

ESR 5: Analysis of Compounds Promiscuity based on a Bioassay Ontology

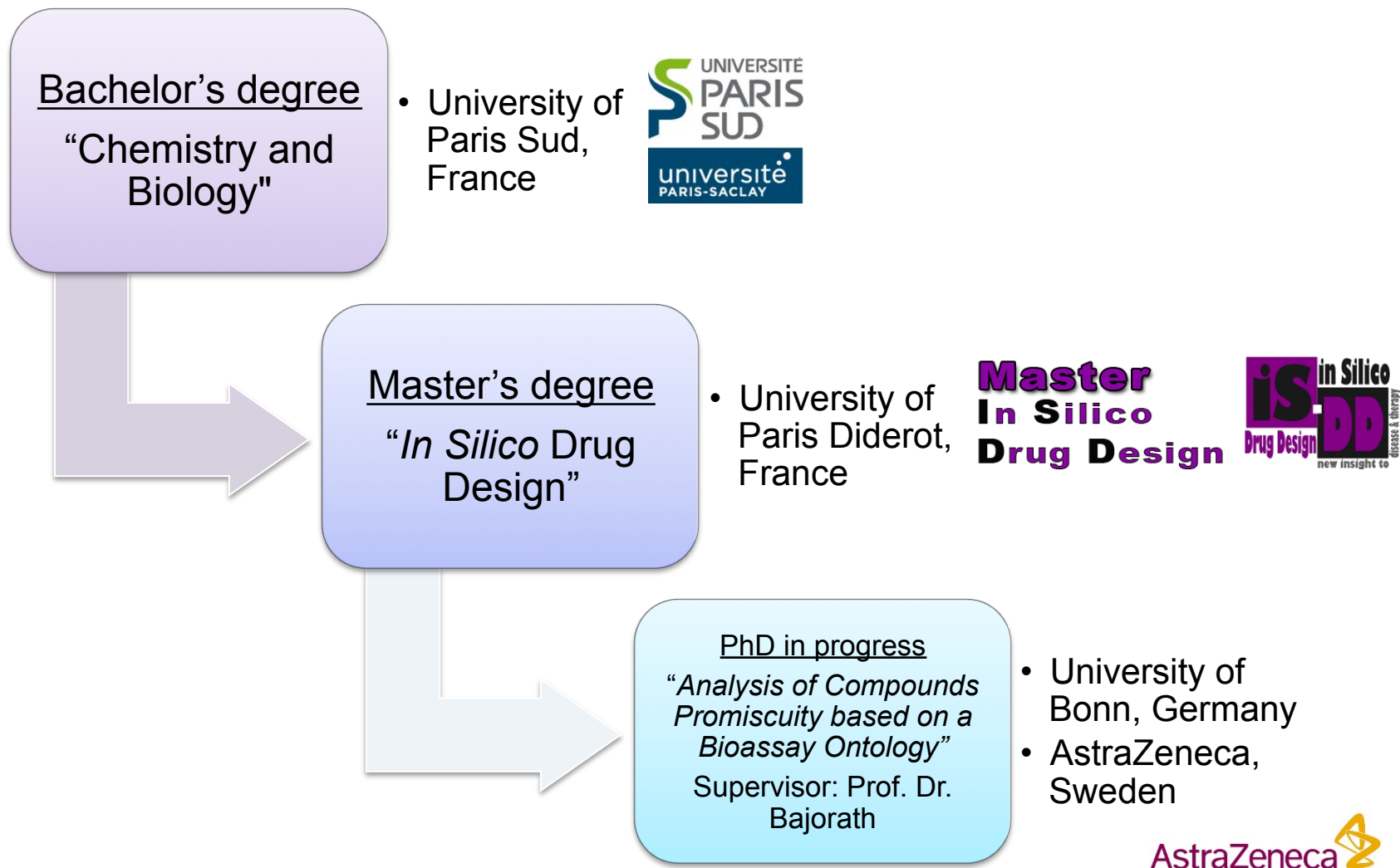
Laurianne David

Supervisor: Prof. Dr. Jürgen Bajorath

B-IT Life Science Informatics

Rheinische Friedrich-Wilhelms-University Bonn

Background

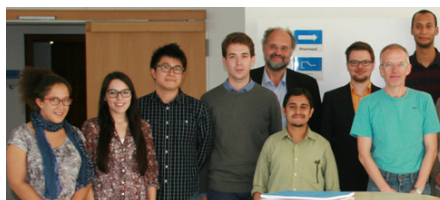


Training Experiences



Bigchem

- Online lectures on various topics related to drug discovery
- First Bigchem School "*Introduction to Chemoinformatics*" hold in Helmholtz Zentrum Munich
- Second Bigchem School "*Chemical databases*" hold in Barcelona



University of Bonn

- Bridging courses in chemistry
- German course (level A1)
- B-IT Lecture Series
- Computational Chemogenomics Workshop
- Attendance to the thesis defenses of Dr. Shilva Kayastha and Dr. Antonio De la Vega de Leon



