VIS 2020

# ChemVA: Interactive Visual Analysis of Chemical Compound Similarity in Virtual Screening


María Virginia Sabando

M. Selzer, I. Ponzoni,
A. J. Soto, M. L. Ganuza

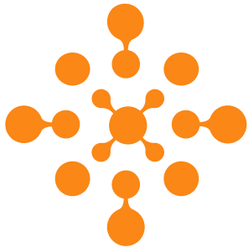
Universidad Nacional del Sur, Argentina



Pavol Ulbrich

J. Byška, J. Mičan,
B. Kozlíková

Masaryk University, Czech Republic



VIS2020

# ChemVA: Interactive Visual Analysis of Chemical Compound Similarity in Virtual Screening


María Virginia Sabando

M. Selzer, I. Ponzoni,
A. J. Soto, M. L. Ganuza

Universidad Nacional del Sur, Argentina

Pavol Ulbrich

J. Byška, J. Mičan,
B. Kozlíková

Masaryk University, Czech Republic

# Motivation

- Virtual screening → accelerate drug discovery
- Large sets of potential drug candidates
- Visualization tools:
  - Explore compounds and their features
  - Select the best candidates for testing *in-vitro*

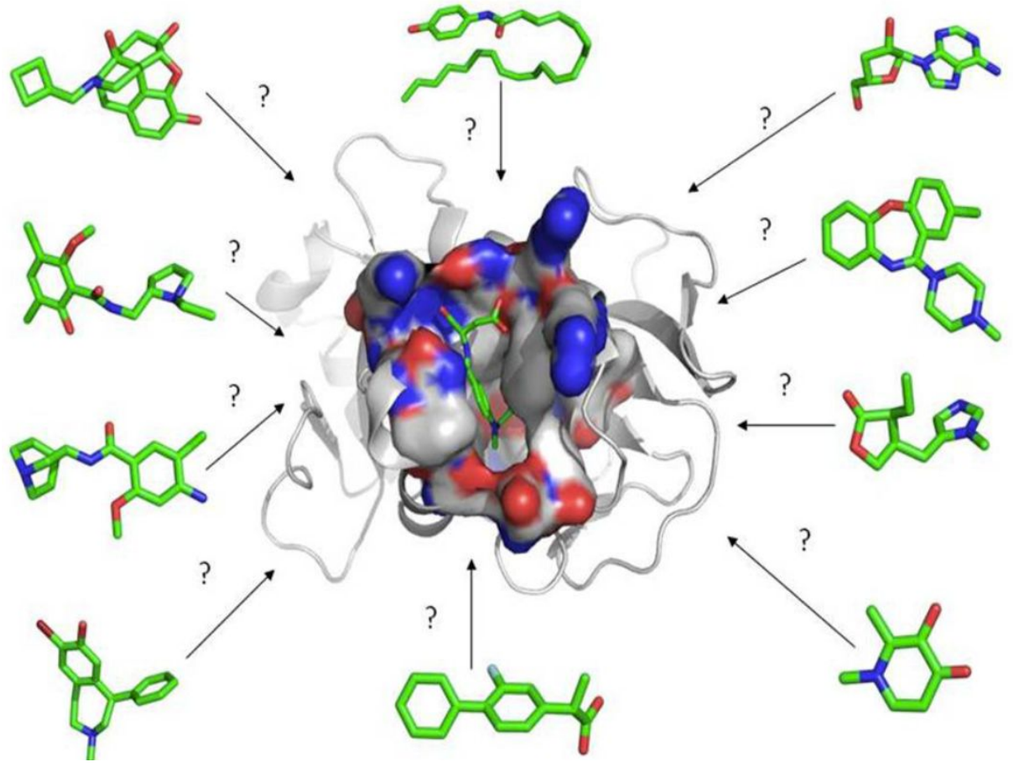


Image source:
https://cdn.rcsb.org/rcsb-pdb/general_information/news_publications/newsletters/2014q2/corner.html

# Motivation

- Virtual screening → accelerate drug discovery
- Large sets of potential drug candidates

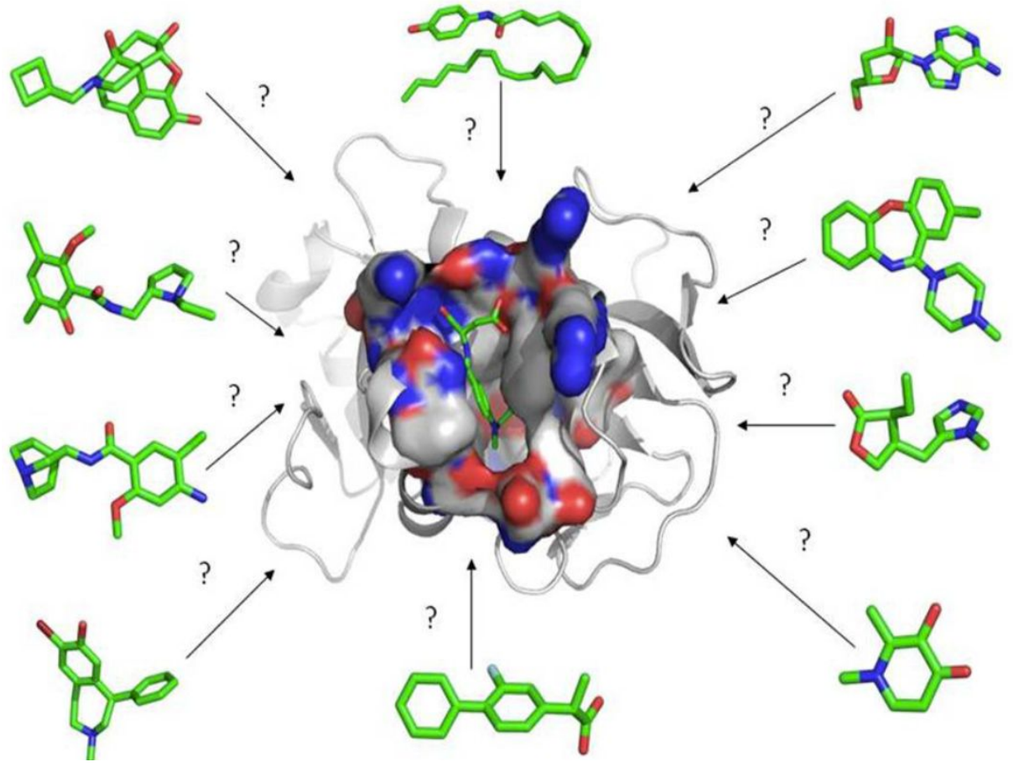


Image source:
https://cdn.rcsb.org/rcsb-pdb/general_information/news_publications/newsletters/2014q2/corner.html

