



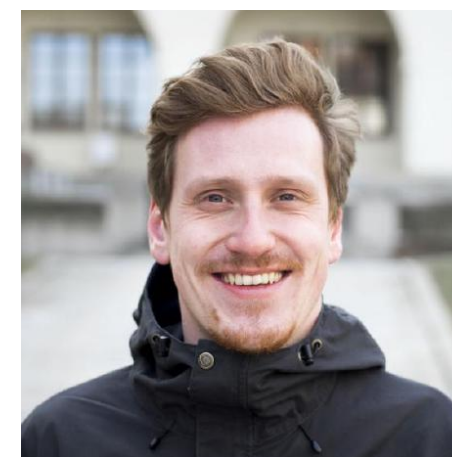
ChemVA: Interactive Visual Analysis of Chemical Compound Similarity in Virtual Screening



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



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

A decorative orange molecular structure is located in the bottom right corner of the slide, consisting of a central node with several smaller nodes connected by lines, similar in style to the main logo.

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*

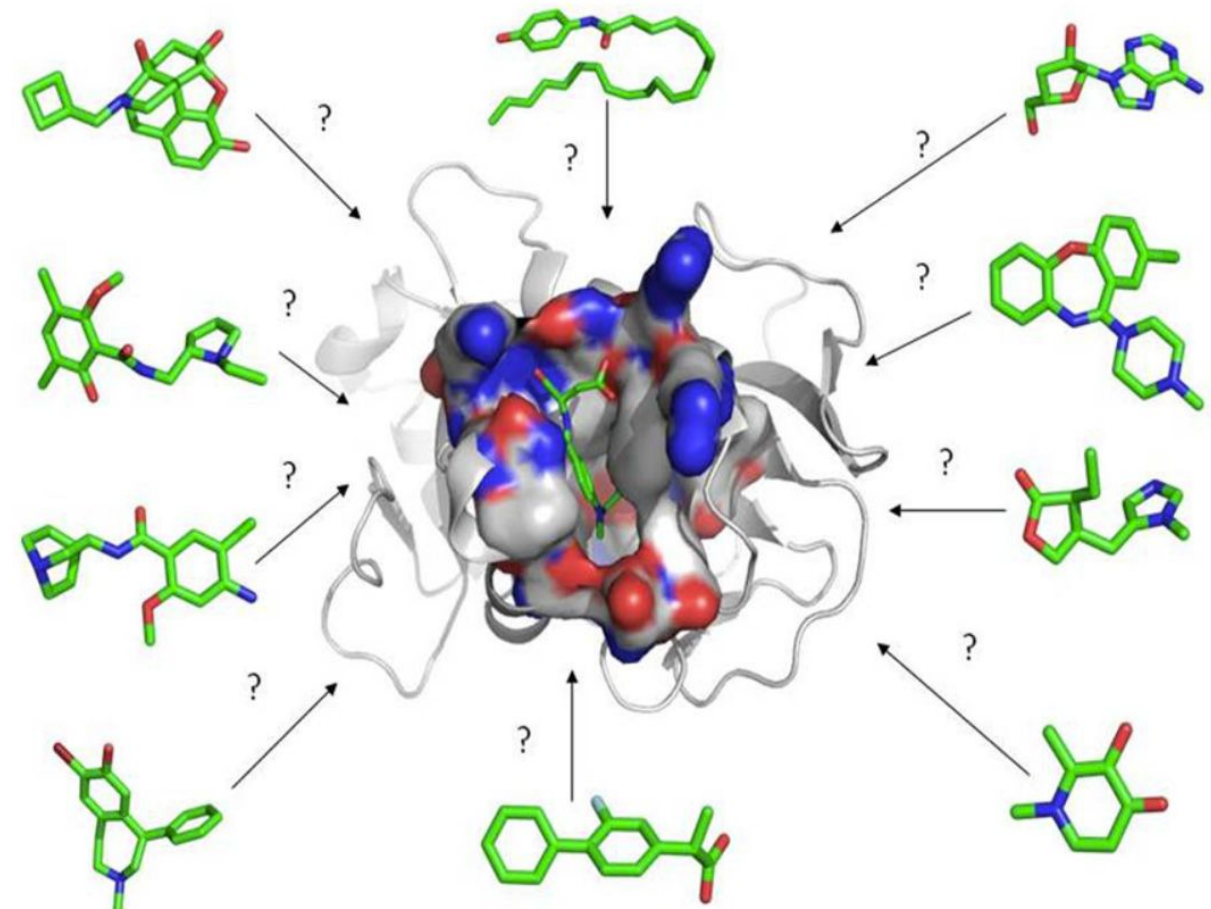


Image source:
https://cdn.rcsb.org/rcsb-pdb/general_information/news_publications/newsletters/2014q2/corner.html