Systematic Evaluation of Analog Series containing Promiscuous Compounds

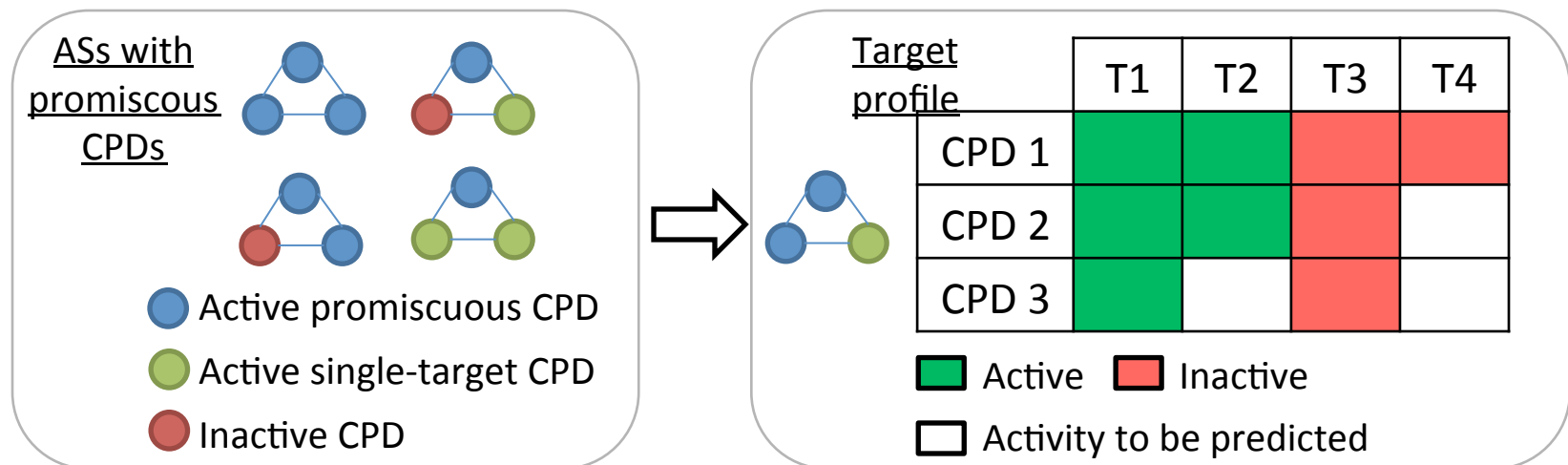
Laurianne David
B-IT Life Science Informatics
Rheinische Friedrich-Wilhelms-University Bonn

Introduction

- **Target identification** is a major topic in medicinal chemistry and chemogenomics
- New target hypotheses for compounds (CPDs) can be inferred from ligand **structural similarity**
- **Analog series** (ASs) are series of CPDs generated on the basis of the matched molecular pair (MMP) formalism
- **Promiscuous CPDs** interact with multiple biological targets