- Visualization tools:
  - Explore compounds and their features
  - Select the best candidates for testing *in-vitro*

# Challenge 1: Multidimensional Nature of Molecules

- Similarity of compounds depends on many factors

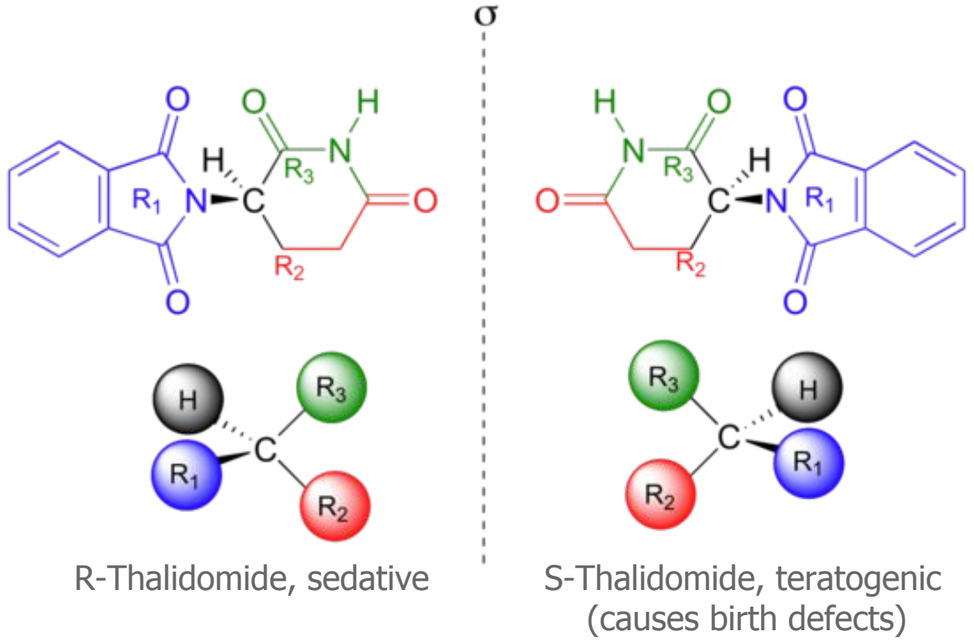



Image source: https://courses.lumenlearning.com/suny-mcc-organicchemistry/chapter/introduction-to-cycloalkanes/

! Structural similarity does not guarantee similar bioactivity!

- Multiple high-dimensional vector-based abstractions of a compound
- Visual support for the exploration of similarity

# Challenge 2: Dimensionality Reduction and its Trustworthiness

- Dimensionality reduction to visualize the molecular space

# Challenge 2: Dimensionality Reduction and its Trustworthiness

- Dimensionality reduction to visualize the molecular space

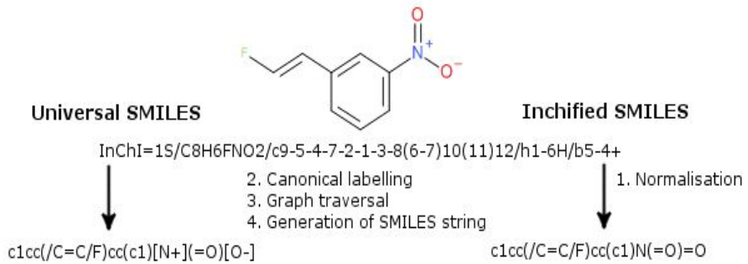



Image source: Towards a Universal SMILES representation - A standard method to generate canonical SMILES based on the InChI, Journal of Cheminformatics, 2012

# Challenge 2: Dimensionality Reduction and its Trustworthiness

- Dimensionality reduction to visualize the molecular space

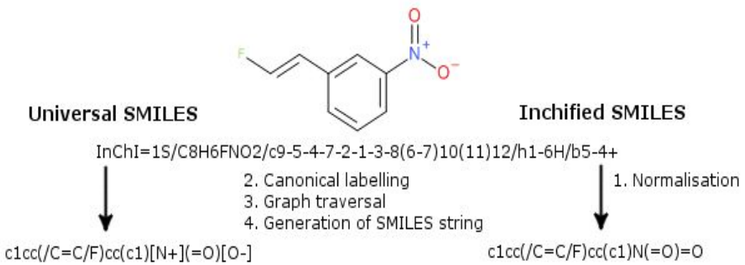



Image source: Towards a Universal SMILES representation - A standard method to generate canonical SMILES based on the InChI, Journal of Cheminformatics, 2012

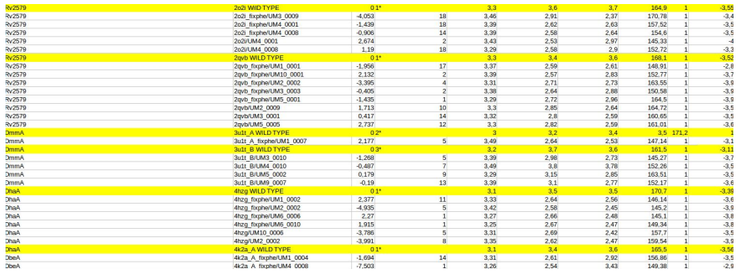


Image source: courtesy of Jan Mičan, researcher at Loschmidt Laboratories and Faculty of Medicine, Masaryk University, Czech Republic

# Challenge 2: Dimensionality Reduction and its Trustworthiness

- Dimensionality reduction to visualize the molecular space

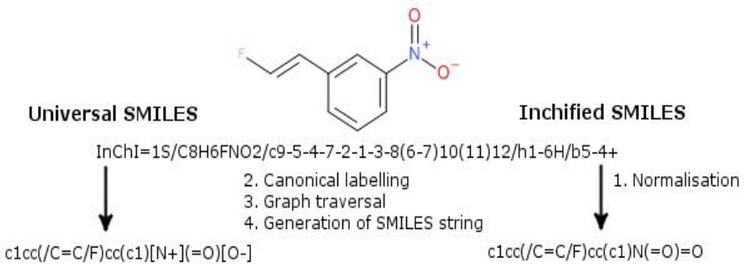



Image source: Towards a Universal SMILES representation - A standard method to generate canonical SMILES based on the InChI, Journal of Cheminformatics, 2012