Motivation

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

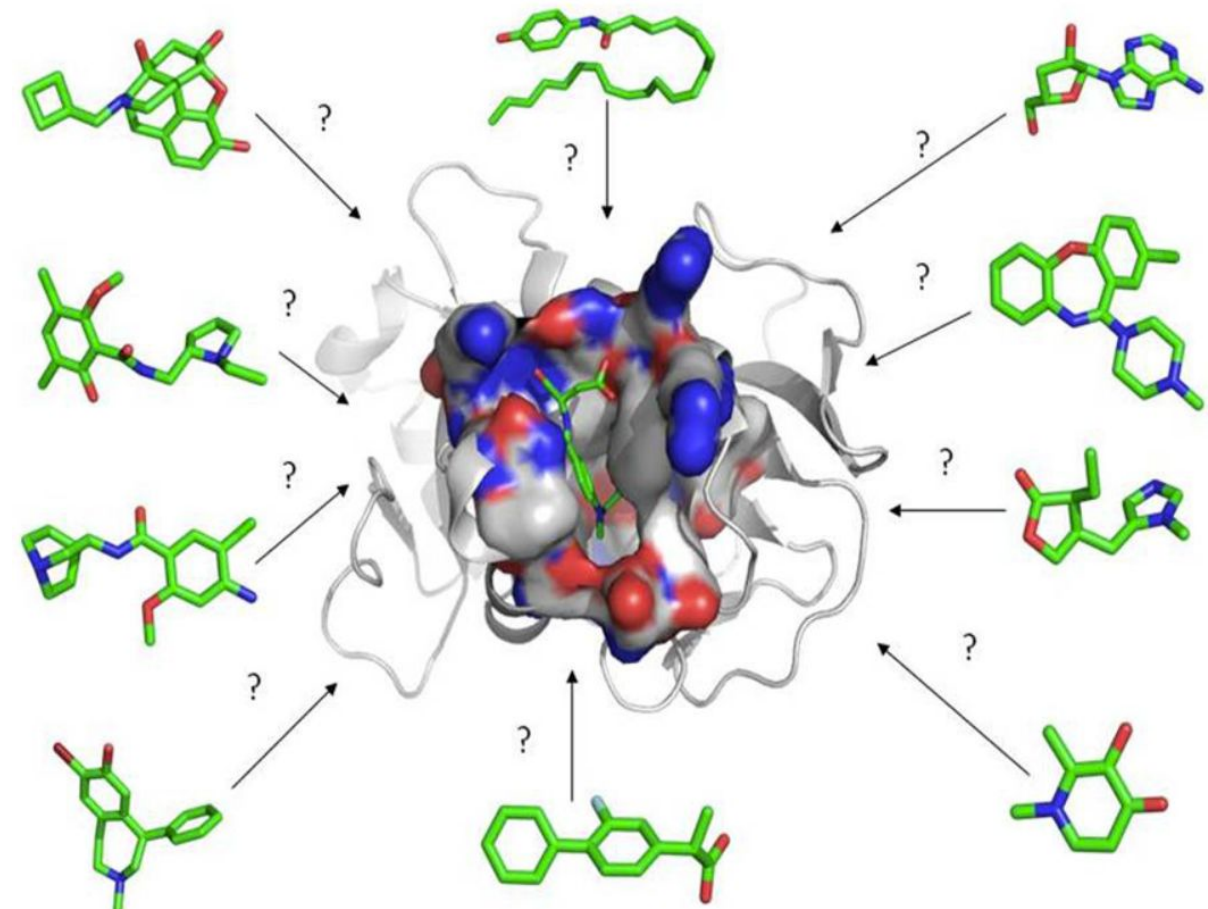
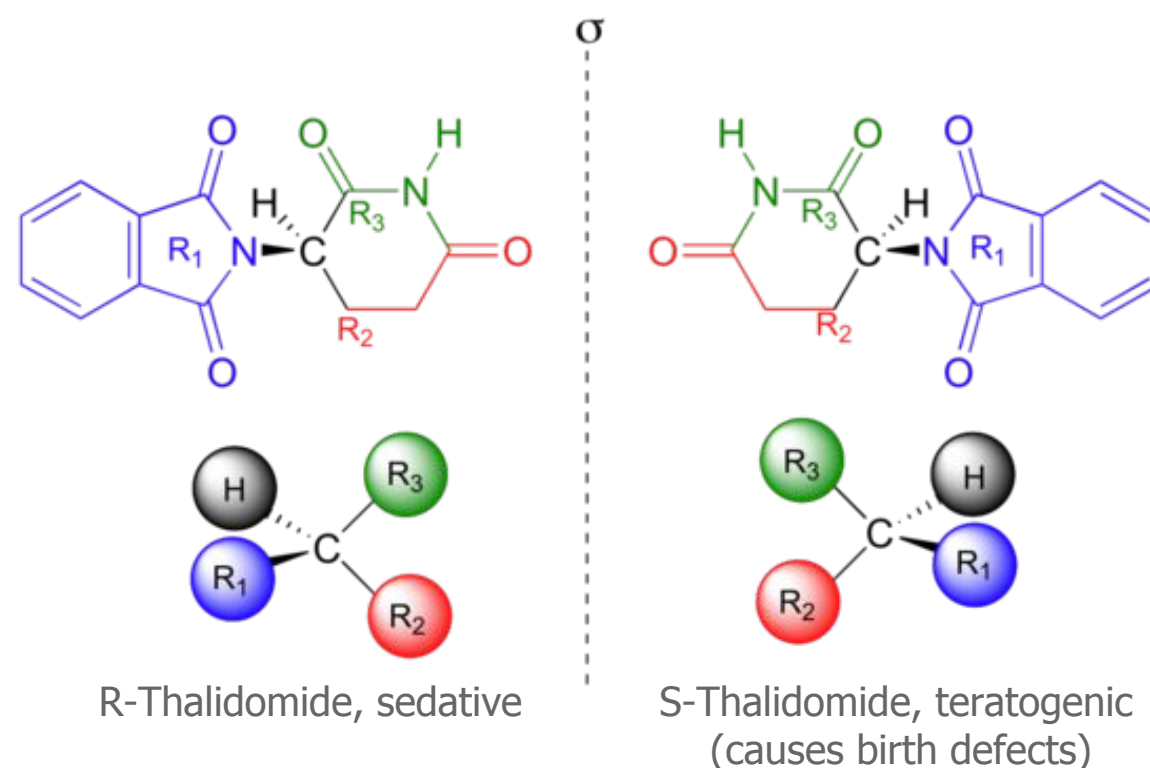


Image source:

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

Challenge 1: Multidimensional Nature of Molecules

- Similarity of compounds depends on many factors



! Structural similarity does not guarantee similar bioactivity!

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

- 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

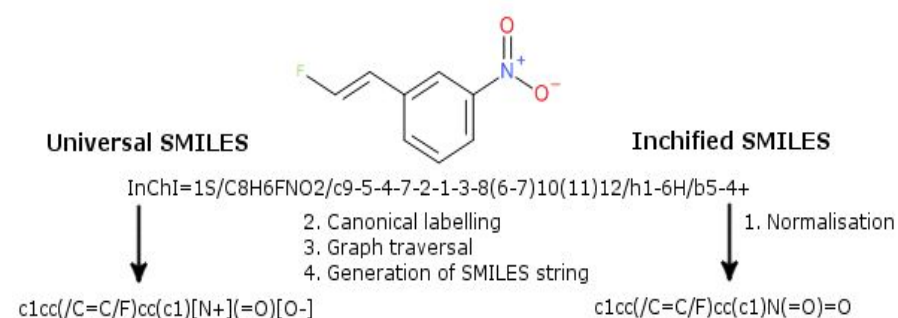


Image source: Towards a Universal SMILES representation - A standard method to generate canonical SMILES based on the InChI, Journal of Cheminformatics, 2012

Challenge 2: Dimensionality Reduction and its Trustworthiness

- Dimensionality reduction to visualize the molecular space

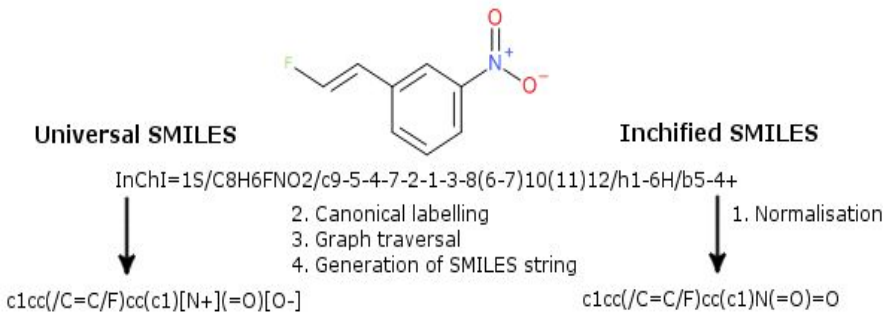


Image source: Towards a Universal SMILES representation - A standard method to generate canonical SMILES based on the InChI, Journal of Cheminformatics, 2012