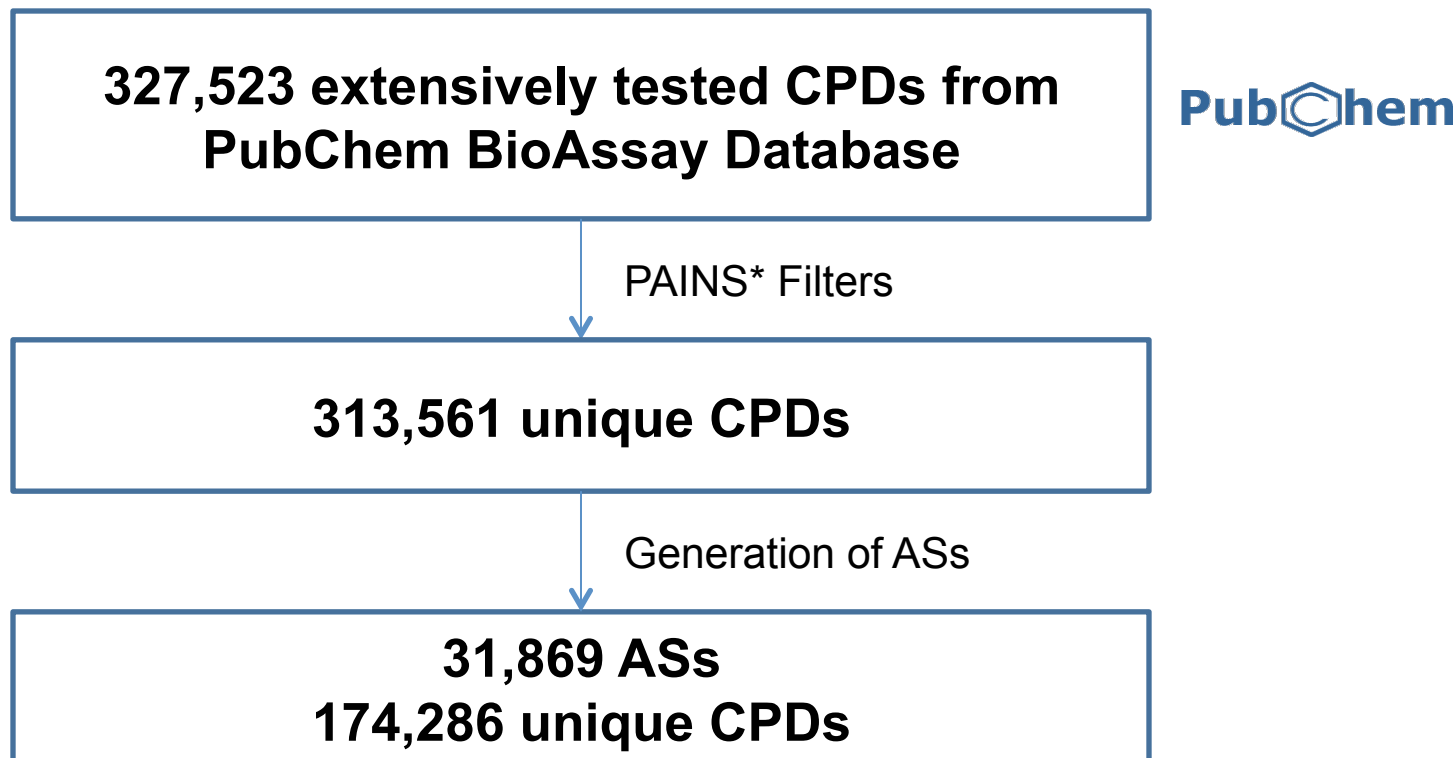
Aim of the Study

- Investigate the potential of **ASs** containing promiscuous CPDs for target identification
- Derive new target hypotheses for analogs



Data Selection

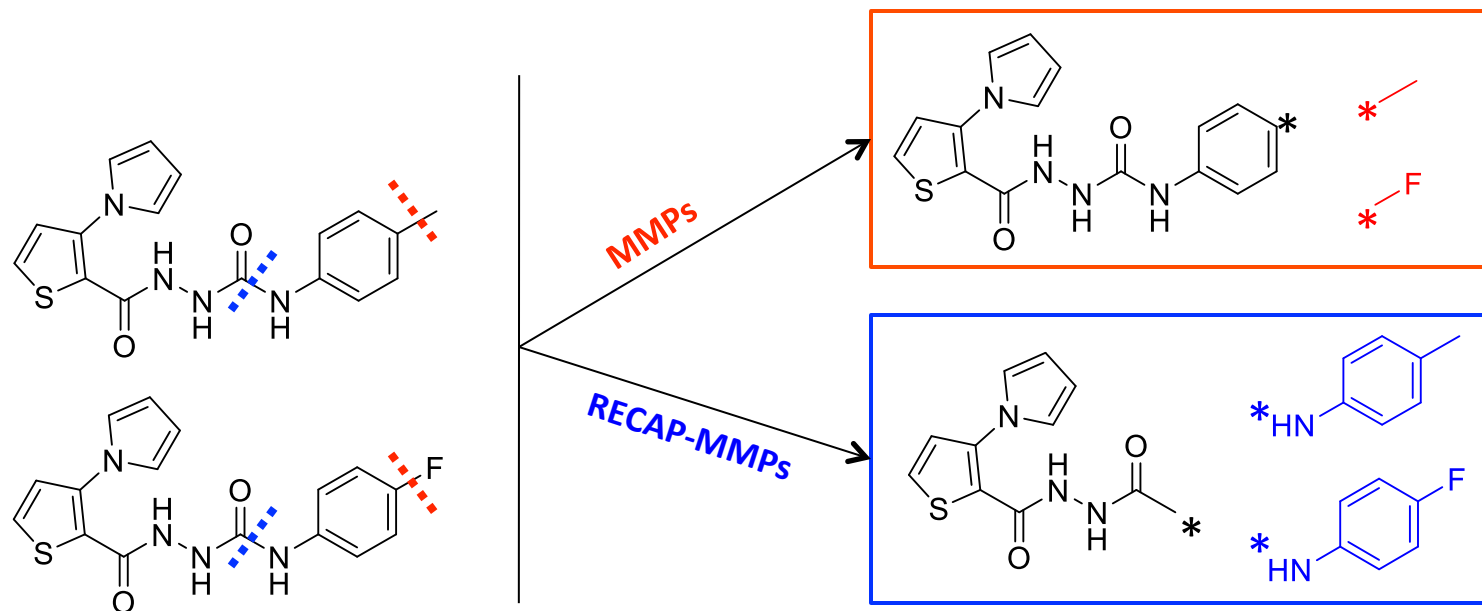
Data Selection



* Pan-Assay Interference Compounds

Small molecule, contained as substructure in a larger compound, presenting reactivity or other liability