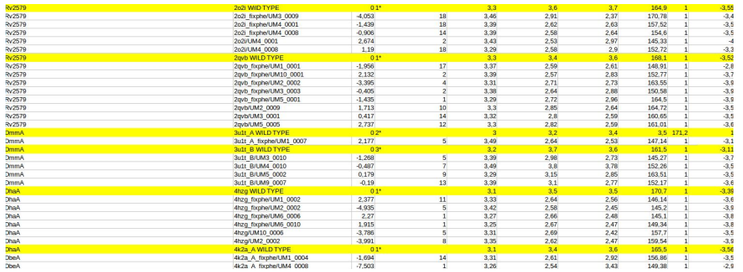


Image source: courtesy of Jan Mičan, researcher at Loschmidt Laboratories and Faculty of Medicine, Masaryk University, Czech Republic

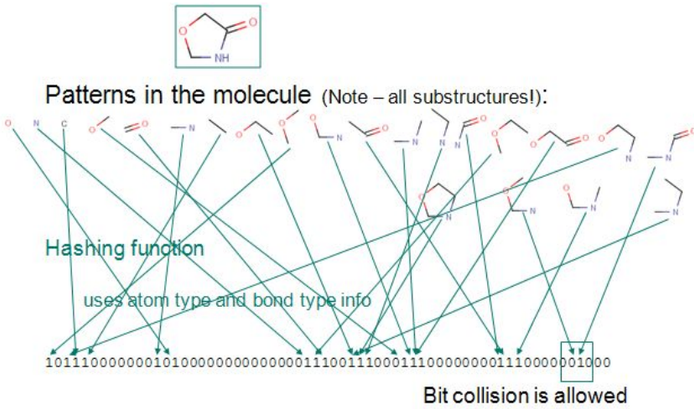



Image source: https://docs.chemaxon.com/Chemical_Hashed_Fingerprint.html

# Challenge 2: Dimensionality Reduction and its Trustworthiness

- Dimensionality reduction to visualize the molecular space

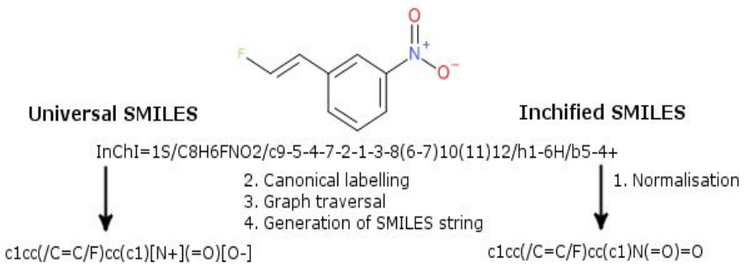



Image source: Towards a Universal SMILES representation - A standard method to generate canonical SMILES based on the InChI, Journal of Cheminformatics, 2012

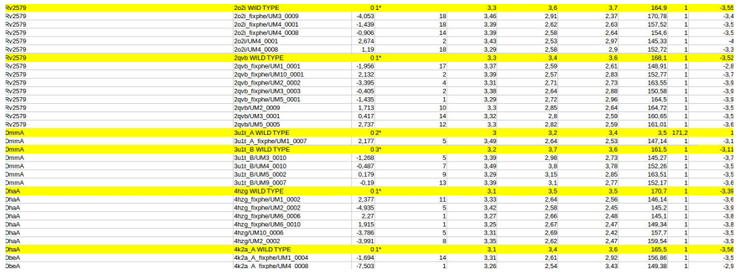


Image source: courtesy of Jan Mičan, researcher at Loschmidt Laboratories and Faculty of Medicine, Masaryk University, Czech Republic