Rv2579	2cd_WILD TYPE	0.17	3.3	3.6	3.7	164.9	1	-3.9	
Rv2579	2cd_TxphelUM3_0009	-4.053	18	3.46	2.91	2.37	170.78	1	-3.4
Rv2579	2cd_TxphelUM4_0001	-1.439	18	3.39	2.62	2.63	157.52	1	-3.5
Rv2579	2cd_TxphelUM4_0008	-0.906	14	3.39	2.58	2.64	154.16	1	-3.5
Rv2579	2cd/UM4_0001	2.674	2	3.43	2.53	2.97	145.33	1	-4
Rv2579	2cd/UM4_0009	1.19	18	3.29	2.58	2.9	152.72	1	-3.3
Rv2579	2pbt_WILD TYPE	0.17	3.3	3.4	3.6	168.1	1	-3.55	
Rv2579	2pbt_TxphelUM1_0001	-1.956	17	3.37	2.59	2.61	149.91	1	-2.8
Rv2579	2pbt_TxphelUM10_0001	2.132	2	3.39	2.97	2.63	152.77	1	-3.7
Rv2579	2pbt_TxphelUM2_0002	-3.995	4	3.31	2.71	2.73	163.55	1	-3.9
Rv2579	2pbt_TxphelUM3_0003	-0.405	2	3.38	2.64	2.88	150.58	1	-3.9
Rv2579	2pbt_TxphelUM5_0001	-1.435	1	3.29	2.72	2.96	164.5	1	-3.9
Rv2579	2pbt/UM2_0009	1.713	10	3.3	2.85	2.64	164.72	1	-3.5
Rv2579	2pbt/UM3_0001	0.417	14	3.32	2.8	2.99	160.69	1	-3.5
Rv2579	2pbt/UM5_0005	2.737	12	3.3	2.82	2.99	161.01	1	-3.6
DmnaA	3u1_A_WILD TYPE	0.27	5	3.2	3.4	3.5	171.2	1	-3.1
DmnaA	3u1_A_TxphelUM1_0007	2.177	5	3.49	2.64	2.53	147.14	1	-3.1
DmnaA	3u1_B_WILD TYPE	0.37	5	3.2	3.7	3.6	161.5	1	-3.11
DmnaA	3u1_B/UM3_0010	-1.298	5	3.39	2.98	2.73	145.27	1	-3.7
DmnaA	3u1_B/UM4_0010	-0.487	7	3.49	3.8	3.78	152.26	1	-3.5
DmnaA	3u1_B/UM5_0002	0.179	9	3.29	3.15	2.85	163.51	1	-3.5
DmnaA	3u1_B/UM6_0007	-0.19	13	3.39	3.1	2.77	152.17	1	-3.6
DnaA	4tsg_WILD TYPE	0.17	13	3.1	3.5	3.5	170.7	1	-3.39
DnaA	4tsg_TxphelUM1_0002	2.377	11	3.33	2.64	2.56	149.14	1	-3.6
DnaA	4tsg_TxphelUM2_0002	-4.935	5	3.42	2.58	2.45	145.2	1	-3.9
DnaA	4tsg_TxphelUM5_0006	2.27	1	3.27	2.66	2.48	145.1	1	-3.8
DnaA	4tsg_TxphelUM6_0010	1.915	1	3.25	2.67	2.47	149.34	1	-3.8
DnaA	4tsg/UM2_0005	-3.786	5	3.31	2.69	2.42	157.7	1	-3.5
DnaA	4tsg/UM2_0002	-3.991	8	3.35	2.62	2.47	159.54	1	-3.9
DnaA	4k2a_A_WILD TYPE	0.17	14	3.1	3.4	3.6	165.5	1	-3.59
DnaA	4k2a_A_TxphelUM1_0004	-1.694	14	3.31	2.81	2.92	156.86	1	-3.5
DnaA	4k2a_A_TxphelUM4_0008	-7.503	1	3.26	2.54	3.43	149.38	1	-2.9

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

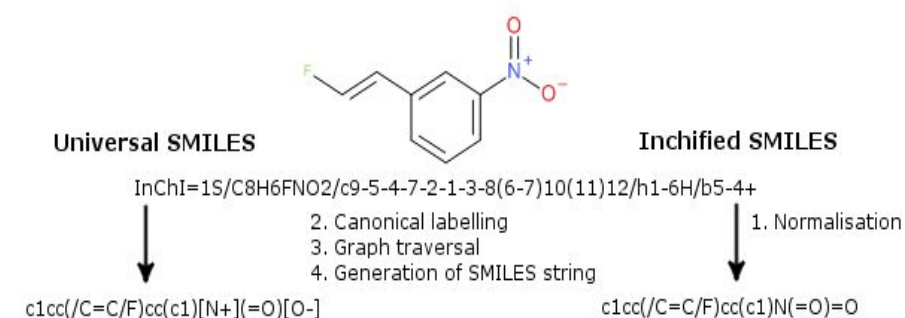


Image source: Towards a Universal SMILES representation - A standard method to generate canonical SMILES based on the InChI, Journal of Cheminformatics, 2012