Matched Molecular Pairs (MMPs) and RECAP-MMPs*



MMPs

Pairs of CPDs that differ by a structural change at a single point

- Compounds are not related by chemical reactions

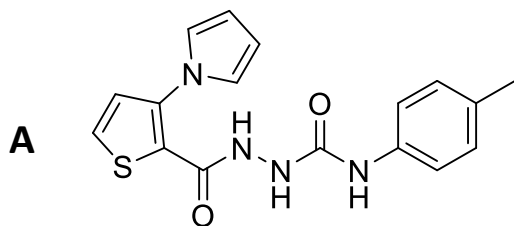
RECAP-MMPs*

- The structural change is based on a specific chemical reaction

* Retrosynthetic Combinatorial Analysis Procedure

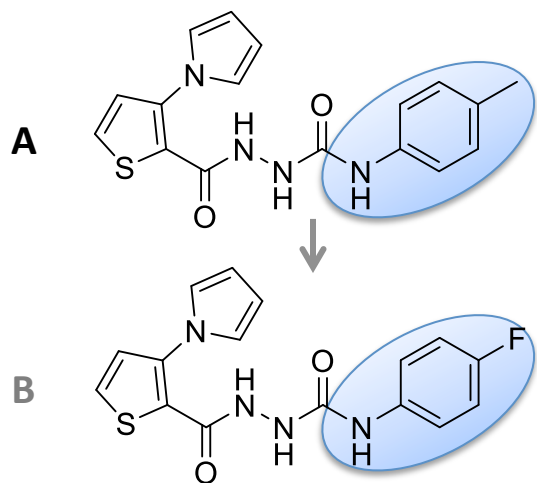
Generation of Analog Series (ASs)

- All pairwise RECAP-MMP relationships for one compound



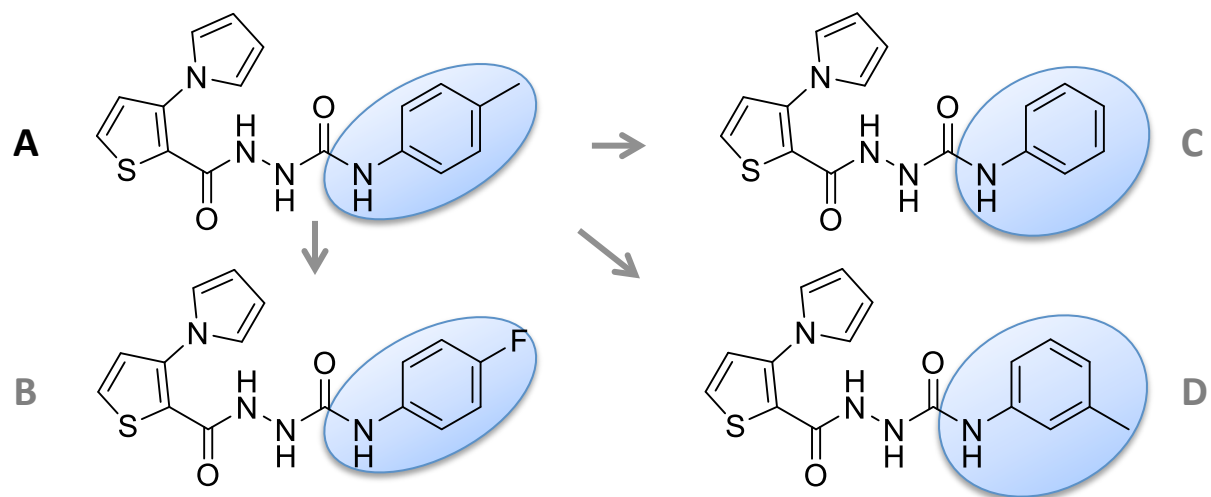
Generation of Analog Series (ASs)

- All pairwise RECAP-MMP relationships for one compound



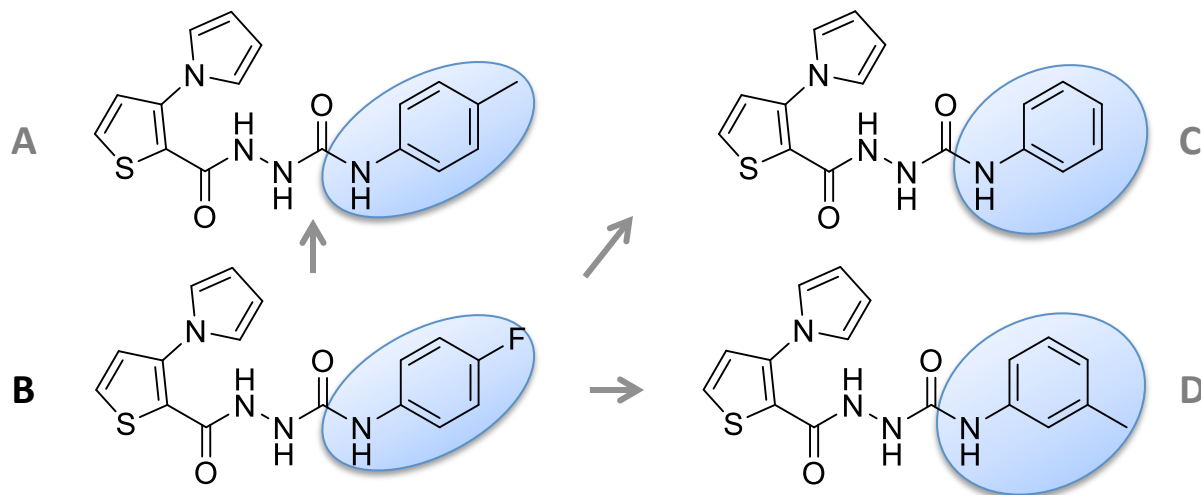
Generation of Analog Series (ASs)