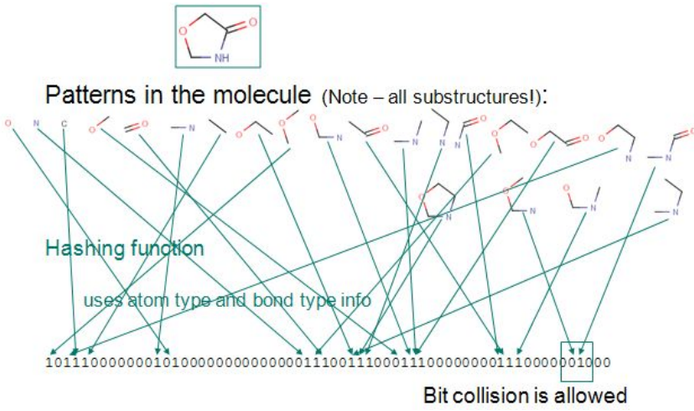



Image source: https://docs.chemaxon.com/Chemical_Hashed_Fingerprint.html


Dimensionality Reduction

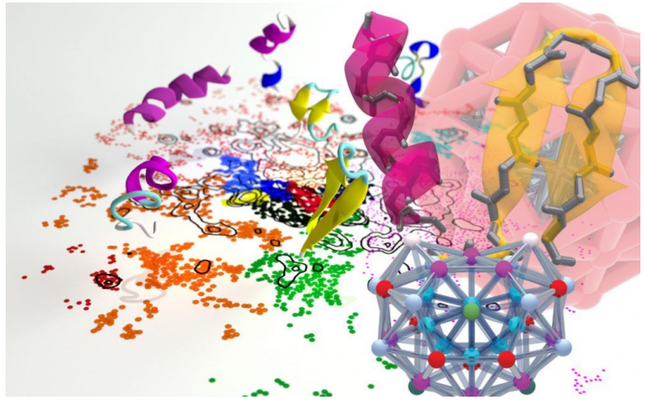


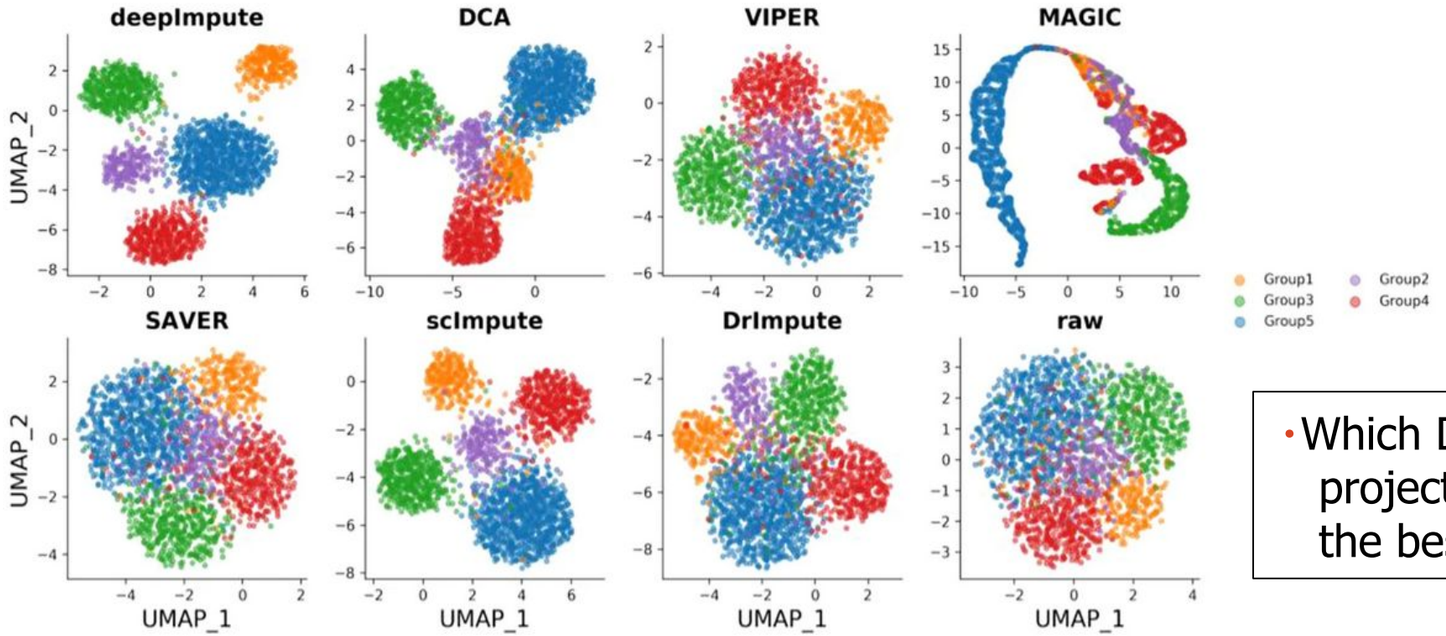
Image source: https://www.epfl.ch/labs/cosmo/research/page-148964-en-html/

# Challenge 2: Dimensionality Reduction and its Trustworthiness

- …but how to tell if a DR projection is trustworthy?

# Challenge 2: Dimensionality Reduction and its Trustworthiness

- …but how to tell if a DR projection is trustworthy?




- Which DR projection is the best one?

Image source: DeepImpute: an accurate, fast, and scalable deep neural network method to impute single-cell RNA-seq data, Genome Biology, 2019

- Comparative visualization for assessing trustworthiness

# Challenge 2: Dimensionality Reduction and its Trustworthiness

- …but how to tell if a DR projection is trustworthy?

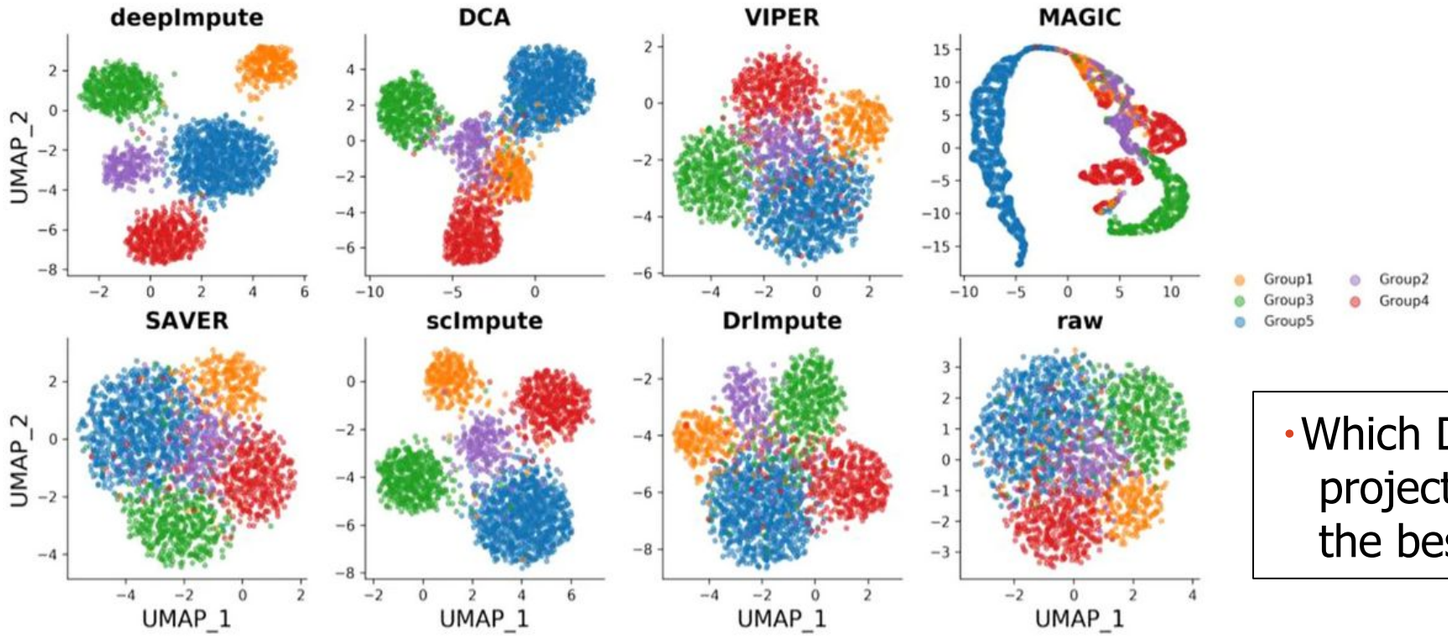
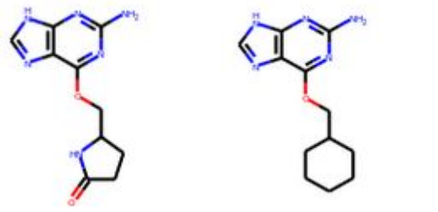



Image source: DeepImpute: an accurate, fast, and scalable deep neural network method to impute single-cell RNA-seq data, Genome Biology, 2019

- Which DR projection is the best one?




