StARs and constellations in chemical space:

A visual representation of Structure-Activity Relationships

José L. Medina-Franco

DIFACQUIM research group

Department of Pharmacy, School of Chemistry
National Autonomous University of Mexico (UNAM)
medinajl@unam.mx
www.difaquim.com

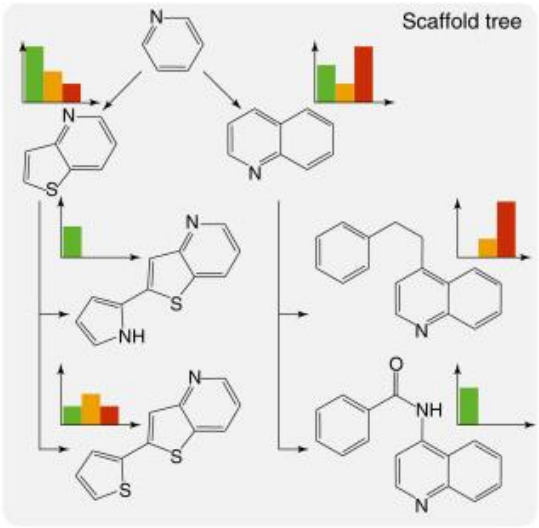
Chemoinformatics Strasbourg Summer School
University of Strasbourg, 29 June - 3 July 2020

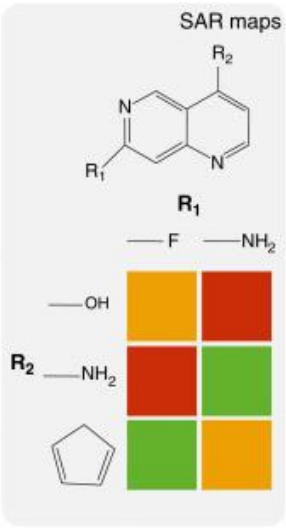
Outline

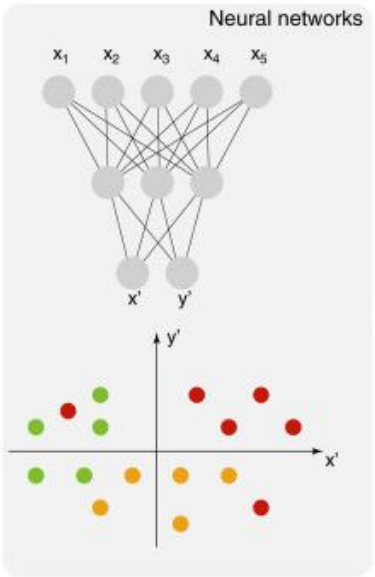
- Background
  - Methods for SAR analysis and visualization
  - Chemical space and activity landscape modeling
- Motivation
- Activity Landscape Plotter
- Constellation Plots
- Concluding remarks and perspectives

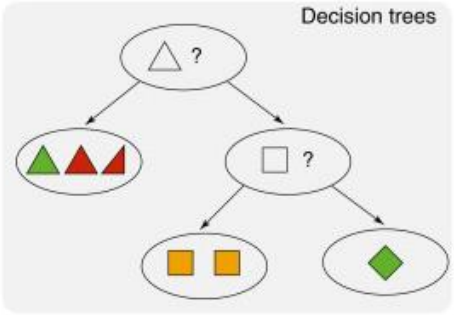
General approaches for SAR analysis

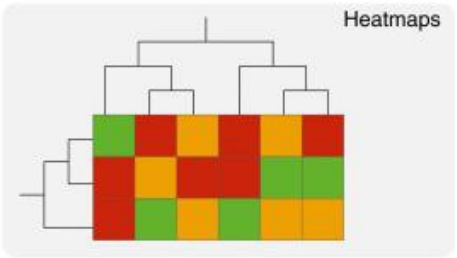
- Dimension reduction.
- Clustering and partitioning.
- Organization and annotation of substructures.
- Structural vs. activity similarity.

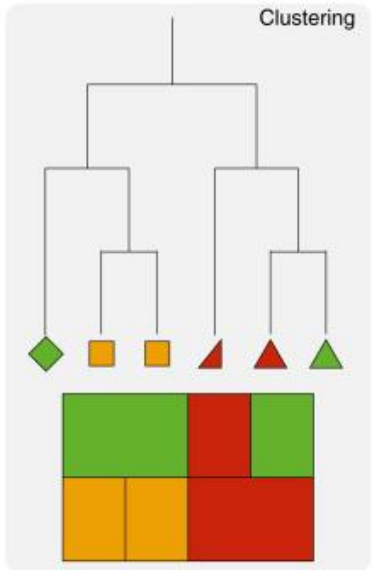


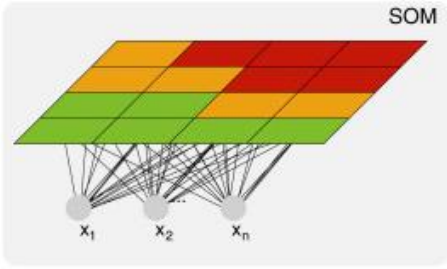


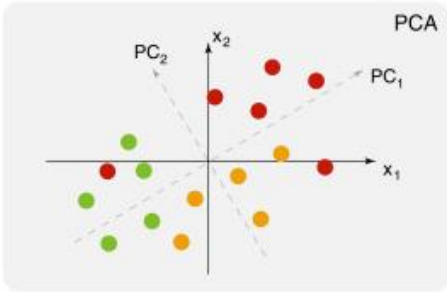












Wawer et al. *Drug Discovery Today* 2010 15:630

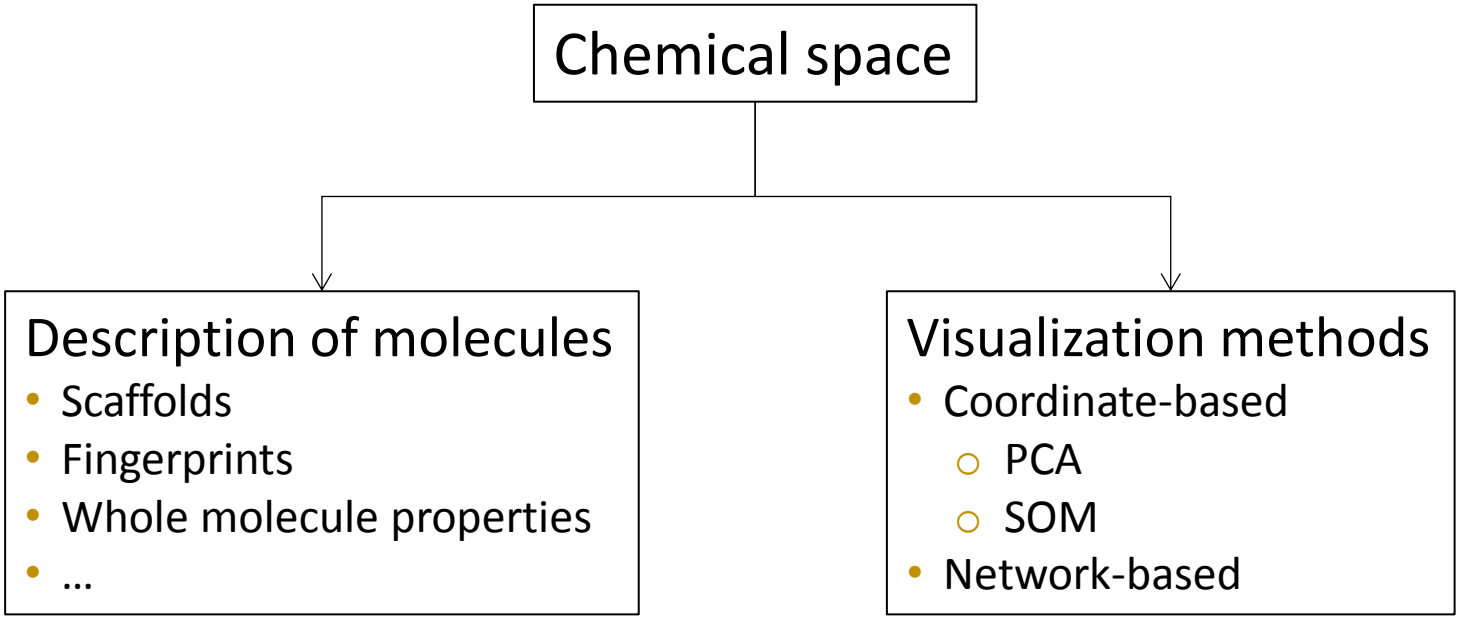
Implementations of methods for SAR

Method	Ref.
SAR matrices (SARMs)	ACS Omega 4:7061
SAR maps and Enhanced SAR maps	J Med Chem 50:5926
SAR Analyzer	J Cheminform 5:31
Consensus Diversity Plots	J Cheminform. 8:63
Chem Maps	JCC 3:157
Shinyheatmap	PLoS ONE. 12:e0176334

Software	Ref.
DataWarrior	Exp Op on Drug Disc 14:335
Tableau	Tableau Soft. WA, USA
SARANEA	JCIM 50:16
Scaffold Hunter	J Cheminf 9 28
Spotfire	TIBCO Soft. CA, USA
Molecular Property eXplorer	JCIM 45: 523
LeadScope	JCICS 40:6
StarDrop	Optibrium. CA,UK
Miner3D	Addinsoft. PA, FR
SAR Report	MOE Soft. MO, CAN
Web-based applications	
VisualiSAR	JMGM 17:85
MOESaic	MOE Soft. MO, CAN