Rv2579	2d2b_WILD TYPE	0.17	3.3	2.6	3.7	164.9	1	-3.9
Rv2579	2d2b_TxphelUM3_0009	-4.053	18	3.46	2.91	170.78	1	-3.4
Rv2579	2d2b_TxphelUM4_0001	-1.439	18	3.39	2.62	157.52	1	-3.5
Rv2579	2d2b_TxphelUM4_0008	-0.906	14	3.39	2.66	154.16	1	-3.5
Rv2579	2d2bUM4_0001	2.674	2	3.43	2.53	145.33	1	-4
Rv2579	2d2bUM4_0009	1.19	18	3.29	2.58	152.72	1	-3.3
Rv2579	2d2b_WILD TYPE	0.17	3.3	2.4	3.6	168.1	1	-3.55
Rv2579	2d2b_TxphelUM1_0001	-1.956	17	3.37	2.59	149.91	1	-2.8
Rv2579	2d2b_TxphelUM10_0001	2.132	2	3.39	2.97	152.77	1	-3.7
Rv2579	2d2b_TxphelUM2_0002	-3.995	4	3.31	2.71	163.55	1	-3.9
Rv2579	2d2b_TxphelUM3_0003	-0.405	2	3.38	2.64	150.58	1	-3.9
Rv2579	2d2b_TxphelUM5_0001	-1.435	1	3.29	2.72	164.5	1	-3.9
Rv2579	2d2bUM2_0009	1.713	10	3.3	2.85	164.72	1	-3.5
Rv2579	2d2bUM3_0001	0.417	14	3.32	2.8	160.89	1	-3.5
Rv2579	2d2bUM5_0005	2.737	12	3.3	2.82	161.01	1	-3.6
Dm1A	3d1_A_WILD TYPE	0.27	3	3.2	3.4	171.2	1	-3.1
Dm1A	3d1_A_TxphelUM1_0007	2.177	5	3.49	2.64	147.14	1	-3.1
Dm1A	3d1_B_WILD TYPE	0.37	5	3.2	3.7	161.5	1	-3.1
Dm1A	3d1_BUM3_0010	-1.298	5	3.39	2.98	145.27	1	-3.7
Dm1A	3d1_BUM4_0010	-0.487	7	3.49	3.8	152.26	1	-3.5
Dm1A	3d1_BUM5_0002	0.179	9	3.29	3.15	163.51	1	-3.5
Dm1A	3d1_BUM6_0007	-0.19	13	3.39	3.1	152.17	1	-3.6
Dm1A	4d2g_WILD TYPE	0.17	13	3.1	3.5	170.7	1	-3.38
Dm1A	4d2g_TxphelUM1_0002	2.377	11	3.33	2.64	146.14	1	-3.6
Dm1A	4d2g_TxphelUM2_0002	-4.905	5	3.42	2.58	145.2	1	-3.9
Dm1A	4d2g_TxphelUM5_0006	2.27	1	3.27	2.66	145.1	1	-3.8
Dm1A	4d2g_TxphelUM6_0010	1.915	1	3.25	2.67	149.34	1	-3.8
Dm1A	4d2gUM2_0005	-3.786	5	3.31	2.69	157.7	1	-3.5
Dm1A	4d2gUM2_0002	-3.991	8	3.35	2.62	159.54	1	-3.9
Dm1A	4k2a_A_WILD TYPE	0.17	14	3.1	3.4	165.5	1	-3.58
Dm1A	4k2a_A_TxphelUM1_0004	-1.694	14	3.31	2.61	146.86	1	-3.5
Dm1A	4k2a_A_TxphelUM4_0008	-7.503	1	3.26	2.54	149.38	1	-2.9

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

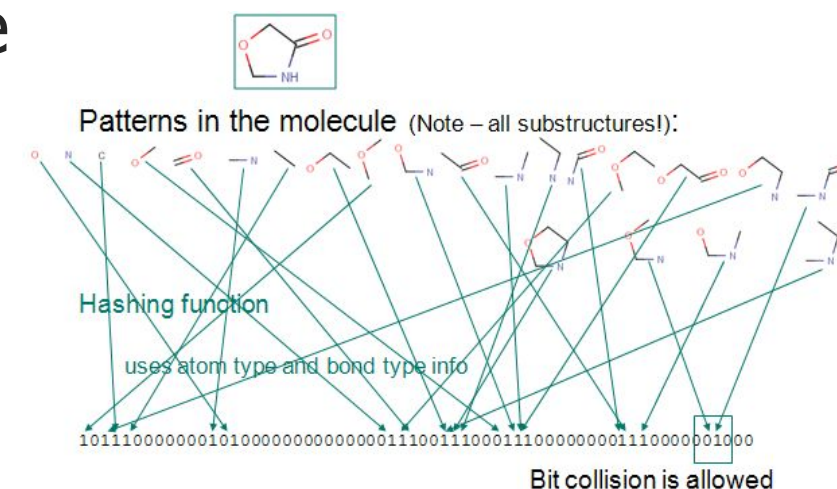


Image source:
https://docs.chemaxon.com/Chemical_Hashed_Fingerprint.html

Challenge 2: Dimensionality Reduction and its Trustworthiness

- Dimensionality reduction to visualize the molecular space

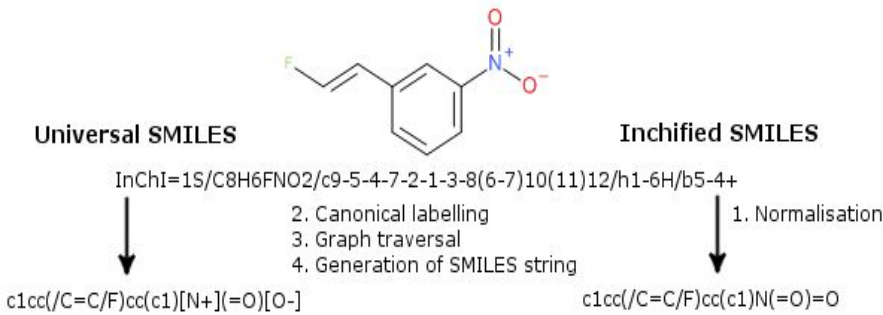


Image source: Towards a Universal SMILES representation - A standard method to generate canonical SMILES based on the InChI, Journal of Cheminformatics, 2012

Rv2579	20b_WILD TYPE	0.1*	3.3	2.6	3.7	164.9	1	-3.9
Rv2579	20b_TspAeUM3_0009	-4.053	18	2.46	2.91	170.78	1	-3.4
Rv2579	20b_TspAeUM4_0001	-1.439	18	3.39	2.62	157.52	1	-3.5
Rv2579	20b_TspAeUM4_0008	-0.906	14	3.39	2.66	154.16	1	-3.5
Rv2579	20b/UM4_0001	2.674	2	3.43	2.53	145.33	1	-4
Rv2579	20b/UM4_0008	1.19	18	3.29	2.58	152.72	1	-3.3
Rv2579	20b_WILD TYPE	0.1*	3.3	2.4	3.6	168.1	1	-3.55
Rv2579	20b_TspAeUM1_0001	-1.956	17	3.37	2.59	149.91	1	-2.8
Rv2579	20b_TspAeUM10_0001	2.132	2	3.39	2.97	152.77	1	-3.7
Rv2579	20b_TspAeUM2_0002	-3.995	4	3.31	2.71	163.55	1	-3.9
Rv2579	20b_TspAeUM3_0003	-0.405	2	3.38	2.64	150.58	1	-3.9
Rv2579	20b_TspAeUM5_0005	-1.435	1	3.29	2.72	164.5	1	-3.9
Rv2579	20b/UM2_0009	1.713	10	3.3	2.85	164.72	1	-3.5
Rv2579	20b/UM3_0001	0.417	14	3.32	2.8	160.89	1	-3.5
Rv2579	20b/UM5_0005	2.737	12	3.3	2.82	161.01	1	-3.6
Dmna	3d1_A_WILD TYPE	0.2*	3	3.2	3.4	171.2	1	-3.1
Dmna	3d1_A_TspAeUM1_0007	2.177	5	3.49	2.64	147.14	1	-3.1
Dmna	3d1_B_WILD TYPE	0.3*	5	3.2	3.7	161.5	1	-3.1
Dmna	3d1_B/UM3_0010	-1.298	5	3.39	2.98	145.27	1	-3.7
Dmna	3d1_B/UM4_0010	-0.487	7	3.49	3.48	152.26	1	-3.5
Dmna	3d1_B/UM5_0002	0.179	9	3.29	3.15	163.51	1	-3.5
Dmna	3d1_B/UM7_0007	-0.19	13	3.39	3.1	152.17	1	-3.6
DnaA	4bzg_WILD TYPE	0.1*	13	3.1	3.5	170.7	1	-3.8
DnaA	4bzg_TspAeUM1_0002	2.377	11	3.33	2.64	149.14	1	-3.6
DnaA	4bzg_TspAeUM2_0002	-4.905	5	3.42	2.56	145.2	1	-3.9
DnaA	4bzg_TspAeUM5_0006	2.27	1	3.27	2.66	145.1	1	-3.8
DnaA	4bzg_TspAeUM6_0010	1.915	1	3.25	2.67	149.34	1	-3.8
DnaA	4bzg/UM2_0005	-3.786	9	3.31	2.69	157.7	1	-3.5
DnaA	4bzg/UM2_0002	-3.991	8	3.35	2.62	159.54	1	-3.9
DnaA	4k2a_A_WILD TYPE	0.1*	14	3.1	3.4	165.5	1	-3.98
DnaA	4k2a_A_TspAeUM1_0004	-1.694	14	3.31	2.81	156.86	1	-3.5
DnaA	4k2a_A_TspAeUM4_0008	-7.503	1	3.26	2.54	149.38	1	-2.9