- Are similar molecules projected near each other in the latent space?

- Comparative visualization for assessing trustworthiness

# Challenge 3: Efficient Exploration of Large Molecular Ensembles

- Molecular ensembles → hundreds to thousands of compounds

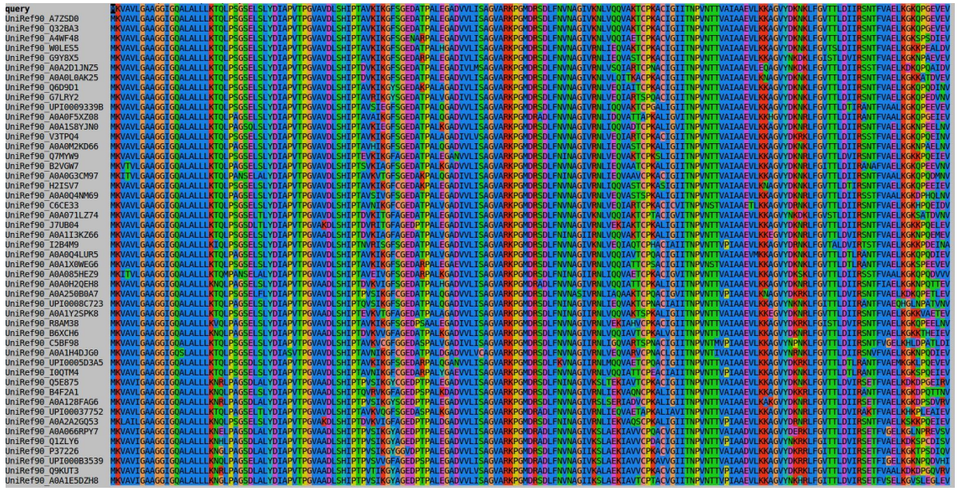


Image source: courtesy of Jan Mičan, researcher at Loschmidt Laboratories and Faculty of Medicine, Masaryk University, Czech Republic

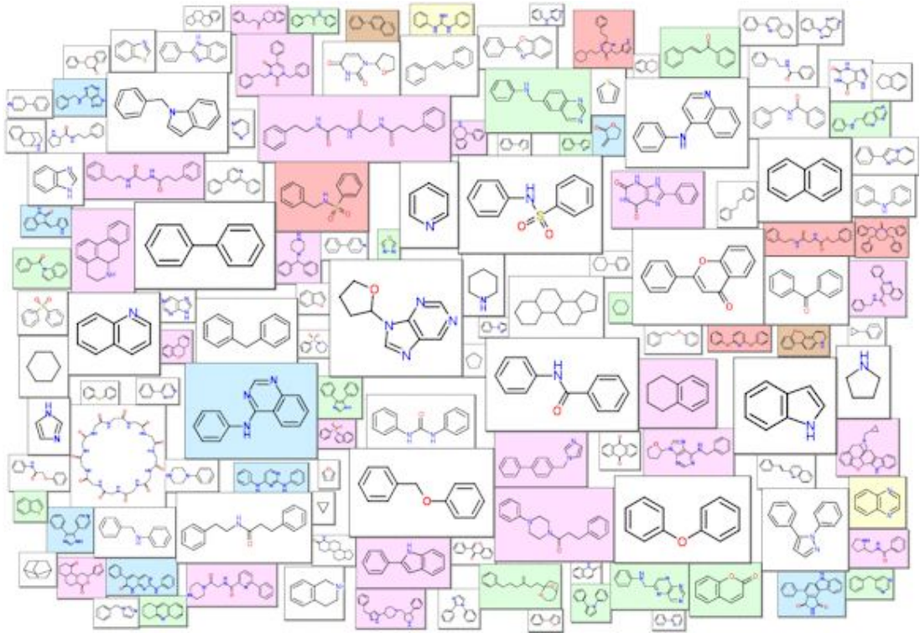
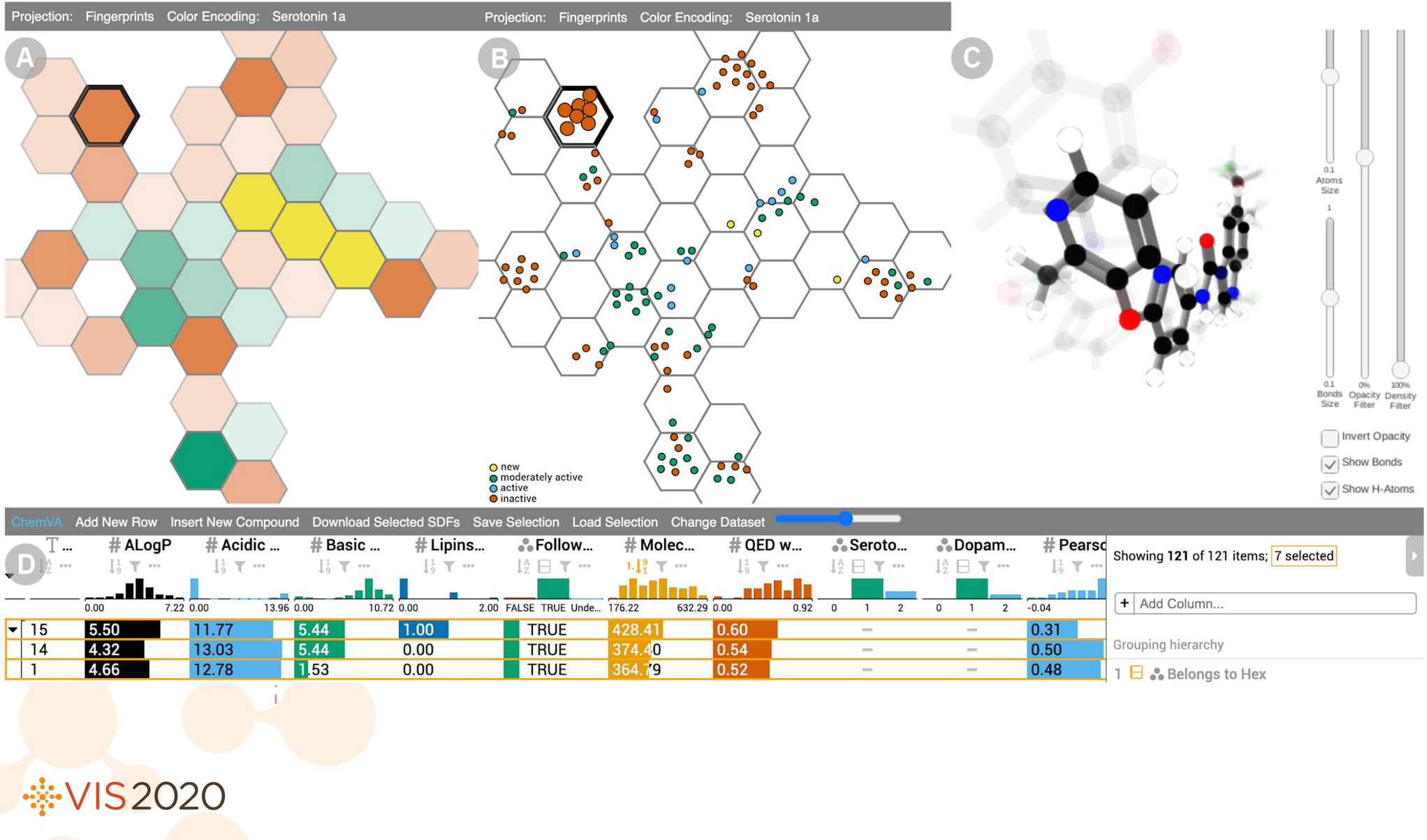