Chemical space

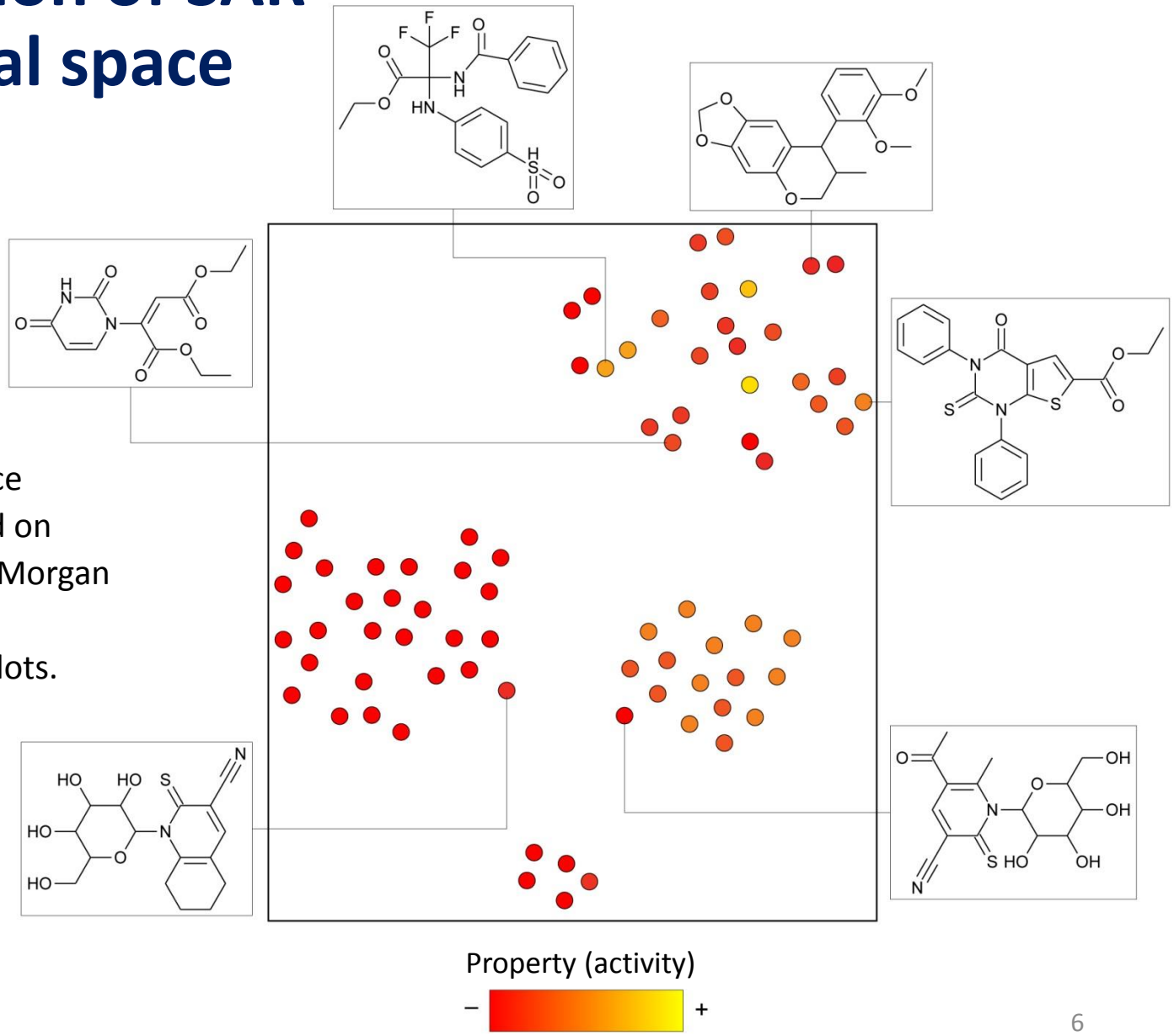
One of many definitions: 'an M-dimensional Cartesian space in which compounds are located by a set of M physicochemical and/or chemoinformatic descriptors'.



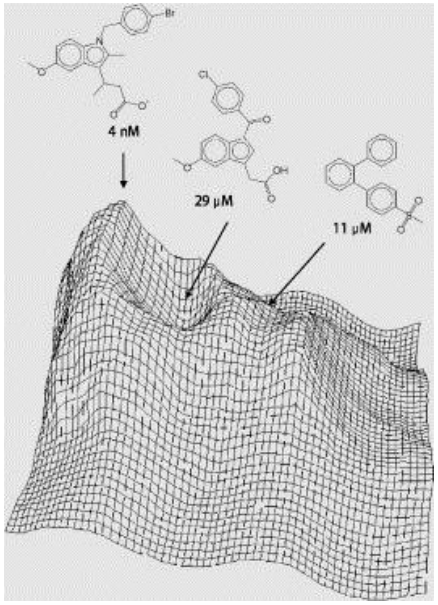
Virshup AM et al. *J Am Chem Soc* 2013 135:7296

Visualization of SAR in chemical space

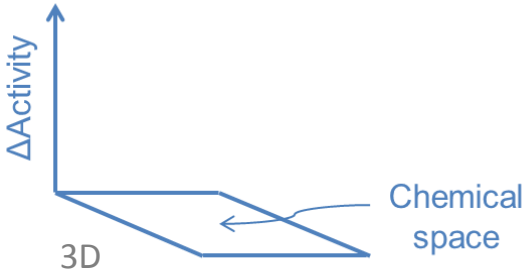
- Typical chemical space representation based on coordinates (t-SNE / Morgan fingerprints).
- Single molecules as dots.

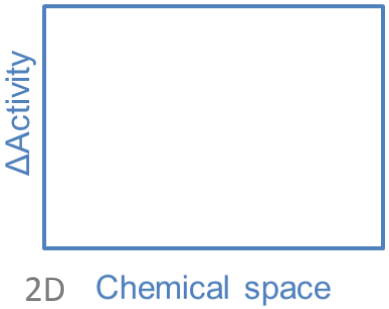


Activity landscape modeling



"Any representation that integrates the analysis of the structure similarity and potency differences between compounds sharing the same biological activity."





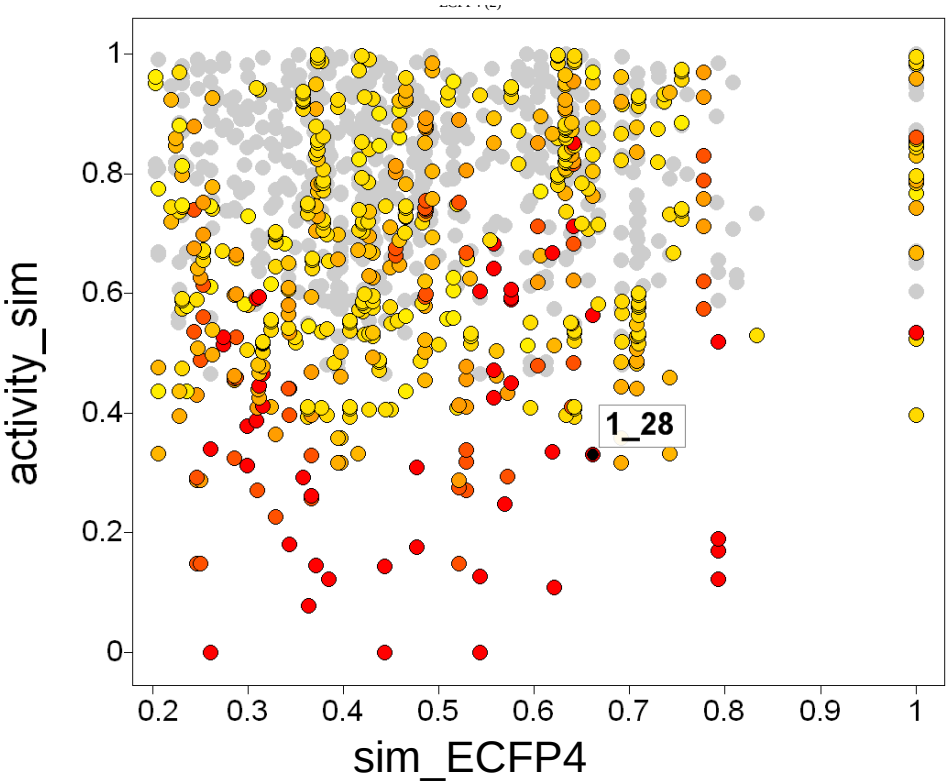
Several methods available:*

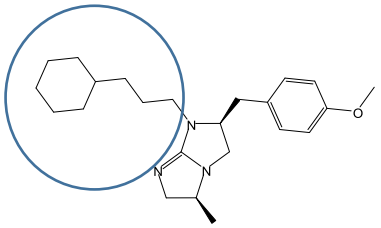
- SALI: Structure-Activity Selectivity Index
- SAS: Structure-Activity Similarity maps
- SARANEA …

*Stumpfe D. Bajorath J *J Med Chem* 2014 57:18

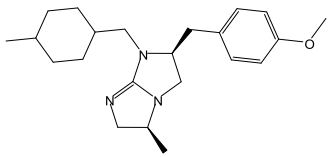
*Iqbal J. Vogt M, Bajorath J. *J. Cheminformatics* 2020 12:34

Structure-Activity Similarity (SAS) Map



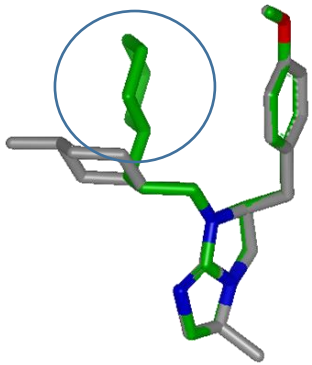


1: 37 nM



28: 1,568 nM

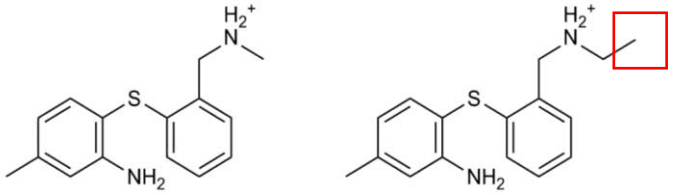
Activity cliff: Similar structures, very different activity



Shanmugasundaram V, Maggiora GM. CINF-032. 222 ACS National Meeting, Chicago, IL, Aug 26–30, 2001