- All pairwise RECAP-MMP relationships for one compound



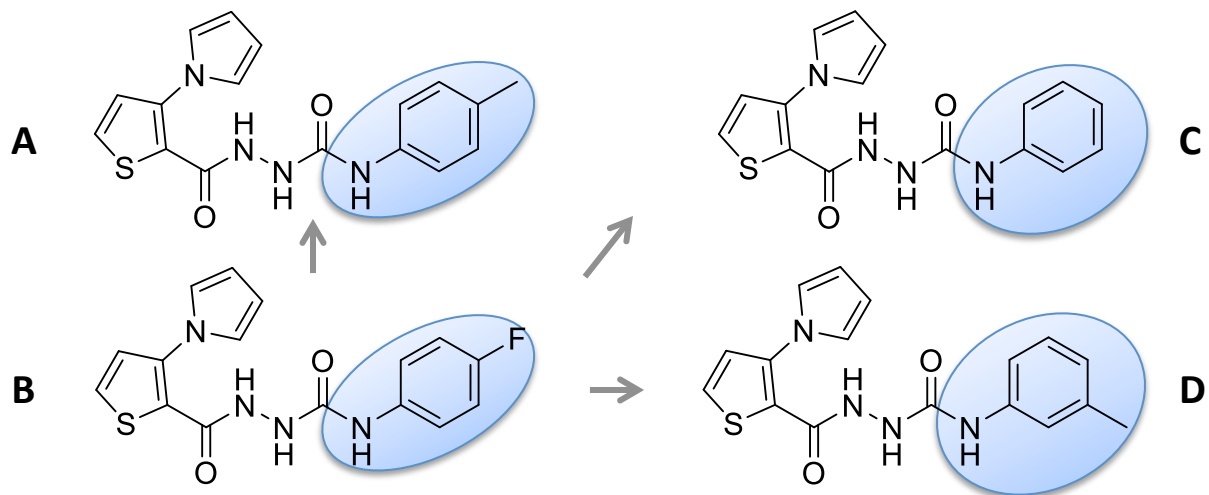
Generation of Analog Series (ASs)

- All pairwise RECAP-MMP relationships for one compound and **all its neighbors**



Generation of Analog Series (ASs)

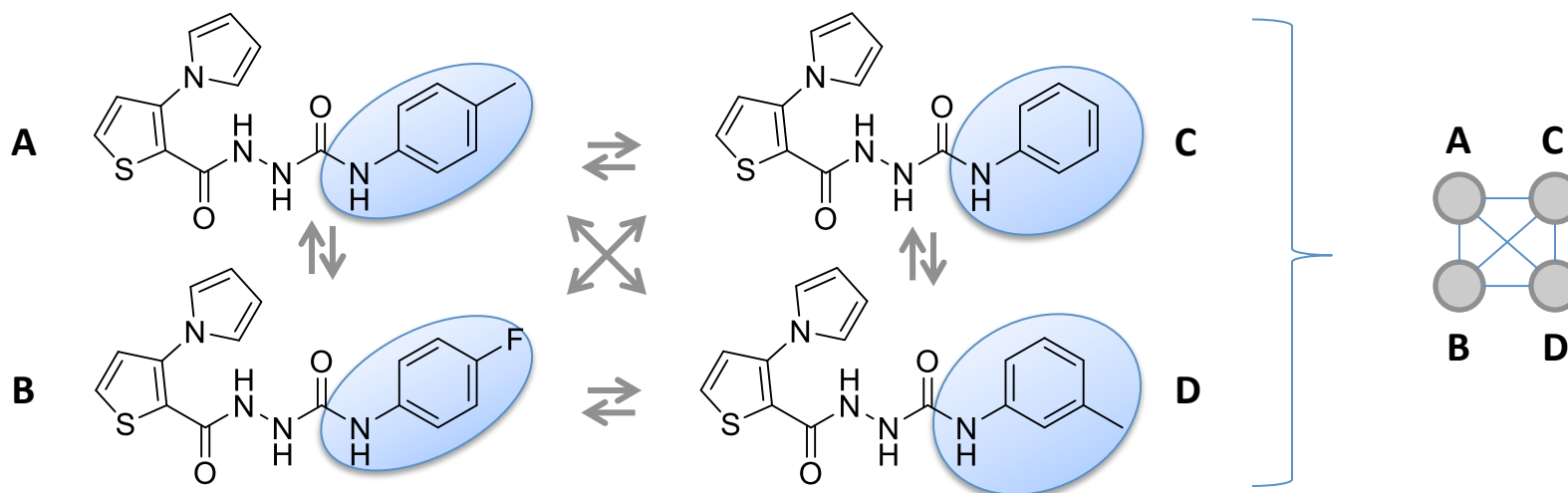
- All pairwise RECAP-MMP relationships for one compound and **all its neighbors**



- Systematic analog search until no additional analog compounds are detected

Generation of Analog Series (ASs)

- All pairwise RECAP-MMP relationships for one compound and all its neighbors

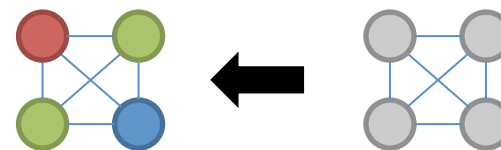


- Systematic analog search until no additional analog compounds are detected

Generation of Analog Series (ASs)