Image source: Ertl and Rohde, The Molecule Cloud - compact visualization of large collections of molecules, Journal of Cheminformatics, 2012

- Views and interactions for efficiently exploring molecular features

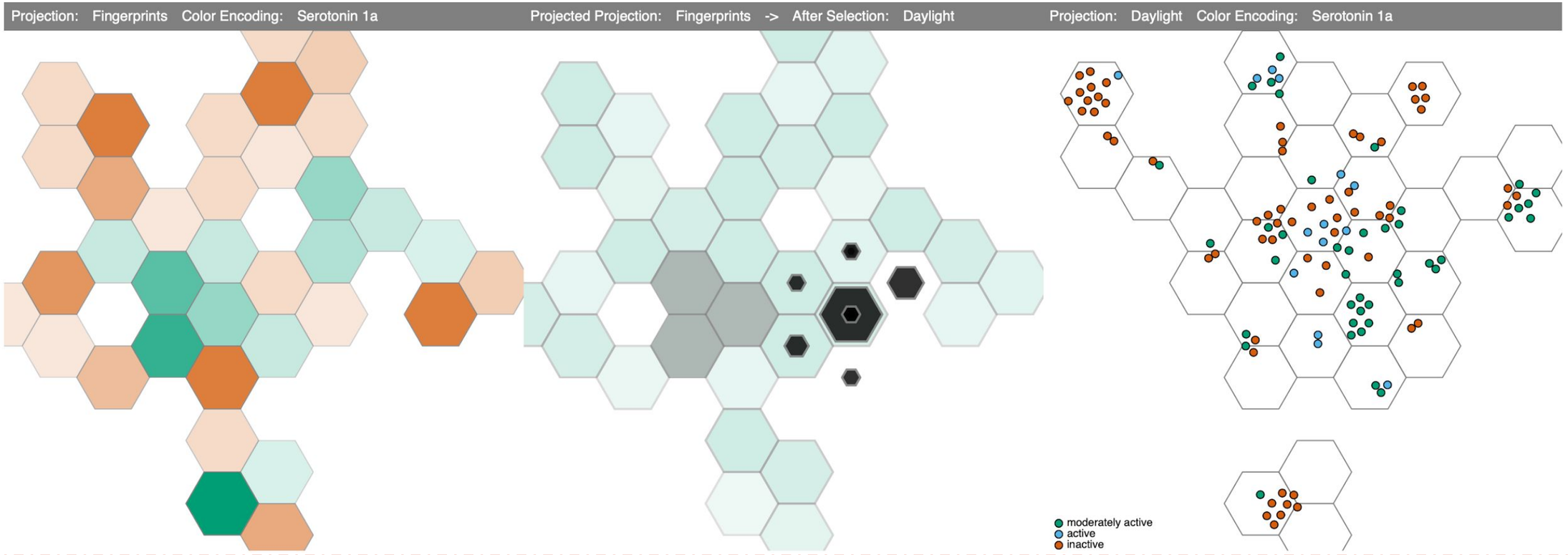

Projection: Fingerprints Color Encoding: Serotonin 1a
A
Projection: Fingerprints Color Encoding: Serotonin 1a
B
new
moderately active
active
inactive
C
0.1
Atoms Size
1
0.1
Bonds Size
0%
Opacity Filter
100%
Density Filter
Invert Opacity
Show Bonds
Show H-Atoms
ChemVA
Add New Row
Insert New Compound
Download Selected SDFs
Save Selection
Load Selection
Change Dataset
D
# ALogP
# Acidic ...
# Basic ...
# Lipins...
Follow...
# Molec...
# QED w...
Seroto...
Dopam...
# Pearsc
0.00 7.22
0.00 13.96
0.00 10.72
0.00 2.00
FALSE TRUE Unde...
176.22 632.29
0.00 0.92
0 1 2
0 1 2
-0.04
15 5.50 11.77 5.44 1.00 TRUE 428.41 0.60 0.31
14 4.32 13.03 5.44 0.00 TRUE 374.40 0.54 0.50
1 4.66 12.78 1.53 0.00 TRUE 364.79 0.52 0.48
Showing 121 of 121 items; 7 selected
Add Column...
Grouping hierarchy
1 Belongs to Hex