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

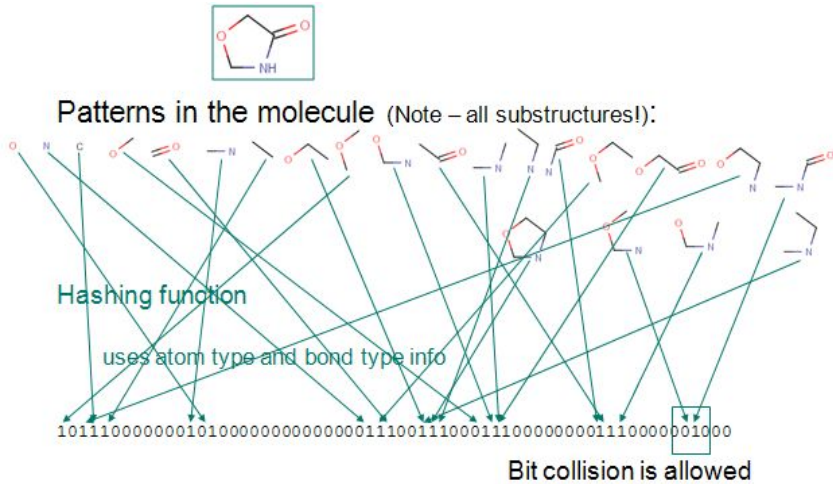
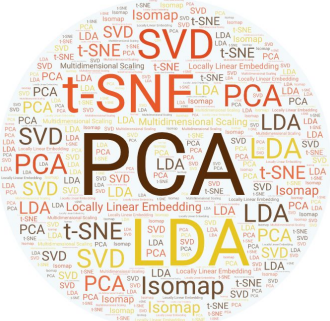


Image source: https://docs.chemaxon.com/Chemical_Hashed_Fingerprint.html



Dimensionality Reduction

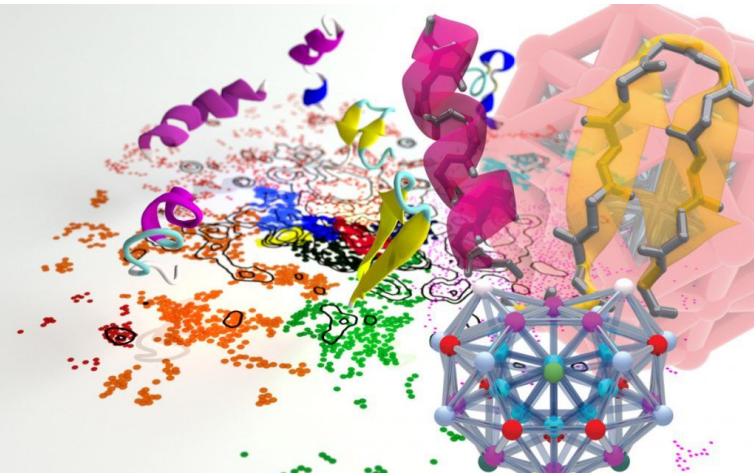


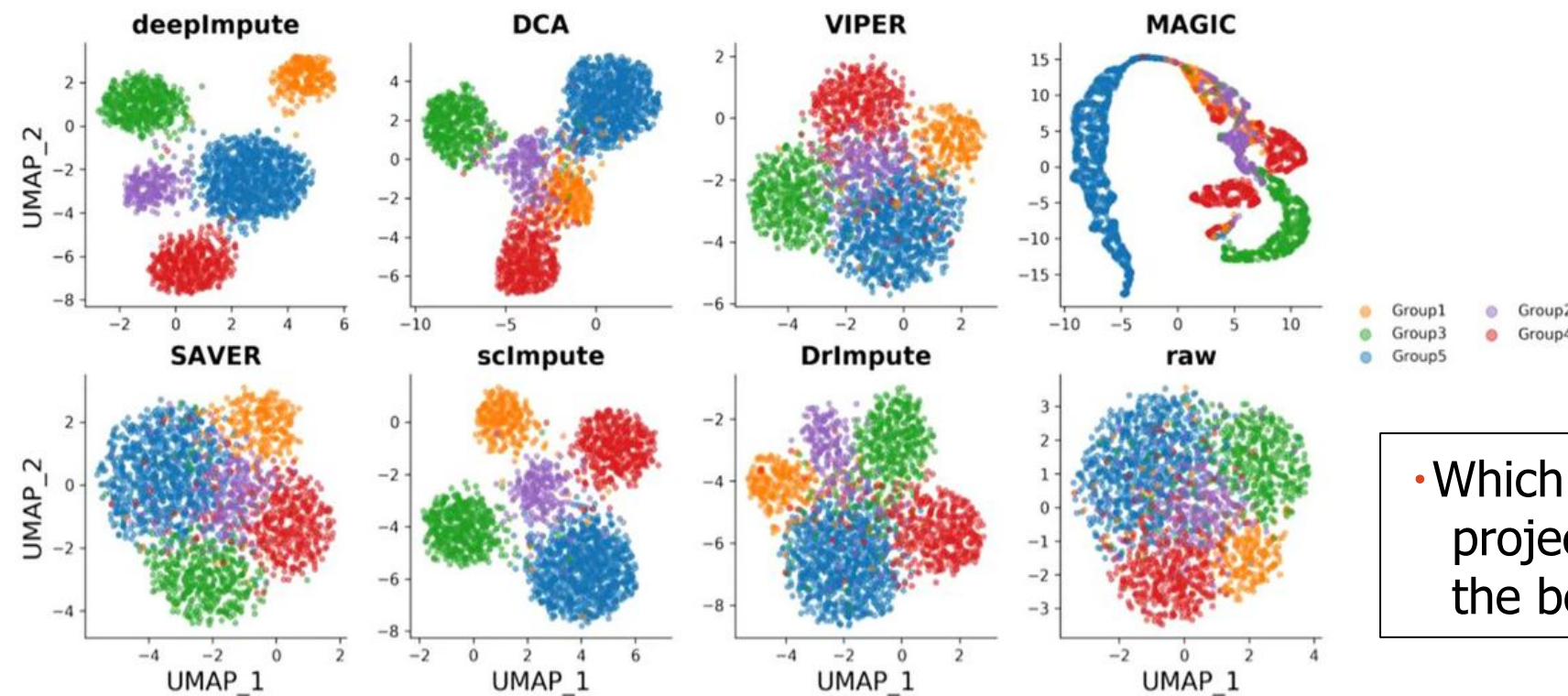
Image source: <https://www.epfl.ch/labs/cosmo/research/page-148964-en-html/>

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?