Dual Activity Difference (DAD) Map

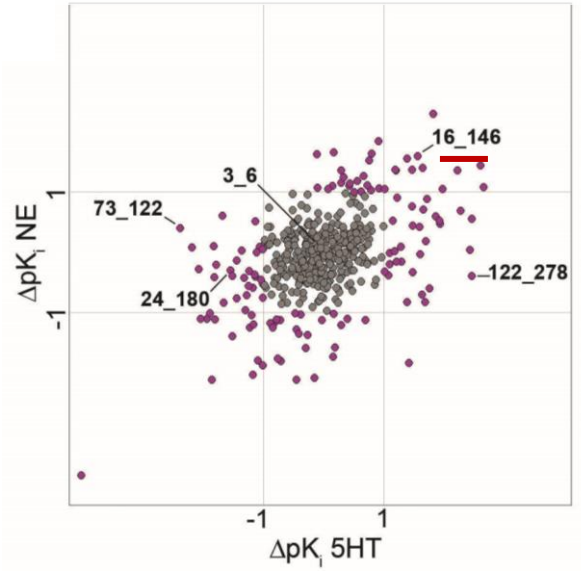
Dual activity cliff with similar SAR



16 **146**

DA: 698 nM
NE: 191 nM
5HT: 1.94 nM

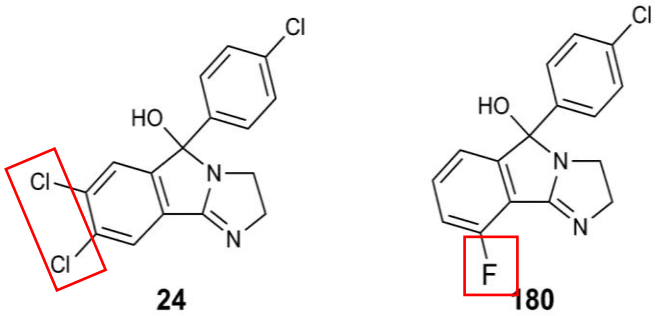
DA: 2121 nM
NE: 7552 nM
5HT: 70.2 nM



Activity cliffs
Continuous SAR

Dual cliff with opposite SAR

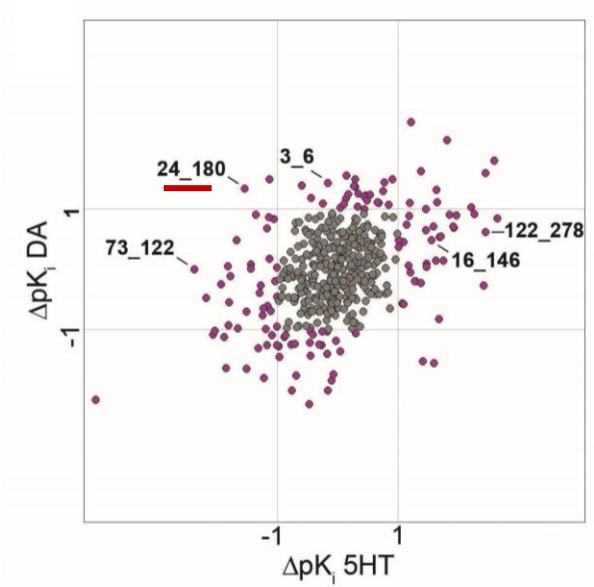
Selectivity switch



24 **180**

DA: 1.05 nM
NE: 10 nM
5HT: 1347 nM

DA: 23.1 nM
NE: 5 nM
5HT: 38.1 nM



Medina-Franco JL. *J Chem Inf Model* 2012 52:2485

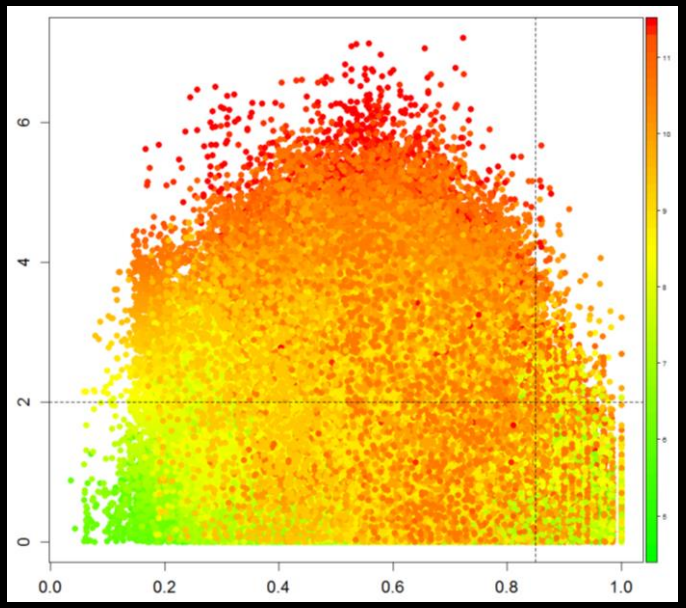
Motivation

- SPR/SAR is a fundamental practice in chemistry.
- Data visualization enhance interpretation and pattern recognition.
  - Not necessarily prediction.

➢ Increase web-based services.

➢ Combine scaffold- with coordinate-based chemical representations.

Arkadii Lin, A, Baskin, IB, Marcou G, Horvath D, Beck B, Varnek A. *Mol Informatics* 2020, in press



Activity Landscape Plotter

A web-based implementation of SAS and DAD maps for SAR analysis and rapid detection of activity cliffs

Activity Landscape Plotter

www.difacquim.com/d-tools/

Activity Landscape Plotter V.1

About | Instructions | Template | SAS map | DAD map | Contact | Acknowledgements | D-Tools

DIFACQUIM

Welcome to this first online Activity Landscape Plotter. Here you will be able to perform analyses of activity landscape using Structure-Activity Similarity (SAS) maps, Dual Activity-Difference (DAD) maps and other metrics.

If you use this app please cite this reference:

Gonzalez-Medina M, Mendez-Lucio O, Medina-Franco JL. Activity Landscape Plotter: A Web-based Application for the Analysis of Structure-Activity Relationships. *J Chem Inf Model* (2017) 57(3), 397-402.

This App was developed by Mariana Gonzalez-Medina Oscar Mendez-Lucio and Jose L. Medina-Franco members of DIFACQUIM The group is based in the Pharmacy Department of the School of Chemistry in Universidad Nacional Autonoma de Mexico.

DIFACQUIM UNAM

What are they for?

Structure-Activity Similarity (SAS) maps.

SAS maps were introuced to find a relationship between structure and activity, based on a systematic pairwise comparison of all the compounds in a data set. You can find more information here.

With this App you can automatically obtain all the information required to generate and analyze SAS maps using your own activity data for one or more biological targets. You can customize the SAS maps and download the raw data as described in the Instructions.

Dual Activity-Difference (DAD) maps.

DAD maps depict pairwise activity differences for each possible pair of compounds in a data set against two targets. These maps are helpful to differentiate when a strucutural modification increases or decreases the activity for one target or the other. You can find more information here.

With this App you can automatically obtain all the information required to plot and analyze DAD maps, using the biological activities in your input file. You can customize the DAD maps and download the raw data as described in the Instructions.

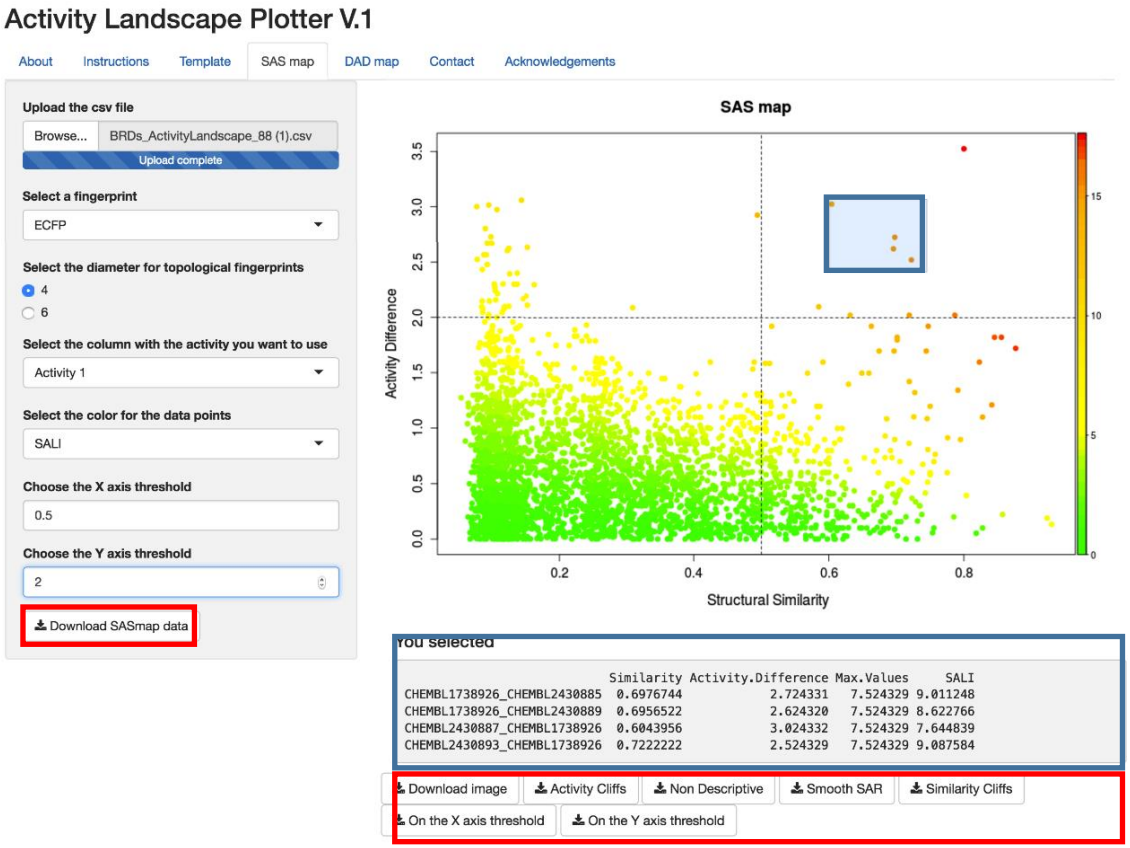
Mariana González

Shiny → User interface

Rcdk → Chemoinformatics

González-Medina M et al. *J Chem Inf Model* 2017 57:397

Activity Landscape Plotter
SAS maps



The user can choose between:

Three **molecular fingerprints**:

- ECFP4 and ECFP6
- MACCS keys 166 bits
- PubChem 881 bits

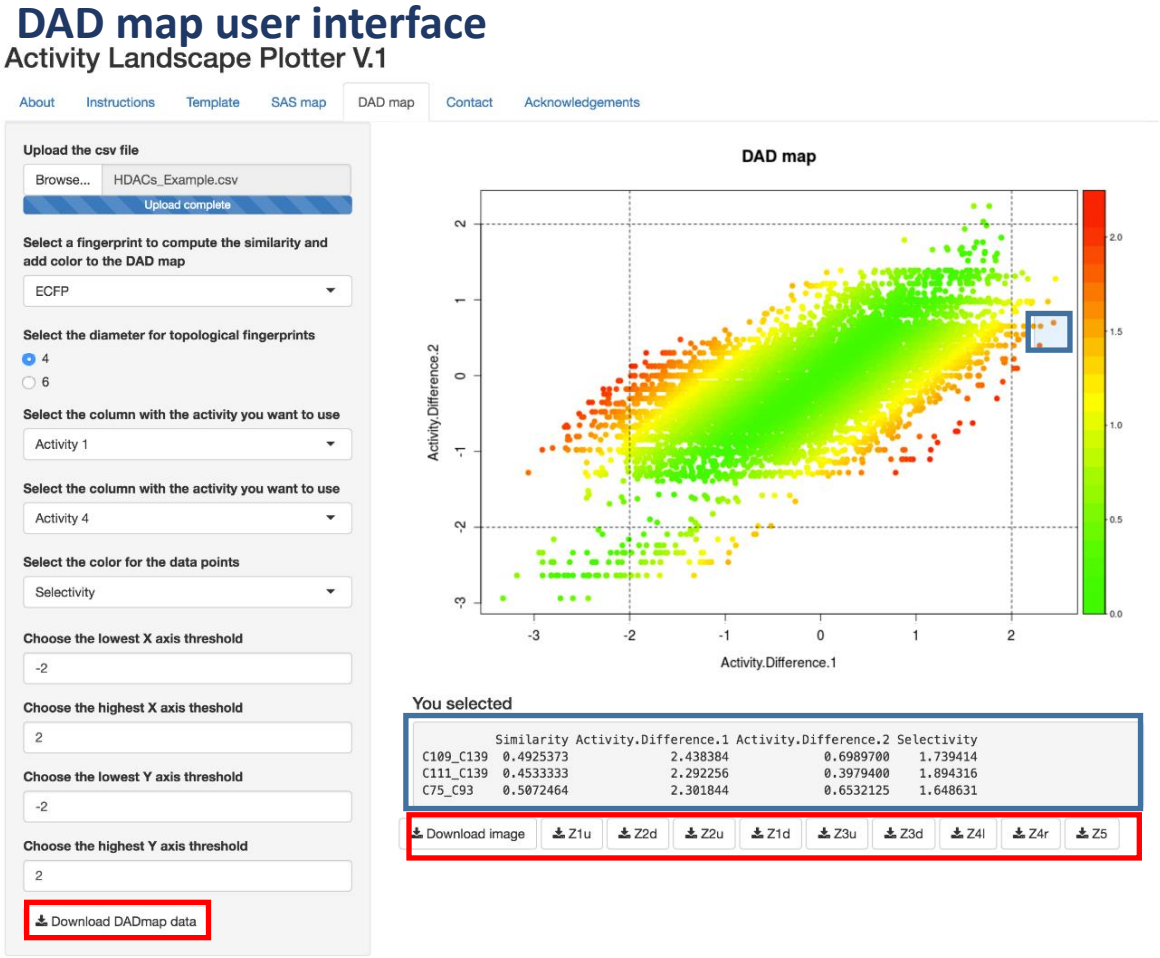
Different activities from their input file

Set the x and y thresholds

Download raw data

Activity Landscape Plotter
DAD maps

DAD map user interface



The user can choose between:

Different activities from their input file

Set the x and y thresholds

Download raw data

González-Medina M et al. *J Chem Inf Model* 2017 57:397–402

Other free web services

D-TOOLS

DIFACQUIM Tools for Cheminformatics

D-Peptide Builder

A chemoinformatic tool to enumerate combinatorial libraries

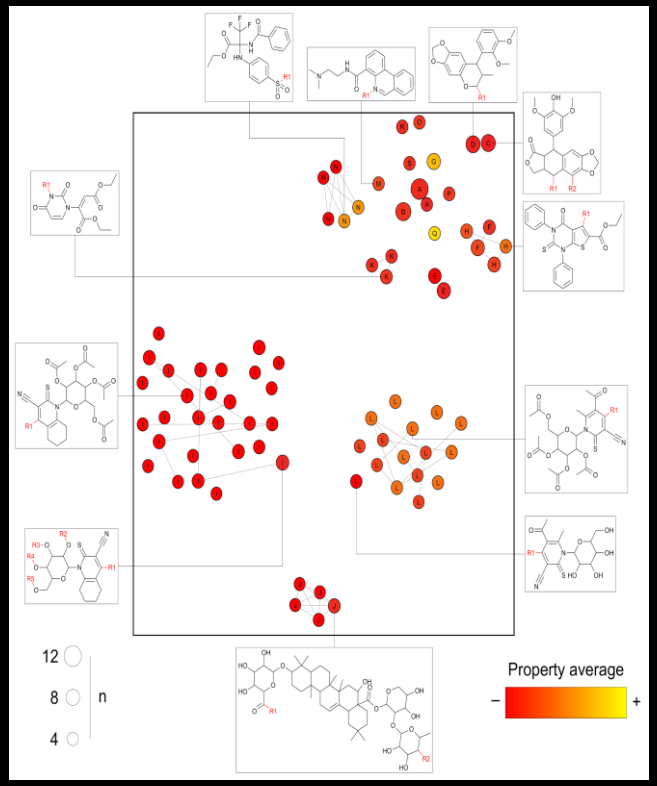
Díaz-Eufracio BI, Palomino-Hernández O, Arredondo-Sánchez A, Medina-Franco JL. D-Peptide Builder: A Web Service to Enumerate, Analyze, and Visualize the Chemical Space of Combinatorial Peptide Libraries. *Molecular Informatics* 2020, in press, doi: 10.1002/minf.202000035

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Platform for Unified Molecular Analysis: PUMA

Mariana González-Medina* and José L. Medina-Franco*

www.difacquim.com/d-tools/

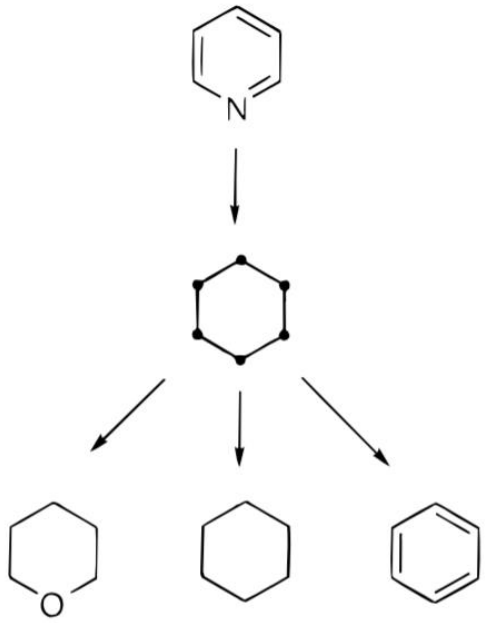


Constellation Plots

A combination of scaffold-based and coordinate-based representation of chemical space for SAR analysis

Scaffold: A simple concept

- Bemis and Murcko's: simple and useful representation for scaffolds based on ring identification.
- Concept broadly applied in multiple diversity and SAR analyses.



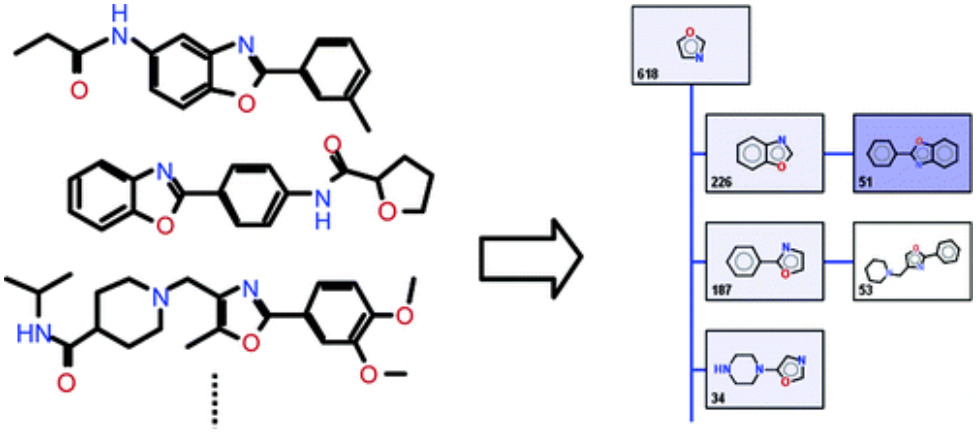
Graph representation of molecules

Bemis and Murcko. *J. Chem. Inf. Model.* 1996 39:2887

Scaffold: A simple concept?

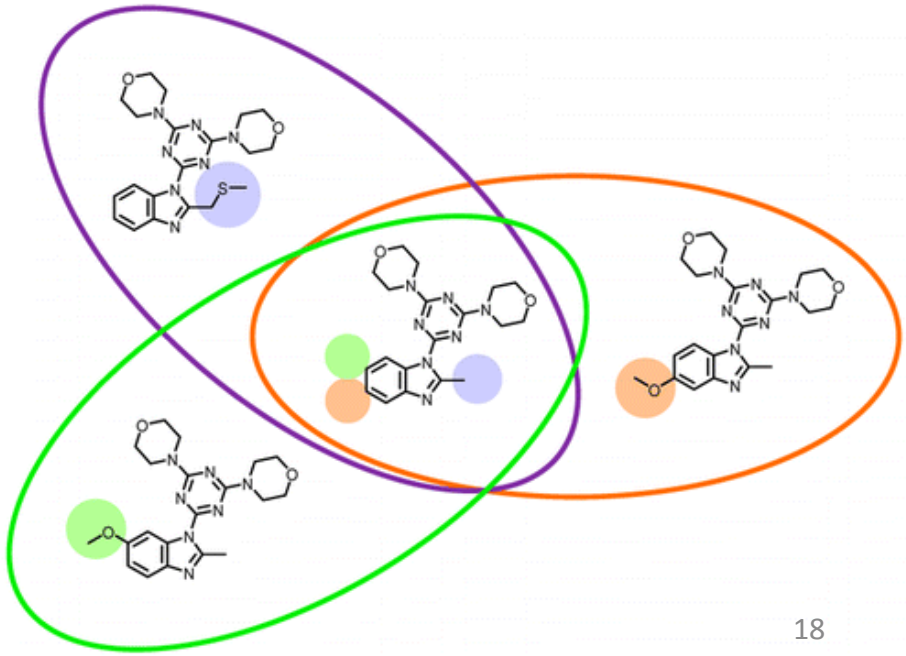
Scaffold trees

A hierarchical scaffold classification



Schuffenhauer et al. *J. Chem. Inf. Model.* 2007:47 47

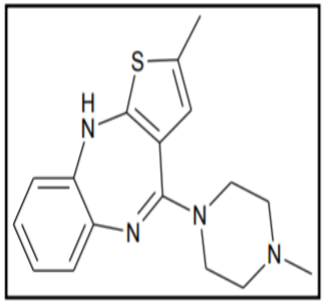
Analog series based scaffolds



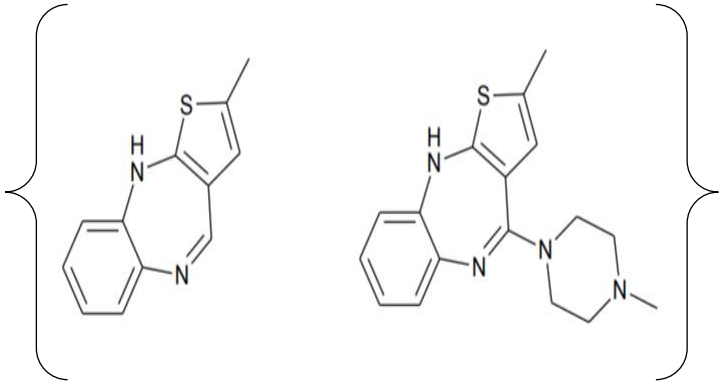
Stumpfe et al. *J. Med. Chem.* 2016:59 7667