VIS2020


VIS2020

# Hexagonal View & Detail View

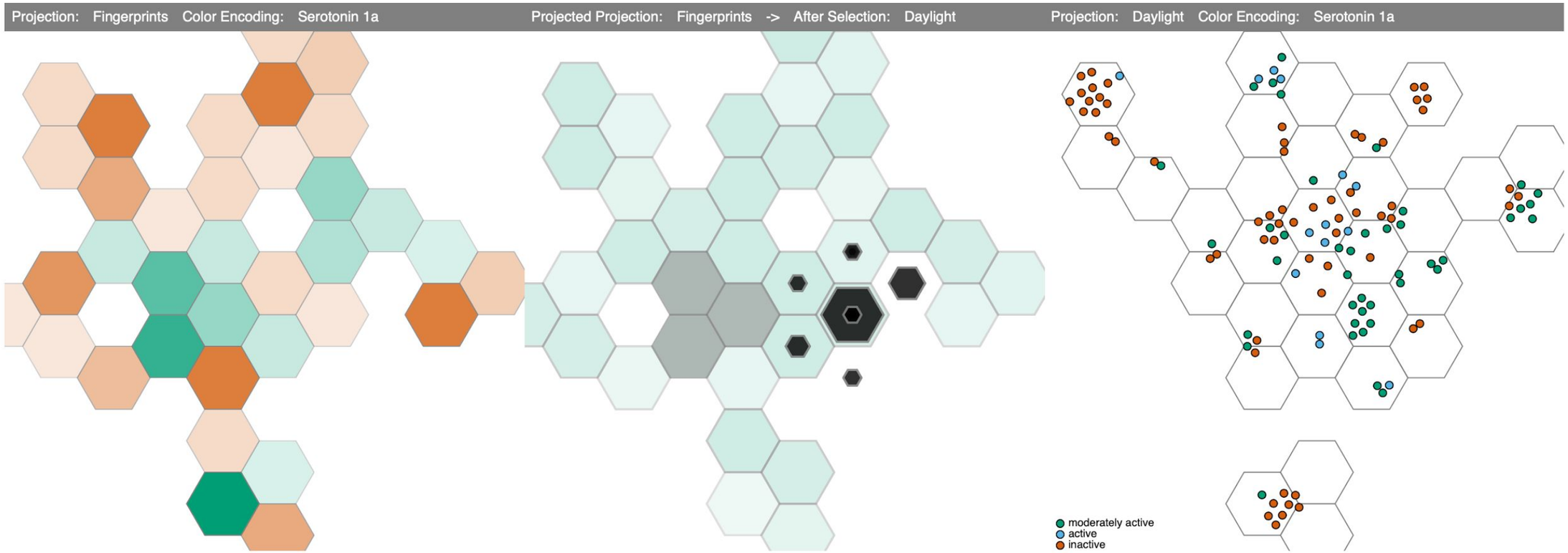
- Data from multiple input sources viewed from different perspectives
  - Variable color and opacity encoding



# Hexagonal View & Detail View

# Difference View for Evaluation of Projections

- Trustworthiness of DR projections

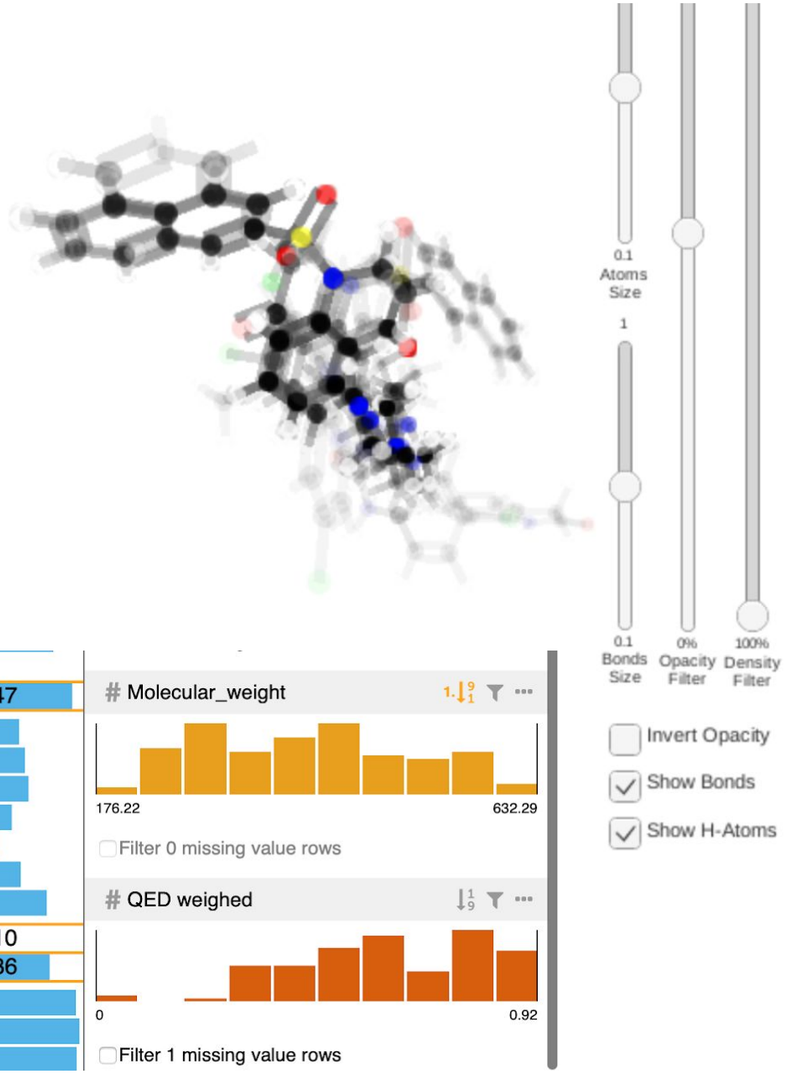


# Difference View for Evaluation of Projections

# Correlation Encoding

# Table View & 3D View

- Provides additional insight into the data
  - Spatial alignment with opacity modulation
  - Sorting, filtering, advanced selection