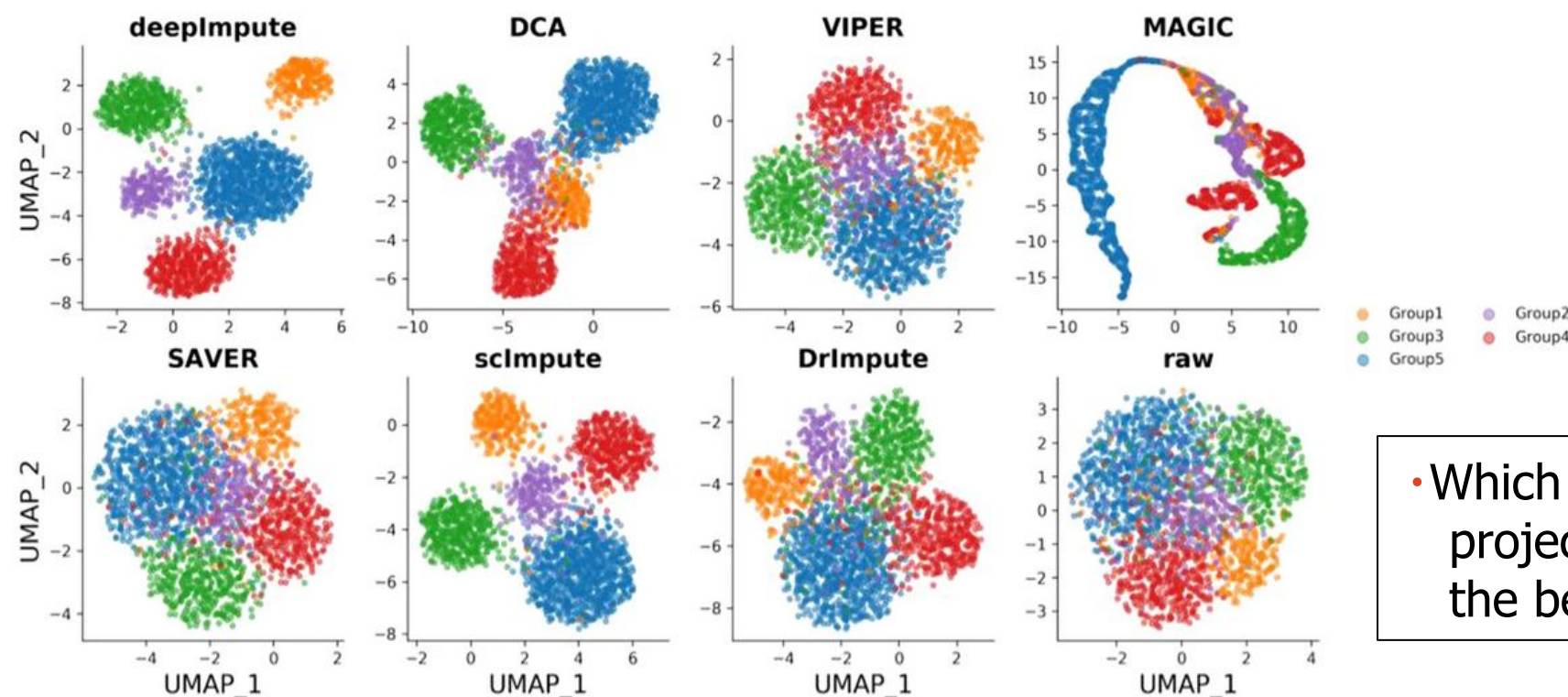
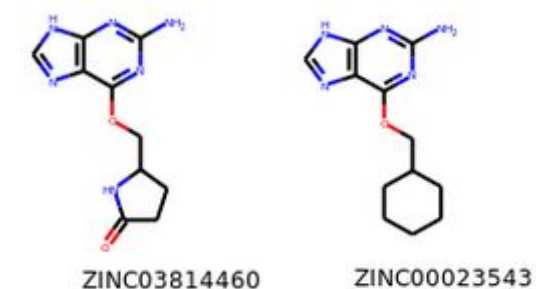


Image source: DeepImpute: an accurate, fast, and scalable deep neural network method to impute single-cell RNA-seq data, Genome Biology, 2019



- 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

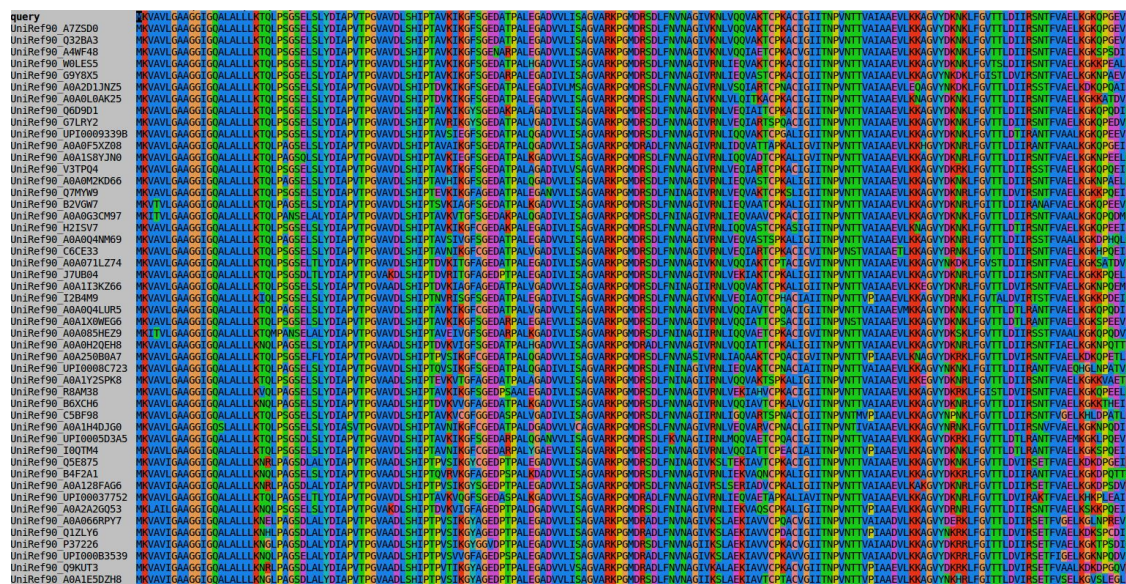


Image source: courtesy of Jan Mičan, researcher at Loschmidt Laboratories and Faculty of Medicine, Masaryk University, Czech Republic

