specific fragments influence binding of the molecules to targets more than other fragments and are the heart of SAFAN-ISP technology. In order to evaluate their importance the Pearson correlation between predicted and experimental data were computed in function of the fragment and compound similarity in presence (1) and absebce (2) of SPECIFIC FRAGMENTS.

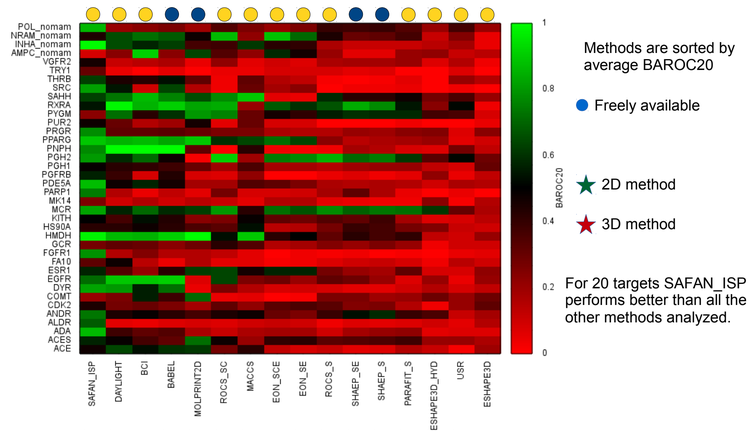


Comparison with similar technologies

Not many data, coming from different providers are available to compare quantitative SAFAN-ISP. However the comparison of several methods, currently used for virtual screening, was published on Venkatraman et al. J. Chem. Inf. Model., 2010, 50 (12), pp 2079–2093.
Running each method on the DUD (A Directory of Useful Decoys) dataset available at http://dud.docking.org/r2/ Venkatraman et al. were able to rank them.
In order to compare SAFAN-ISP with all the methods reported by Venkatram et al. SAFAN-ISP was applied to all the compounds of the DUD dataset.
Results were evaluated using the BAROC(α) metric:

$$BAROC(\alpha) = \sum_{i=1}^{n_s} e^{-\alpha f_i} \qquad \text{Where } \alpha=20$$

high if the ranking is below α and dropping quickly to 0 if it is higher.



SAFAN-ISP in vitro validation

SAFAN-ISP was tested with compounds from Boehringer Ingelheim
For one compound, 13 targets predicted by SAFAN-ISP were then validated in vitro.
JACK3, MK08 and MK0,1highlighted in cyano, were chosen because the fragments similarity is high, the compound similarity is low and the predicted correlation is good because this compound contains a SELECTIVE FRAGMENT for those targets.
The relatively low error between the predictions and the experimental data validates the SELECTIVE FRAGMENT selection algorithm.

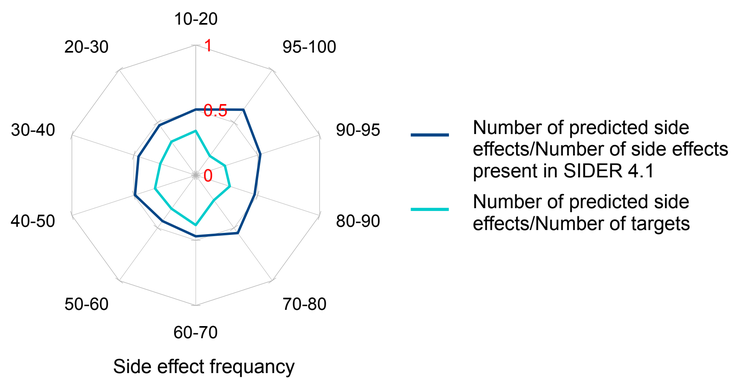
Target: Mnemonic Name	Target: Gene Name	Predicted pK	Experimental results pK	Average Similarity with SAFAN-ISP fragments	Average Similarity with SAFAN-ISP compounds	Expected Correlation with in Vitro Data	Indications	Side Effects
1 target								
FA10	F10	7.501	<5 Not detected	0.92	0.63	0.857		Cardiovascular disorder* Diseases of the circulatory system* Hypogonadism
1 target								
AGTR2	AGTR2	7.393	<5 Not detected	0.658	0.523	0.613		Cardiovascular disorder* Diseases of the circulatory system* Diseases of the digestive system* Diseases of the eye and adnexa* Diseases of the respiratory system* Endocrine, nutritional and metabolic diseases* Hypertension* Neoplasms
1 target								
FLT3	FLT3	6.925	<5 Not detected	0.721	0.438	0.766	Neoplasms	Diseases of the musculoskeletal system and connective tissue
CSF1R	CSF1R	6.86	<5 Not detected	0.595	0.479	0.766	Neoplasms	Diseases of the nervous system* Diseases of the respiratory system
3 targets								
AURKB	AURKB	6.633	5.818	0.68	0.632	0.412	Neoplasms	
IKKB	IKBKB	6.538	5.058	0.697	0.379	0.766		Diseases of the respiratory system
JAK3	JAK3	6.485	6.027	0.96	0.395	0.766	Neoplasms	Diseases of the respiratory system
1 target								
5HT1A	HTR1A	6.473	<5 Not detected	0.759	0.511	0.613	Mental, Behavioral and Neurodevelop mental disorders	Diseases of the digestive system* Diseases of the nervous system* Diseases of the skin and subcutaneous tissue
3 targets								
FGFR1	FGFR1	6.424	5.091	0.74	0.56	0.412	Congenital malformations, deformations and chromosomal abnormalities* Endocrine, nutritional and metabolic diseases* Neoplasms	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism* Diseases of the genitourinary system* Diseases of the respiratory system* Diseases of the skin and subcutaneous tissue* Hypertension
1 target								
AKT1	AKT1	6.267	6.194	0.742	0.564	0.412	Diseases of the digestive system* Endocrine, nutritional and metabolic diseases* Neoplasms* Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	Diseases of the nervous system* Diseases of the respiratory system* Diseases of the skin and subcutaneous tissue* Hypertension
4 targets								
MK08	MAPK8	6.224	5.178	0.988	0.416	0.766	Diseases of the circulatory system* Neoplasms	Diseases of the musculoskeletal system and connective tissue* Diseases of the skin and subcutaneous tissue* Endocrine, nutritional and metabolic diseases* Hypertension
9 targets								
IRAK4	IRAK4	6.019	<5 Not detected	0.573	0.403	0.766		Injury, poisoning and certain other consequences of external causes
68 targets with pk beween 6 and 5								
MK01	MAPK1	4.928	5.096	0.808	0.643	0.766	Neoplasms	Diseases of the circulatory system* Diseases of the digestive system* Diseases of the musculoskeletal system and connective tissue* Diseases of the respiratory system* Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
24 targets with pk < 5								