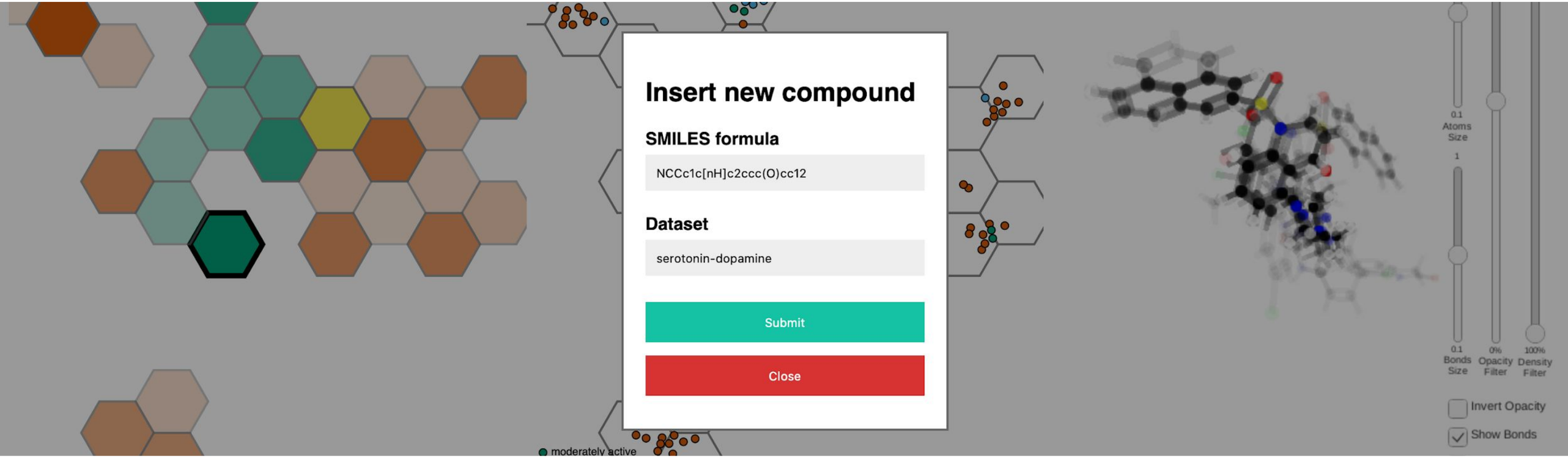

VIS2020

K. Furmanová, S. Gratzl, H. Stitz, T. Zichner, M. Jarešová, A. Lex, and M. Streit. Taggle: Combining overview and details in tabular data visualizations. Information Visualization, 19(2):114–136, 2020. doi: 10.1177/1473871619878085

# Table View & 3D View

# Application Scenario: Adding New Compounds

- Highlighted with yellow color in the dataset



# Case Study:

## Activity towards Serotonin and Dopamine receptors

# Results from Evaluation

- ChemVA was evaluated by three domain experts and one DR visualization expert
  - Five new compounds designed using ChemVA
  - Intuitive and easy to use
  - Saves time and effort in drug design tasks
  - Difference view is effective for comparing DR projections in terms of their trustworthiness

# Results from Evaluation

- ChemVA was evaluated by three domain experts and one DR visualization expert
  - Five new compounds designed using ChemVA
  - Intuitive and easy to use
  - Saves time and effort in drug design tasks
  - Difference view is effective for comparing DR projections in terms of their trustworthiness

# Conclusions and Future Work

- Techniques behind ChemVA would apply to other domains and types of data

- Future extensions of the tool:
  - Load and customize your own datasets
  - Provide options for advanced search: substructures, molecular descriptors.

VIS 2020

# Thank you!

arXiv.org

Please refer to our paper:
ChemVA: Interactive Visual Analysis of Chemical Compound Similarity in Virtual Screening
(preprint available on arXiv)

All resources are publicly available on GitHub.
Check out our repository!

María Virginia Sabando
virginia.sabando@cs.uns.edu.ar
Pavol Ulbrich
paloulbrich@gmail.com

VIS 2020

# Thank you!

arXiv.org

Please refer to our paper:
ChemVA: Interactive Visual Analysis of Chemical Compound Similarity in Virtual Screening
(preprint available on arXiv)

All resources are publicly available on GitHub.
Check out our repository!

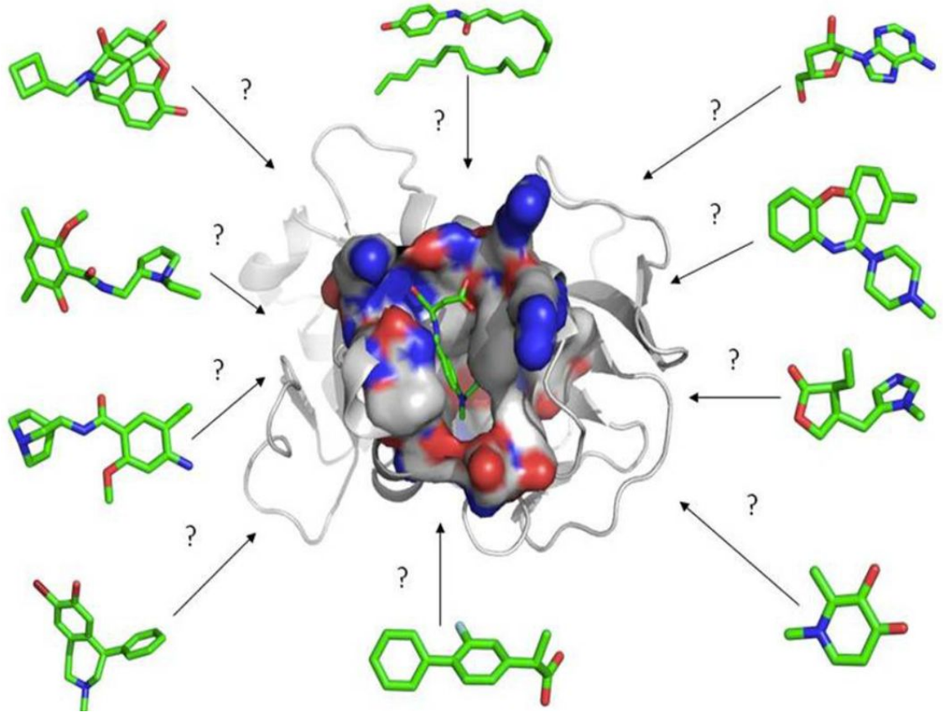
María Virginia Sabando
virginia.sabando@cs.uns.edu.ar
Pavol Ulbrich
paloulbrich@gmail.com

# Agenda

- Challenges
  - Multiple sources of molecular data
  - Efficient exploration of large molecular ensembles
  - DR trustworthiness
- ChemVA overview
  - 2D views
  - 3D view and Table view
  - Adding new compounds
- Case study
- Results from evaluation
- Conclusions and future work

# Motivation

- Virtual screening → accelerate drug discovery
- Large sets of potential drug candidates



- Visualization tools:
  - Explore compounds and their features
  - Select the best candidates for testing *in-vitro*

VIS2020

Image source: https://cdn.rcsb.org/rcsb-pdb/general_information/news_publications/newsletters/2014q2/corner.html