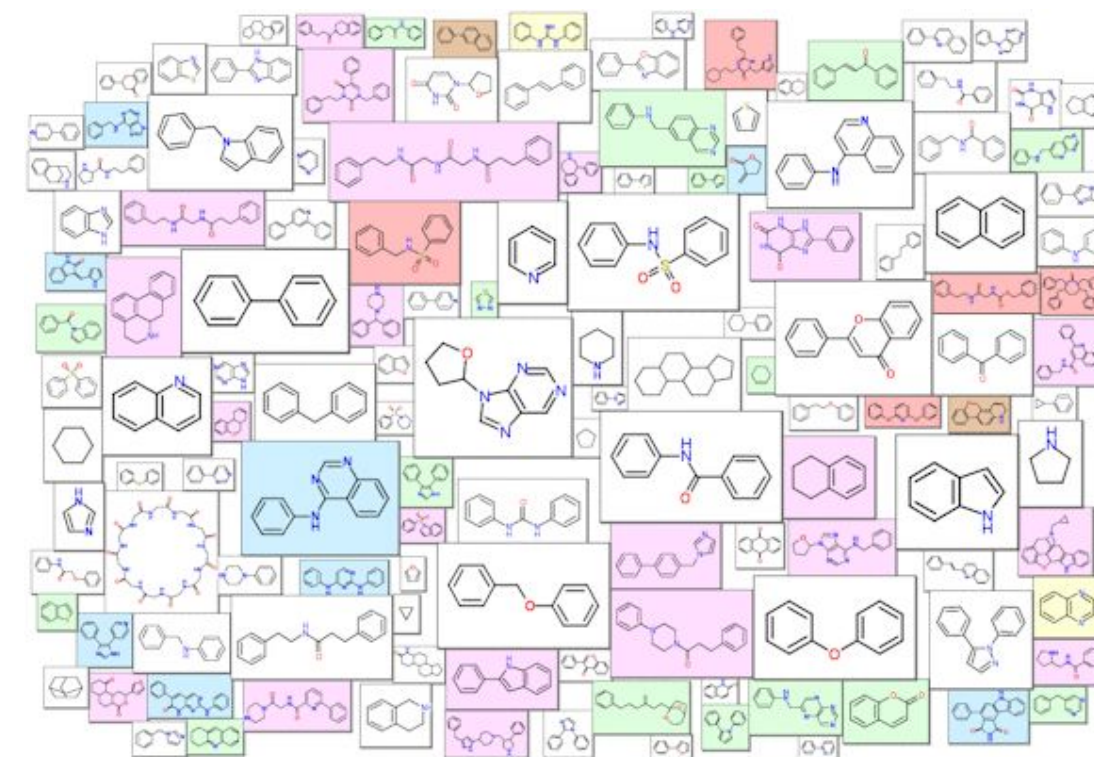
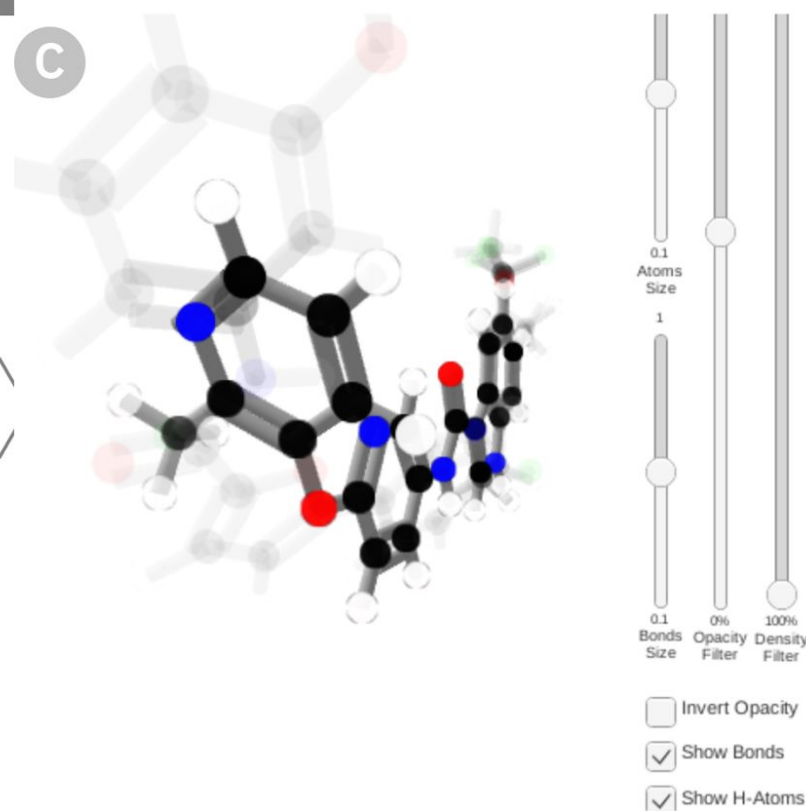
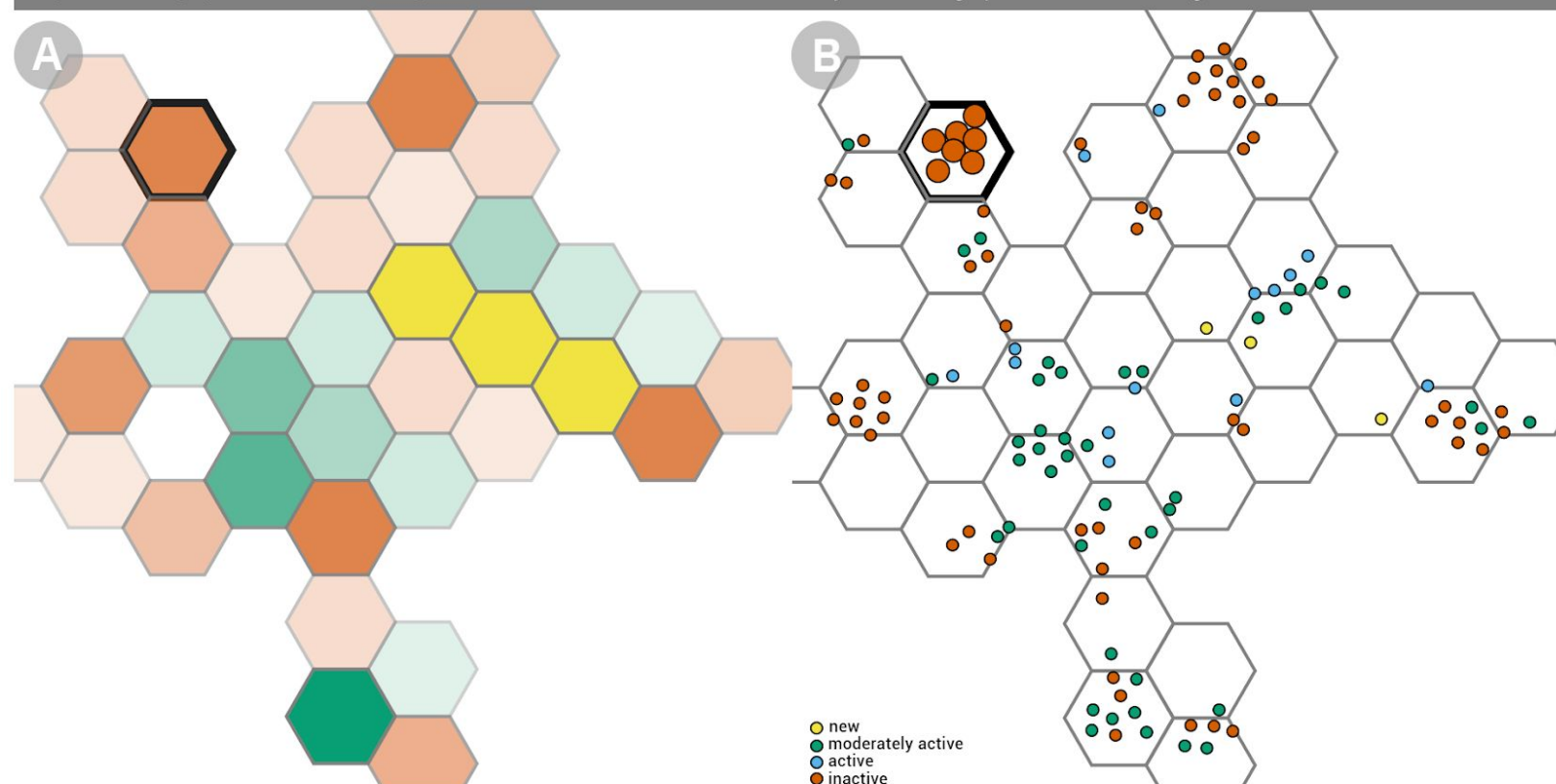