- Classification of analogs within one AS
 - Analogs are classified following their target annotations

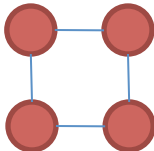
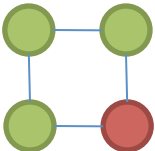
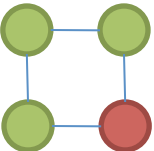
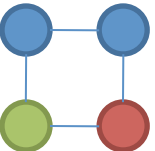
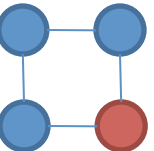
CPD type	# Active target
Inactive	0
Single-target	1
Multi-target	> 1






- Inactive CPD
- Single-target compound (ST-CPD)
- Multi-target compound (MT-CPD)

Classification of Analog Series (ASs)

ASs containing
promiscuous CPDs

AS type	Inactive AS	ST-AS *	MT-AS 1 *	MT-AS 2 *	MT-AS 3 *
AS representation					
Origin of target annotations	NA	ST-CPDs	ST-CPDs	ST-CPDs MT-CPDs	MT-CPDs

* Inactive compounds are not present in every AS

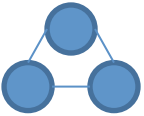
-  Inactive CPD
-  Single-target compound (ST-CPD)
-  Multi-target compound (MT-CPD)

Summary

- 54% of the ASs contained multi-targets annotations based partly on promiscuous compounds
- Can we verify that there is a relation between the structure of the analogs and their activities?
- Can we use ASs to make reliable target predictions?

Assessing the Potential of AS for Target Prediction

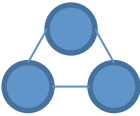
Method for Assessing the Potential of AS for Target Prediction



	T1	T2	T3	T4	T5	T6
CPD 1	Active	Active	Unknown activity	Inactive	Active	Unknown activity
CPD 2	Active	Active	Active	Inactive	Inactive	Unknown activity
CPD 3	Active	Active	Active	Inactive	Inactive	Inactive

 Active  Inactive  Unknown activity

Method for Assessing the Potential of AS for Target Prediction

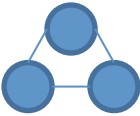


	T1	T2	T3	T4	T5	T6
CPD 1	Active	Active	Unknown activity	Inactive	Active	Unknown activity
CPD 2	Active	Active	Active	Inactive	Inactive	Unknown activity
CPD 3	Active	Active	Active	Inactive	Inactive	Inactive

■ Active ■ Inactive □ Unknown activity

- 1) identify intersecting targets (ITs)
 - ITs are targets on which all the analogs in the series are annotated

Method for Assessing the Potential of AS for Target Prediction



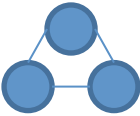
	T1	T2	T3	T4	T5	T6
CPD 1	Active	Active	Unknown activity	Inactive	Active	Unknown activity
CPD 2	Active	Active	Active	Inactive	Inactive	Unknown activity
CPD 3	Active	Active	Active	Inactive	Inactive	Inactive

■ 4 ITs

■ Active ■ Inactive □ Unknown activity

- 1) identify intersecting targets (ITs)
 - ITs are targets on which all the analogs in the series are annotated

Method for Assessing the Potential of AS for Target Prediction



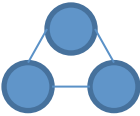
	T1	T2	T3	T4	T5	T6
CPD 1	Active	Active	Unknown activity	Inactive	Active	Unknown activity
CPD 2	Active	Active	Active	Inactive	Inactive	Unknown activity
CPD 3	Active	Active	Active	Inactive	Inactive	Inactive

■ 4 ITs

■ Active ■ Inactive □ Unknown activity

- 1) identify intersecting targets (ITs)
- 2) assess the activity consistency
 - Identify ITs with consistent activity amongst all the compounds in each series

Method for Assessing the Potential of AS for Target Prediction



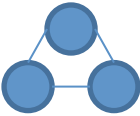
	T1	T2	T3	T4	T5	T6
CPD 1	Active	Active	Unknown activity	Inactive	Active	Unknown activity
CPD 2	Active	Active	Active	Inactive	Inactive	Unknown activity
CPD 3	Active	Active	Active	Inactive	Inactive	Inactive

■ Active ■ Inactive □ Unknown activity

- 4 ITs
- 3 ITs with consistent activity

- 1) identify intersecting targets (ITs)
- 2) assess the activity consistency
 - Identify ITs with consistent activity amongst all the compounds in each series

Method for Assessing the Potential of AS for Target Prediction



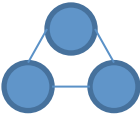
	T1	T2	T3	T4	T5	T6
CPD 1	Active	Active	Unknown activity	Inactive	Active	Unknown activity
CPD 2	Active	Active	Active	Inactive	Inactive	Unknown activity
CPD 3	Active	Active	Active	Inactive	Inactive	Inactive

■ Active ■ Inactive □ Unknown activity

- 4 ITs
- 3 ITs with consistent activity

- 1) identify intersecting targets (ITs)
- 2) assess the activity consistency
- 3) derive new target hypotheses for analogs