Finding putative cores for a single molecule

Initial molecule



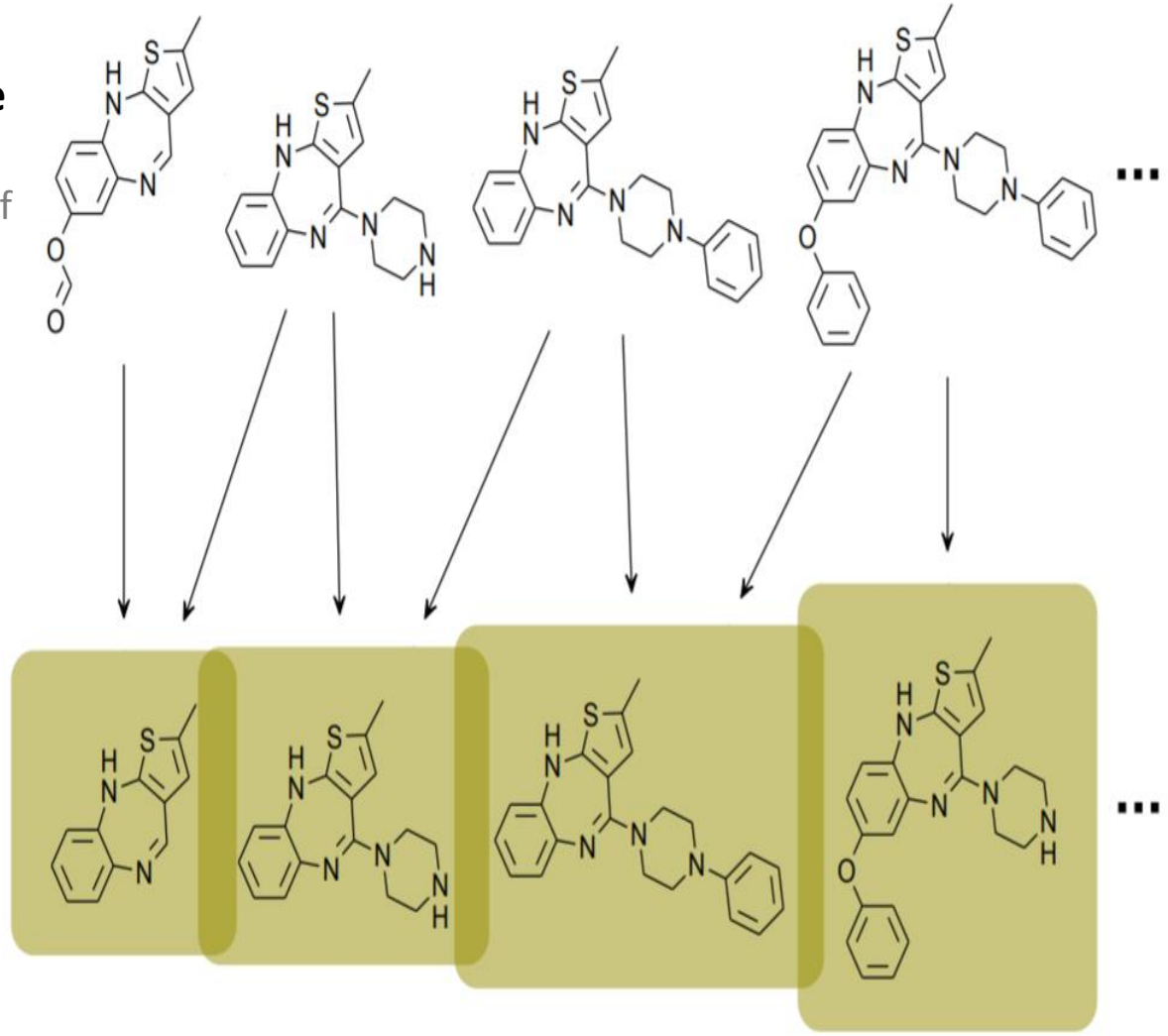
Putative cores



Definition:
Retrosynthesis fragments with at least 2/3 of the size of the initial molecule.

Finding putative cores for a library

Molecule space
example of analogue series



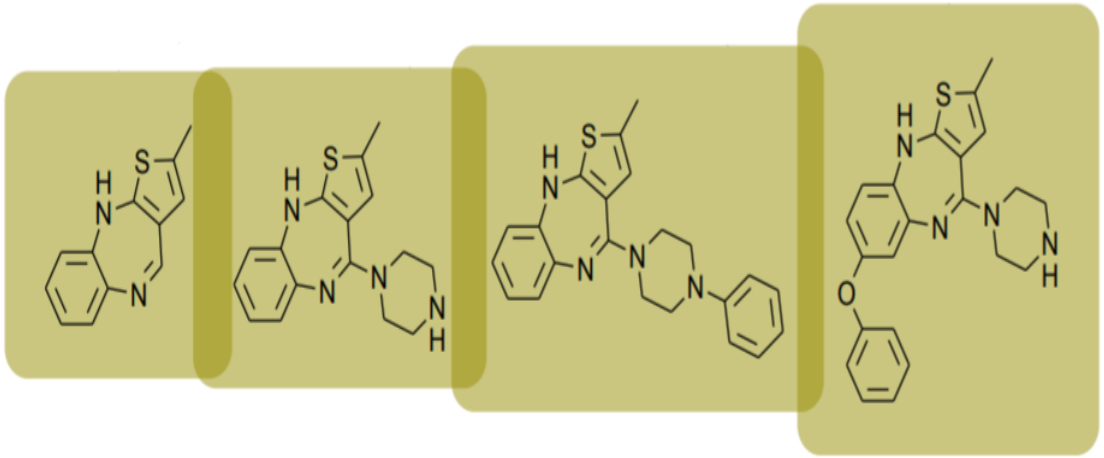
Core space

Each compound is independently processed to find its putative cores.

Analogs: molecules mapping to the same core.

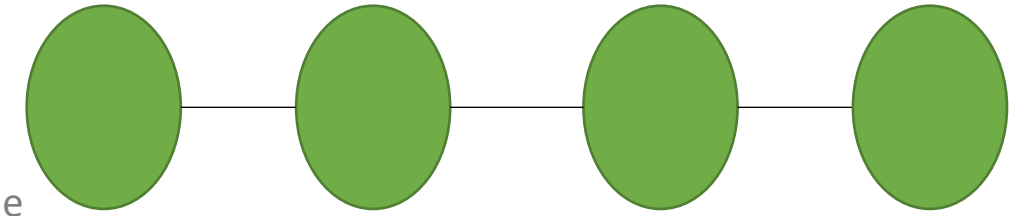
Analog series: a collection of molecules directly or indirectly connected through cores.

Cores as a summary of a series



An **analog series** can be summarized as their cores

Only cores mapping to more than one molecule are considered

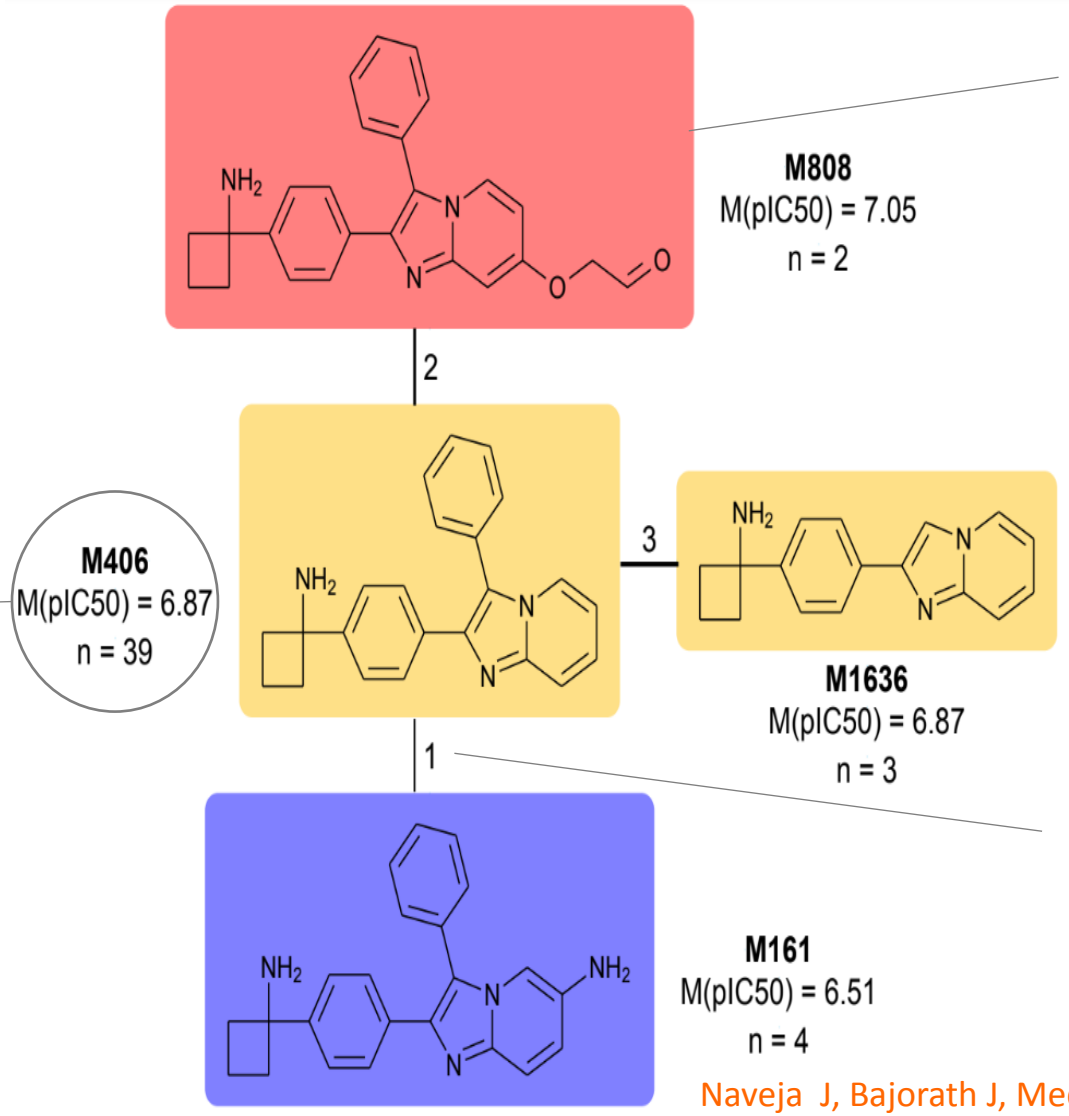


Every core represents a collection of compounds

It is a substructure summary of them

Core SAR (CSAR)