For three other compounds, experimental data on the original target were available to Boehringer Ingelheim but unknown to S.A.F.AN. BIOINFORMATICS when SAFAN-ISP was run.

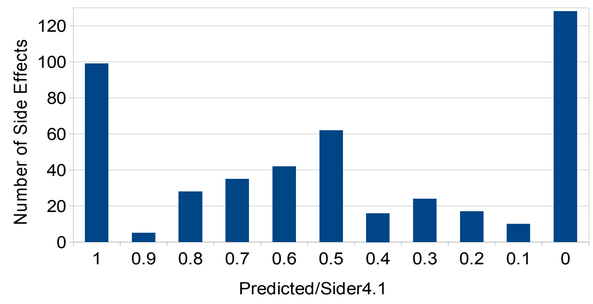
	Predicted pK	Measured pK
Compound A	**9**	**8.6**
Compound B	**6.8**	**7**
Compound C	**6.8**	**7.6**

IS SAFAN-ISP able to forecast side effects?

In order to evaluate if SAFAN-ISP prediction are useful to forecast side effects, we run it on 753 compounds described in SIDER 4.1 Side Effect Resource (http://sideeffects.embl.de/) .
Side displaying a frequency higher than 10% were considered for the analysis. About 50% of the side effects were correctly forecasted. In many cases more than 3 targets are associated with the same side effect.



466 Side Effects retrieved from SIDER4.1 database display a frequency higher Than 10%. The followin graph shows the number of side effects obtained from the predictions as a function of the % of retrieval.



REAL LIFE EXAMPLE:

Rofecoxib is a nonsteroidal anti-inflammatory drug (NSAID) that has now been withdrawn over safety concerns. On September 30, 2004, Merck withdrew rofecoxib from the market because of concerns about increased risk of heart attack and stroke associated with long-term, high-dosage use. Could SAFAN_ISP forecast it?

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
CAH2	CA2	7.657	0.942	0.764	0.789	Congenital malformations, deformations and chromosomal abnormalities* Diseases of the genitourinary system	
SGMR1	SIGMAR1	7.544	0.883	0.634		Diseases of the nervous system	
PD2R2	PTGDR2	7.121	0.904	0.71		Diseases of the respiratory system	
EDNRA	EDNRA	7.071	0.86	0.647	0.78		Diseases of the circulatory system* Diseases of the digestive system* Diseases of the eye and adnexa* Diseases of the nervous system* Diseases of the respiratory system* Neoplasms
DRD3	DRD3	7.064	0.997	0.514	0.78	Diseases of the nervous system	Hypertension* Neoplasms* Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
MMP2	**MMP2**	7.047	0.921	0.676	0.857	Diseases of the circulatory system* Neoplasms	Atrial fibrillation* Connective tissue disorder* Diseases of the genitourinary system* Diseases of the respiratory system* Diseases of the skin and subcutaneous tissue* Endocrine, nutritional and metabolic diseases* Hypertension* Injury, poisoning and certain other consequences of external causes
ACM3	CHRM3	7.029	0.668	0.58	0.828	Congenital malformations, deformations and chromosomal abnormalities* Diseases of the genitourinary system* Pregnancy, childbirth and the puerperium	Diseases of the respiratory system* Hypertension
MMP9	**MMP9**	7.025	0.915	0.666	0.857	Diseases of the circulatory system* Diseases of the digestive system* Neoplasms	Atrial fibrillation* Connective tissue disorder* Diseases of the genitourinary system* Diseases of the musculoskeletal system and connective tissue* Diseases of the respiratory system* Diseases of the skin and subcutaneous tissue* Endocrine, nutritional and metabolic diseases* Hypertension* Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
PD2R	PTGDR	7.018	0.928	0.496	0.78	Diseases of the respiratory system	

A DEMO VERSION OF THE SYSTEM IS AVAILABLE:
you are invited to contact Luisa Pugliese in case of interest.