Method for Assessing the Potential of AS for Target Prediction



	T1	T2	T3	T4	T5	T6
CPD 1	Active	Active	Unknown activity	Inactive	Active	Unknown activity
CPD 2	Active	Active	Active	Inactive	Inactive	Unknown activity
CPD 3	Active	Active	Active	Inactive	Inactive	Inactive

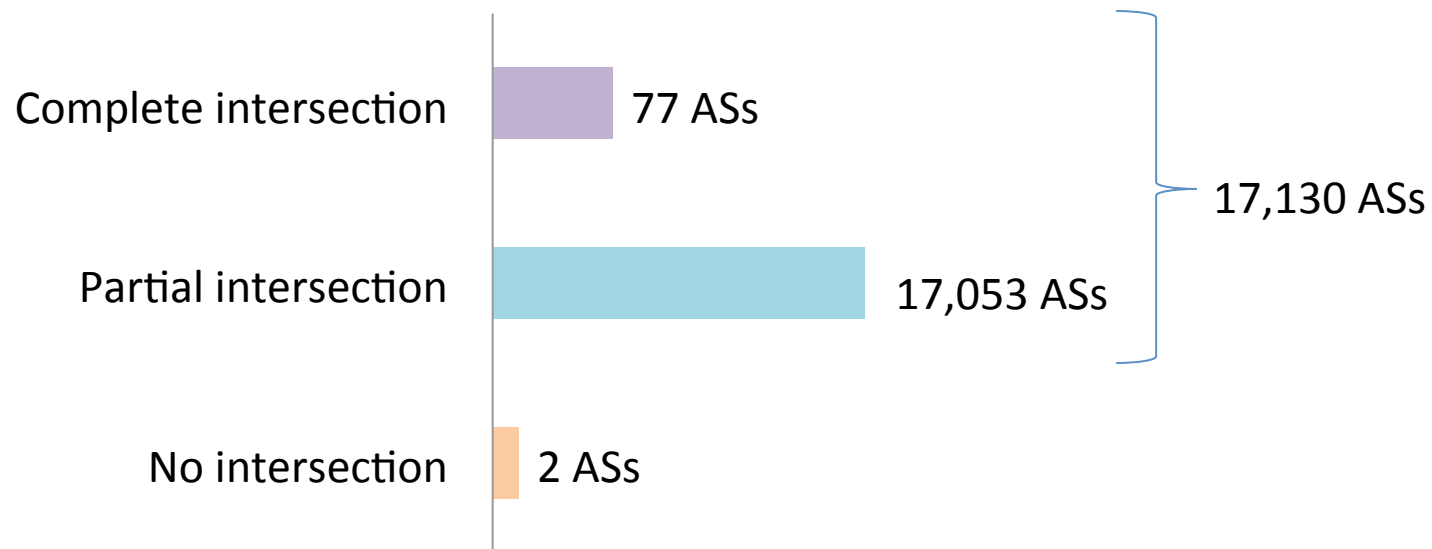
■ Active ■ Inactive □ Unknown activity

- 4 ITs
- 3 ITs with consistent activity
- 3 predictions of compound-target interactions

- 1) identify intersecting targets (ITs)
- 2) assess the activity consistency
- 3) derive new target hypotheses for analogs