An analog series of AKT1 inhibitors



Cores are colored by the potency mean (median)

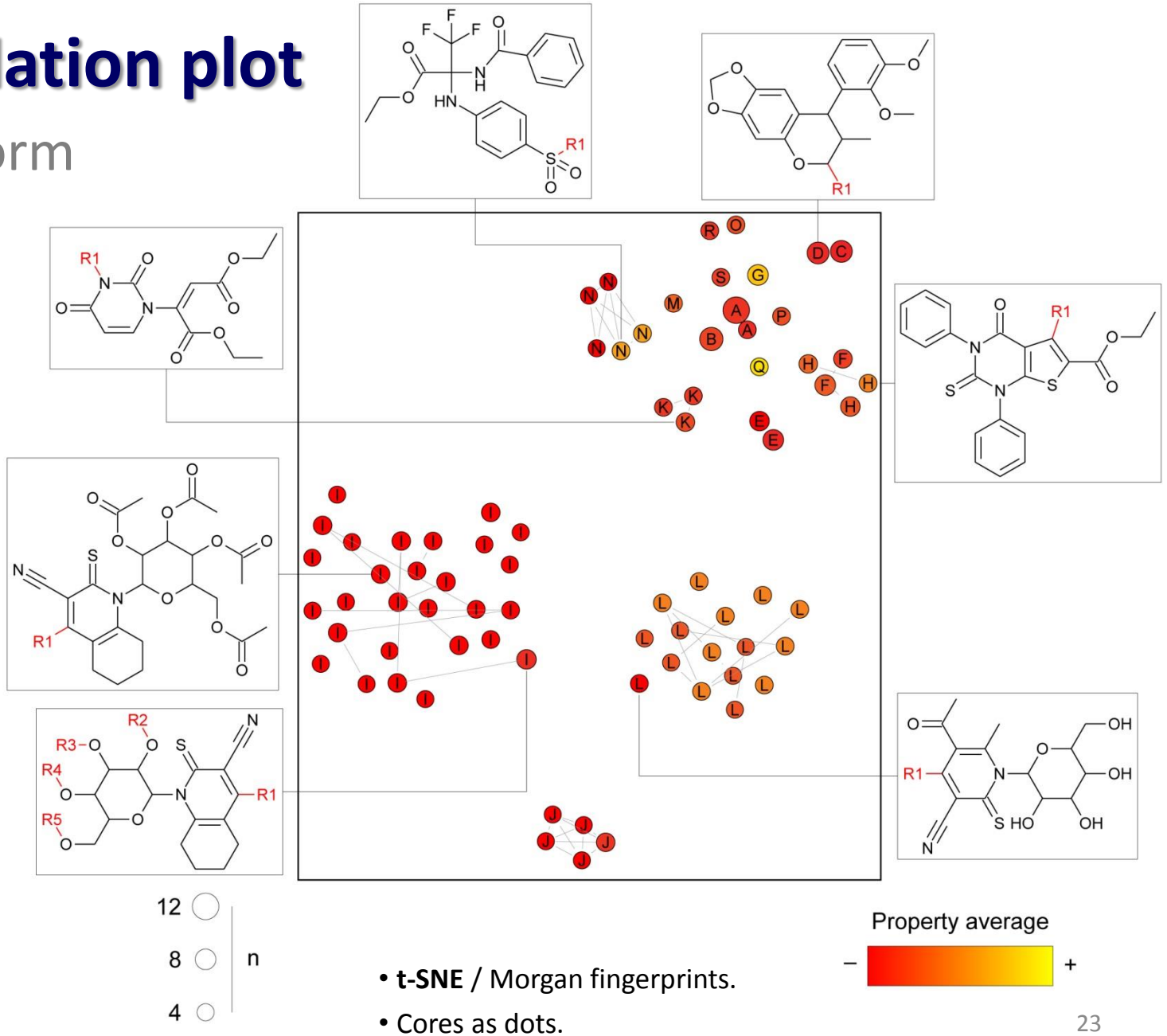
Core ID, potency mean (median) of its molecules, and the number of molecules mapped

The number of molecules mapping to both cores is shown

Naveja J, Bajorath J, Medina-Franco, JL. *J. Cheminforrmatics* 2019 11:61

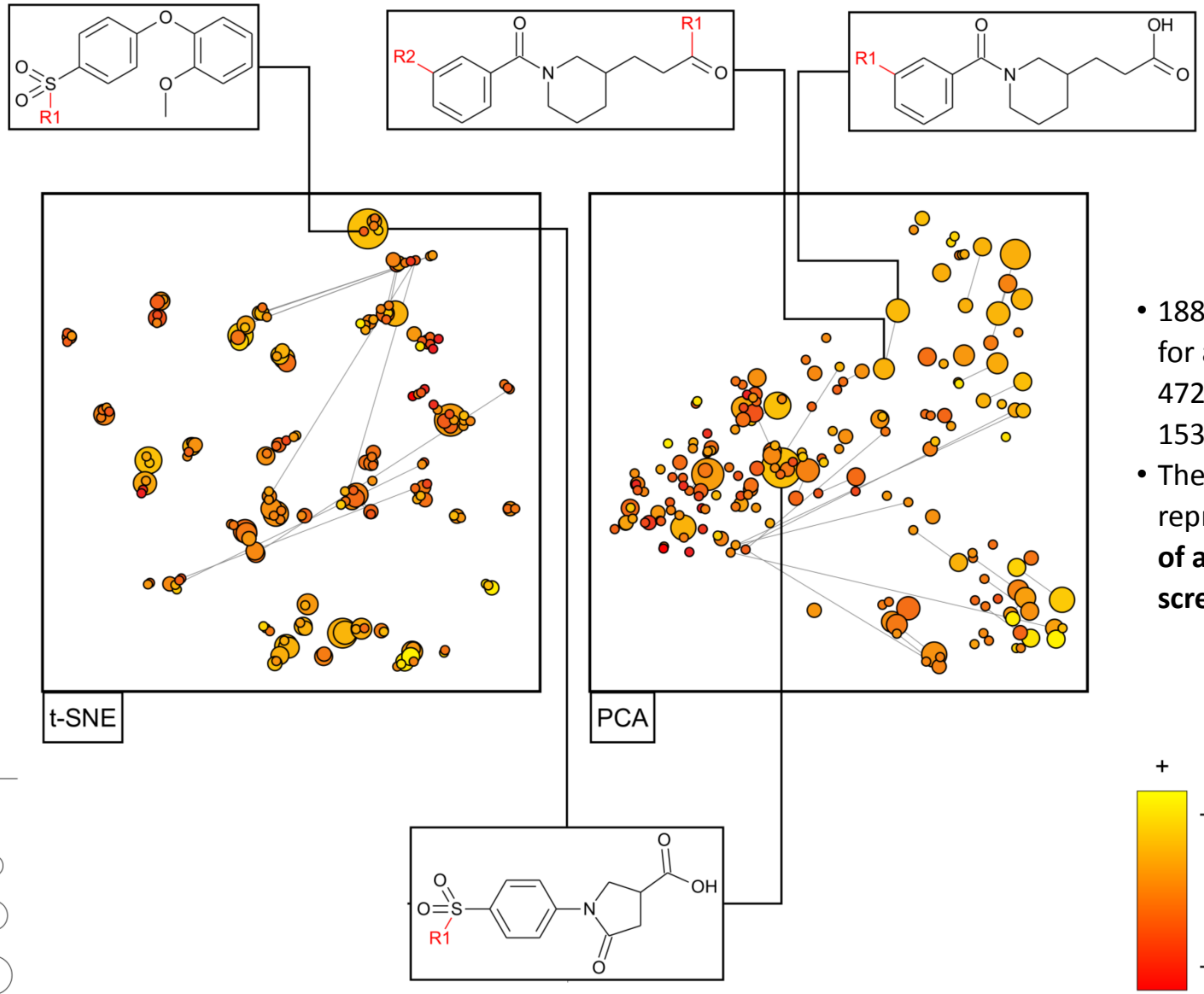
Constellation plot

General form



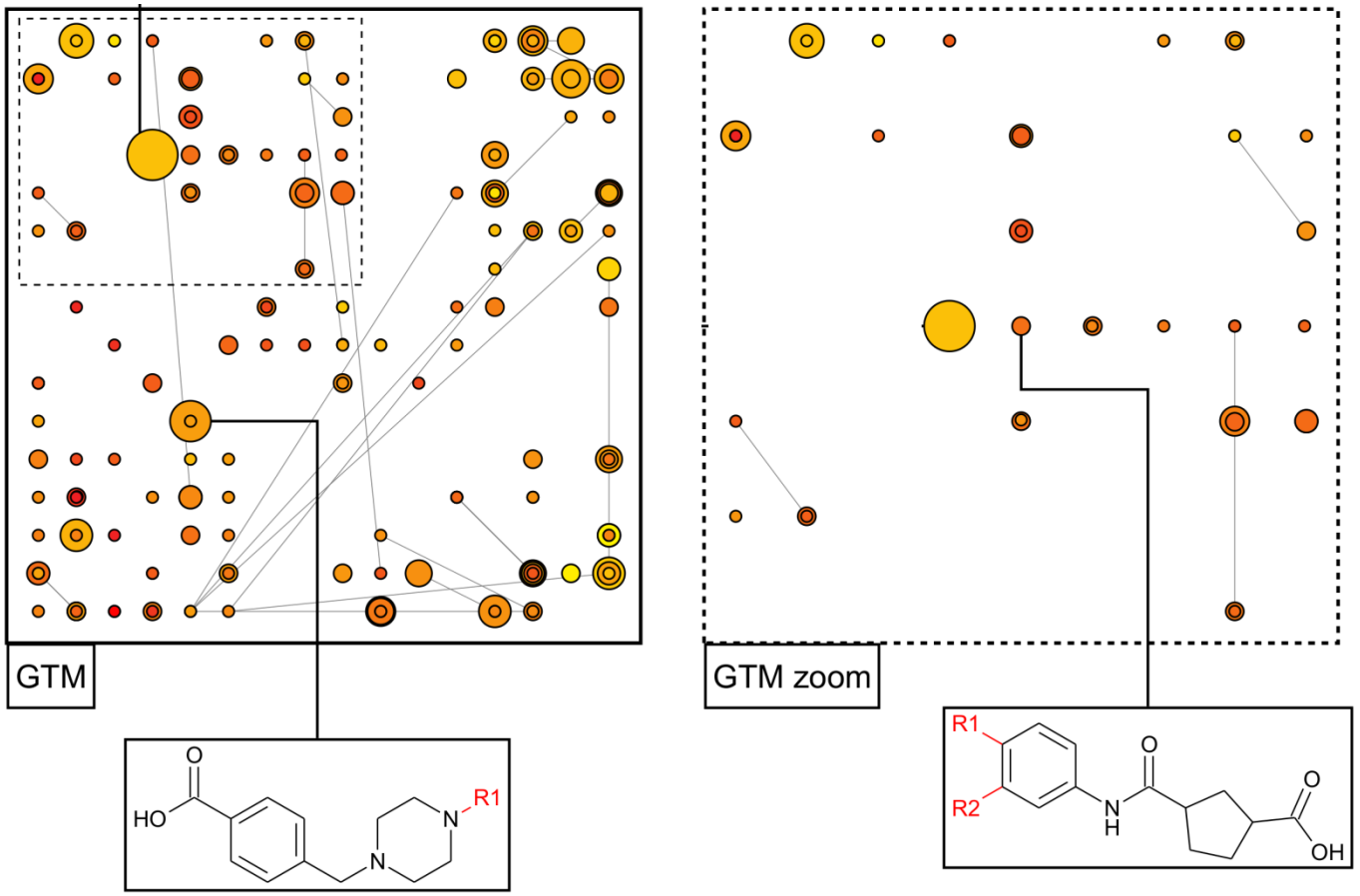
- **t-SNE** / Morgan fingerprints.
- Cores as dots.

Constellation plot using t-SNE and PCA



- 188 cores found for a data set of 472 molecules in 153 analog series.
- The color represent **scores of a virtual screening**.

Constellation plot using GTM



- 188 cores found for a data set of 472 molecules in 153 analog series.
- The color represent **scores of a virtual screening**.

Naveja J, Medina-Franco, JL. *Drug Discovery Today* 2019 11:2162

REVIEWS

Drug Discovery Today • Volume 24, Number 11 • November 2019

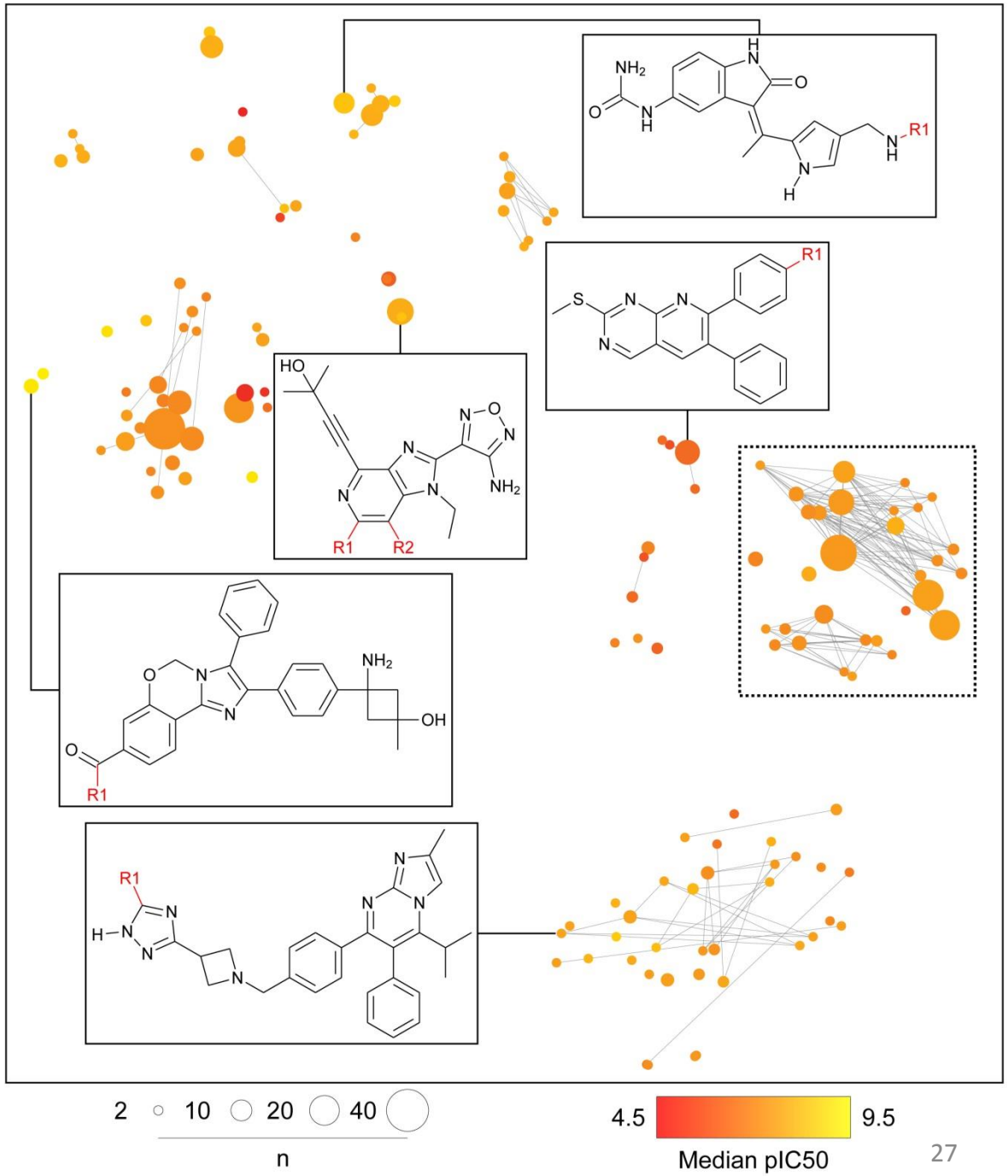
Reviews • INFORMATICS

Reaching for the bright StARs in chemical space

José L. Medina-Franco[1], J. Jesús Naveja[1,2] and Edgar López-López[1]

[1] Department of Pharmacy, School of Chemistry, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Mexico City 04510, Mexico
[2] PECEM, School of Medicine, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico

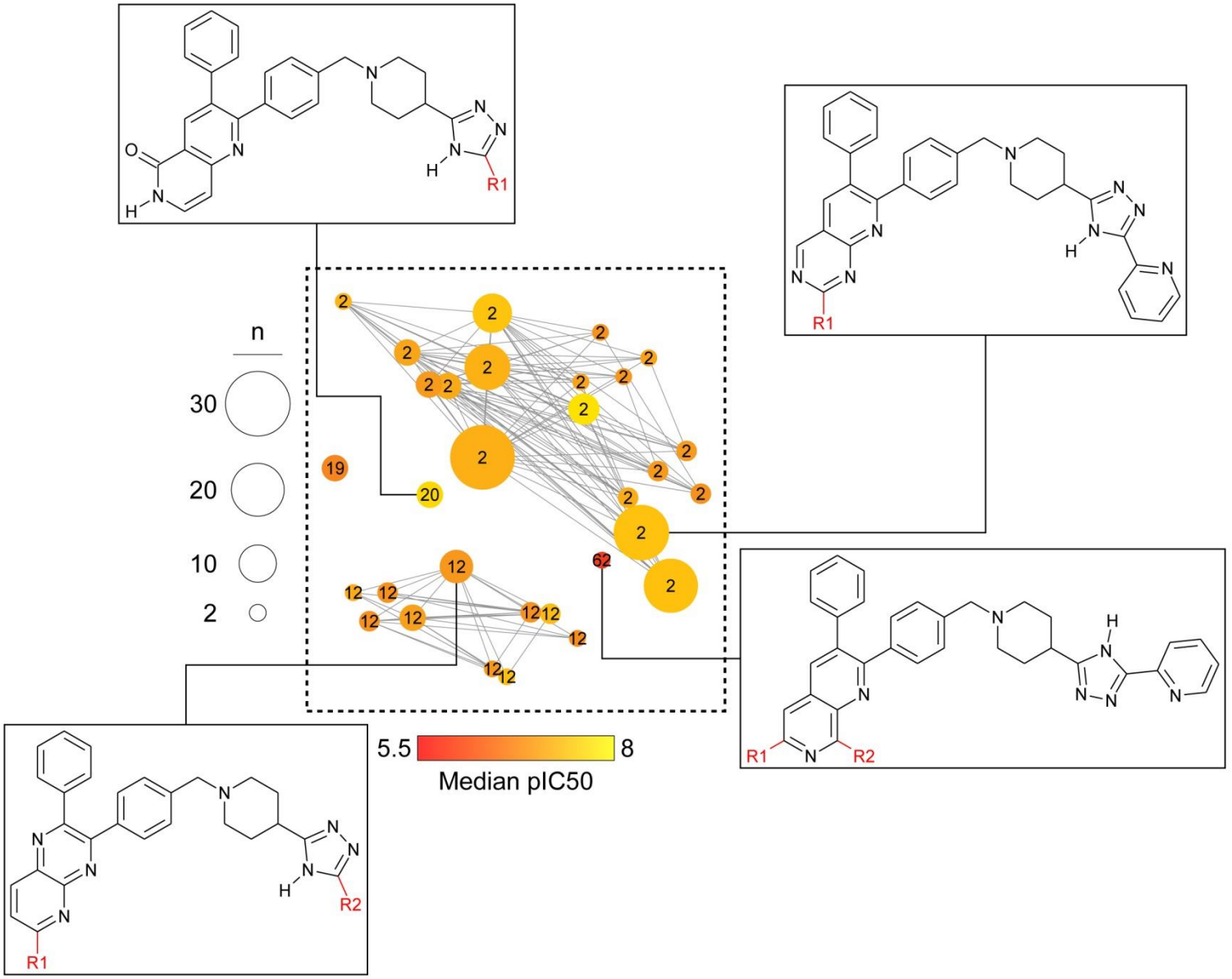
Constellation plot
AKT1 inhibitors



Cores can be mapped to the chemical space

- Comparison of analog series.