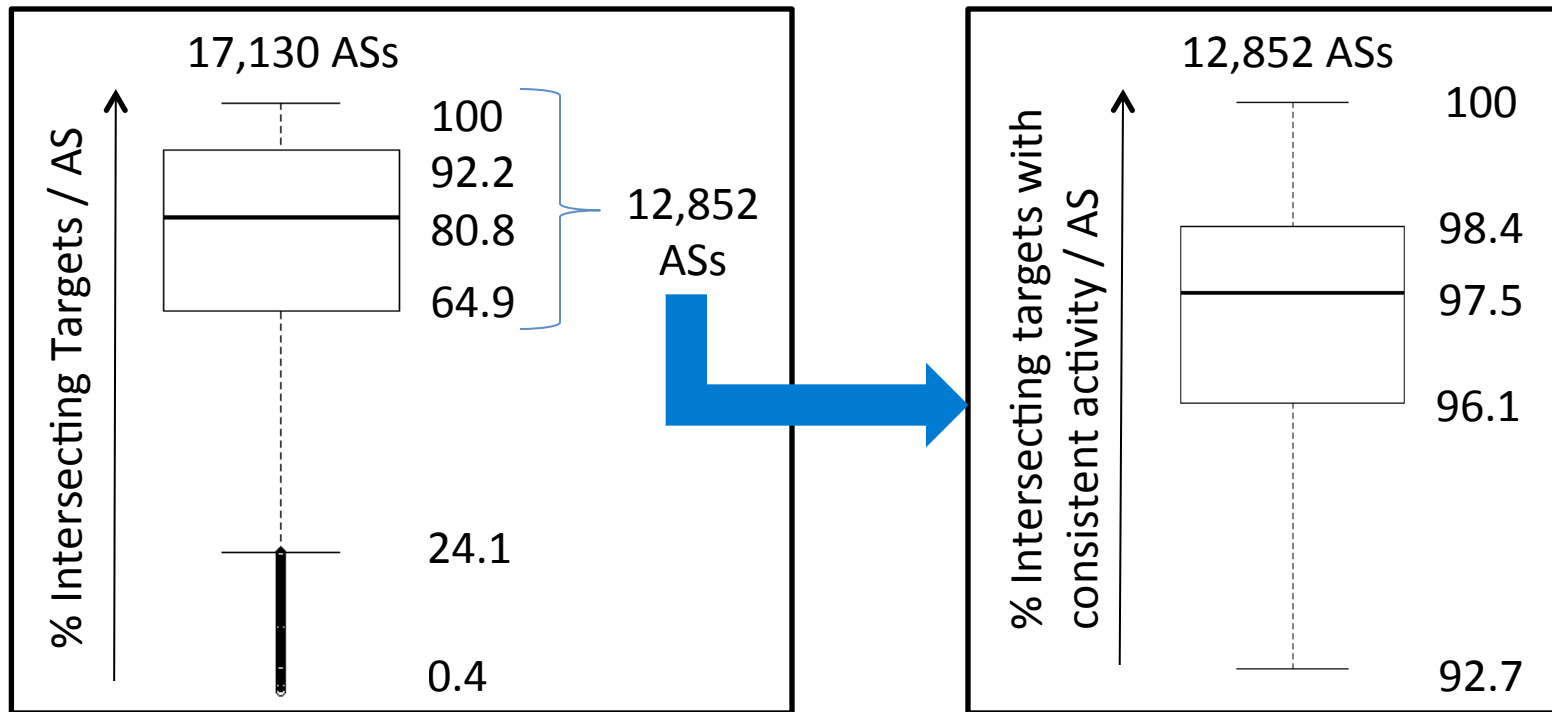
Results

Intersection of Targets in ASs



- For partial intersection, is the number of ITs significant enough compared to the total number of targets?

ASs with Intersecting Targets



- ASs with low distribution of ITs were discarded
- 12,852 ASs with high number of ITs and an average of 97% ITs with consistent activity

Conclusion

Conclusion and Future work

- We investigated the potential of **analog series** containing promiscuous compounds for target identification
- 75% of ASs contained a significant number of intersecting targets with high activity consistency
- There is a correlation between the structure and the activity of the analogs
- ASs containing promiscuous CPDs are **reliable tools for target predictions**
- New target hypotheses will be suggested for analogs

References

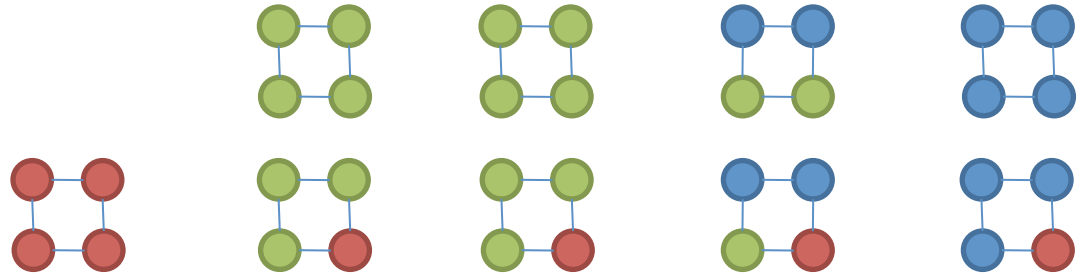
- Kunitomo R, Dimova D & Bajorath J. Application of a new scaffold concept for computational target deconvolution of chemical cancer cell line screens. *ACS Omega* 2, 1463-1468, 2017.
- Keiser MJ, Setola V, Irwin JJ *et al.* Predicting new molecular targets for known drugs. *Nature* 462(7270), 175–181 (2009).
- Jenkins, J. L.; Bender, A.; Davies, J. W. In Silico Target Fishing: Predicting Biological Targets from Chemical Structure. *Drug Discovery Today: Technol.* 2006, 3, 413–421.
- Hu Y, Lounkine E & Bajorath J. Many approved drugs have bioactive analogs with different target annotations. *AAPS J* 16, 847-859, 2014.
- Jasial S, Hu Y & Bajorath J. Determining the degree of promiscuity of extensively assayed compounds. *PLoS One*, 11, e0153873, 2016.
- Gilberg E, Stumpfe D & Bajorath J. Activity profiles of analog series containing pan assay interference compounds. *RSC Adv* 7, 35638-35649, 2017.
- Stumpfe D, Dimova D & Bajorath J. Computational method for the systematic identification of analog series and key compounds representing series and their biological activity profiles. *J Med Chem* 59, 7667-7676, 2016.

Acknowledgment




Dr. Dagmar Stumpfe
Prof. Dr. Jürgen Bajorath


Appendixes

Distribution of Analog Series (ASs)



		Inactive AS	ST-AS	MT-AS 1	MT-AS 2	MT-AS 3
# ASs		5,521	6,211	3,005	9,576	7,556
# CPDs	Inactive CPDs	13,620	10,384	6,757	45,704	6,868
	ST-CPDs	0	6,697	7,363	31,501	0
	MT-CPDs	0	0	0	30,804	14,587
	All	13,620	17,081	14,120	108,009	21,456

-  Active MT-CPD
-  Active ST-CPD
-  Inactive CPD


17,132 ASs

RECAP rules