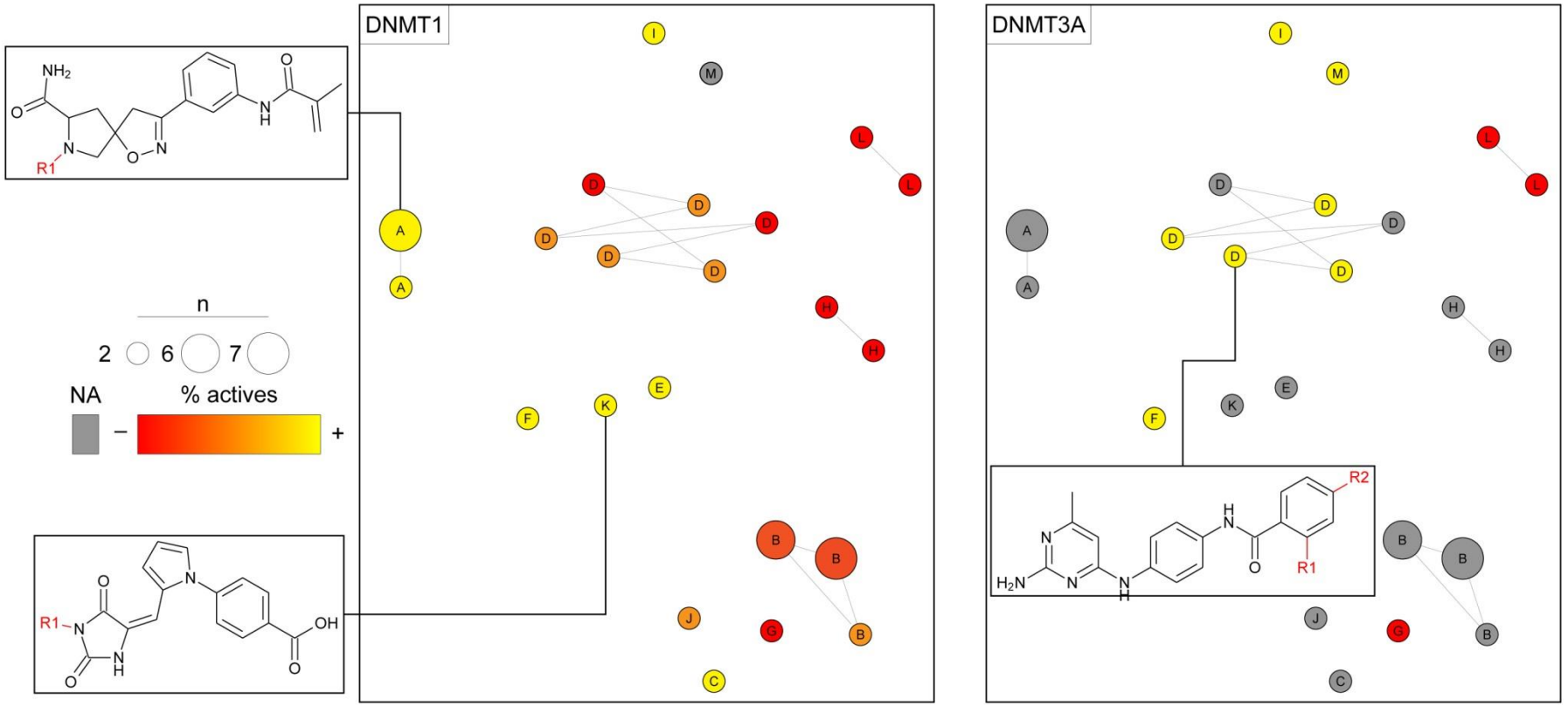
Constellation of AKT1 inhibitors



Multi-target constellations

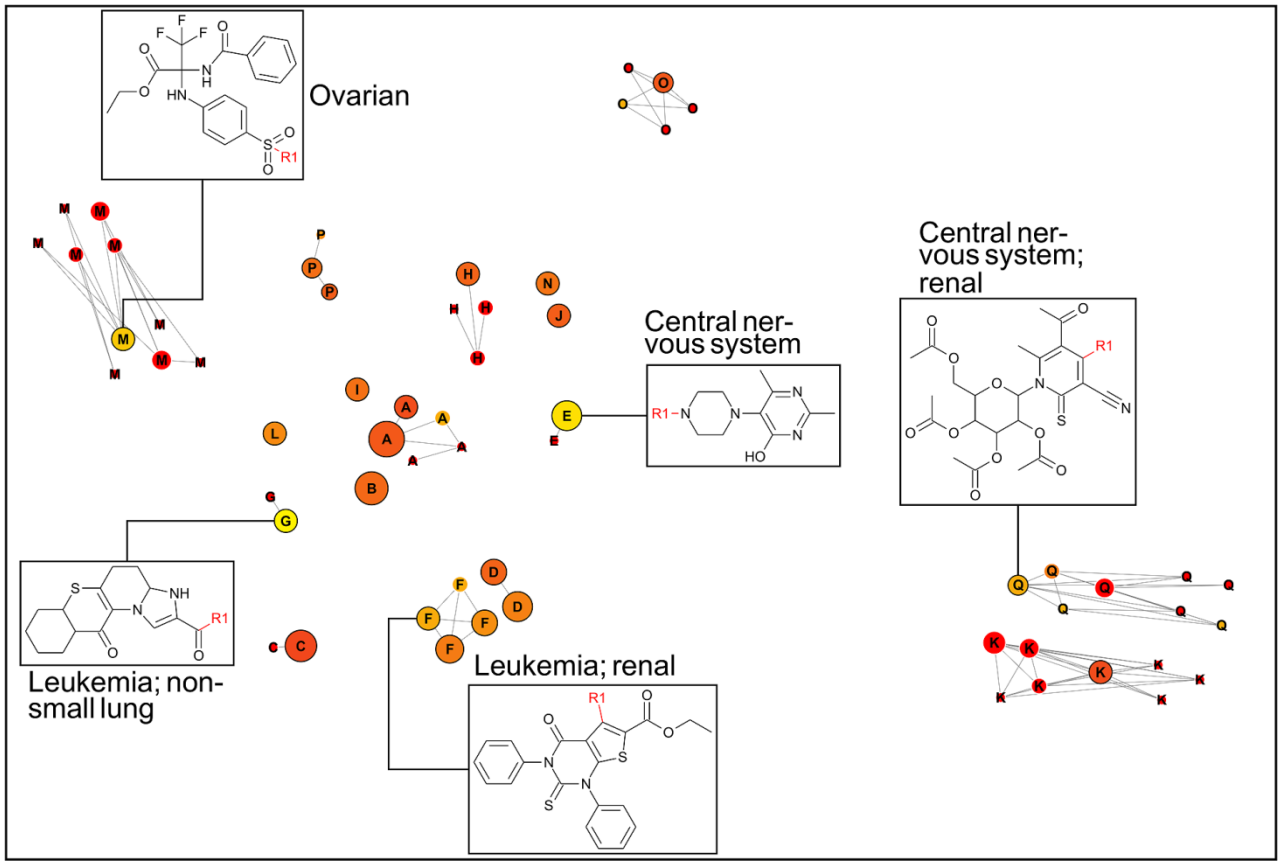
Inhibitors of DNA methyltransferases (DNMTs)

Multi-target data can be visualized by changing the coloring scheme.



Naveja J, Medina-Franco, JL. *Frontiers in Chemistry* 2019 7:510

SAR of compounds consistently tested against the same panel of cell lines



- NCI-60 data
- “active” or “inactive”
- 41,821 compounds
- 73 cell lines.

Constellation plot of 60 cores with high selectivity and consistency scores

Contribution to special issue of *Molecular Informatics*

Chemoinformatics Strasbourg Summer School - 2020

Full Paper www.molinf.com

molecular informatics

Check for updates

DOI: 10.1002/minf.202000061

Consistent Cell-selective Analog Series as Constellation Luminaries in Chemical Space

J. Jesús Naveja*[a] and José L. Medina-Franco*[b]

This manuscript is dedicated to all people affected directly or indirectly by the COVID-19 pandemic.

Jesús Naveja

DIFACQUIM UNAM

Concluding remarks and perspectives

- Activity Landscape Plotter
  - Free on-line resource based on fingerprints.
- Constellation Plots
  - Scaffold- and coordinate-based representation of chemical space.
  - Constellation in chemical space: region formed by core scaffolds with similar structure (as defined by a coordinate-based projection).
  - Enable exploring SARs/SPRs and analyze virtual screening results.
  - Identifies *bright* or *dark* regions.
  - Explore **StARs**: **St**ructure-**A**ctivity **R**elationships in chemical space

Perspectives

- Develop web-based form of Constellation Plots.
- Apply the plots to different types of screening.

Acknowledgments

DIFACQUIM: Computer-Aided Drug Design at the School of Chemistry

Dr. Jesús Naveja

Ph.D. students

Bárbara Díaz

Fernando Prieto

Norberto Sánchez

Ana Chávez

Emma Andrade

Master

Alejandro Gómez

Edgar López

B.Sc.

Eurídice Juárez

Yesenia Cruz

January 2020

www.difacquim.com

Collaborations

Univ. -Prof. Dr. Jürgen Bajorath

University of Bonn, Germany

Dr. Gerald M. Maggiora

University of Arizona, USA

Dr. Ulf Norinder

Stockholm University, Sweden

Grants

- CONACyT Basic Science 282785
- PAPIIT IA203718
- PAPIME PE200118, PAIP 5000-9163
- Miztli - Supercomputer (UNAM)
- Program NUATEI