Image source: Ertl and Rohde, The Molecule Cloud - compact visualization of large collections of molecules, Journal of Cheminformatics, 2012

- Views and interactions for efficiently exploring molecular features

Projection: Fingerprints Color Encoding: Serotonin 1a



ChemVA Add New Row Insert New Compound Download Selected SDFs Save Selection Load Selection Change Dataset

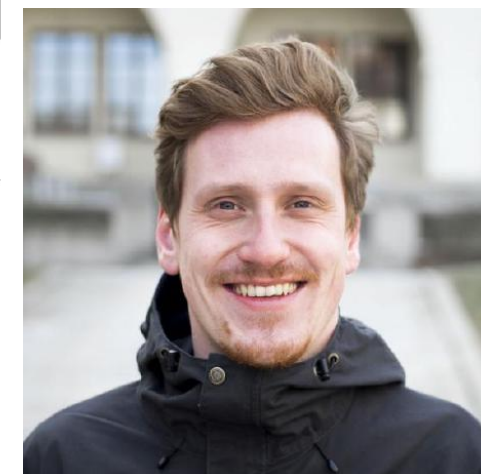
Showing 121 of 121 items; 7 selected

+ Add Column...

Grouping hierarchy

1 ⬢ ⬢ Belongs to Hex

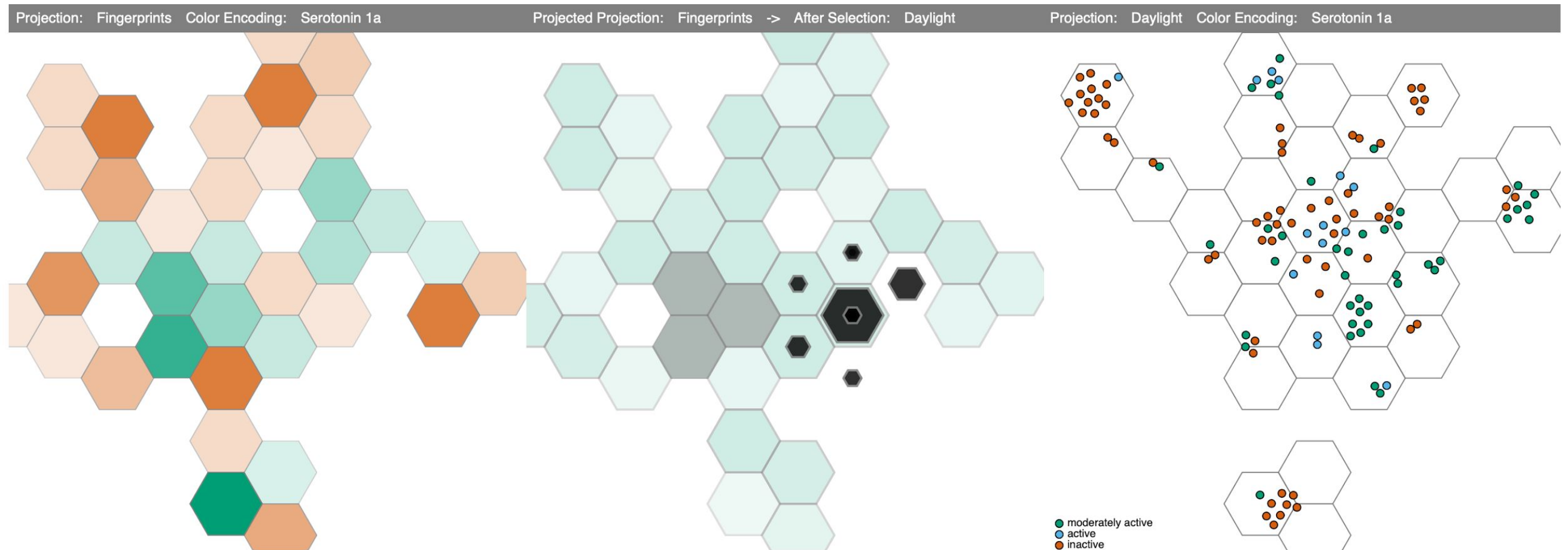
	# ALogP	# Acidic ...	# Basic ...	# Lipins...	Follow...	# Molec...	# QED w...	Seroto...	Dopam...	# Pearso
15	5.50	11.77	5.44	1.00	TRUE	428.41	0.60	—	—	0.31
14	4.32	13.03	5.44	0.00	TRUE	374.40	0.54	—	—	0.50
1	4.66	12.78	1.53	0.00	TRUE	364.79	0.52	—	—	0.48





Hexagonal View & Detail View

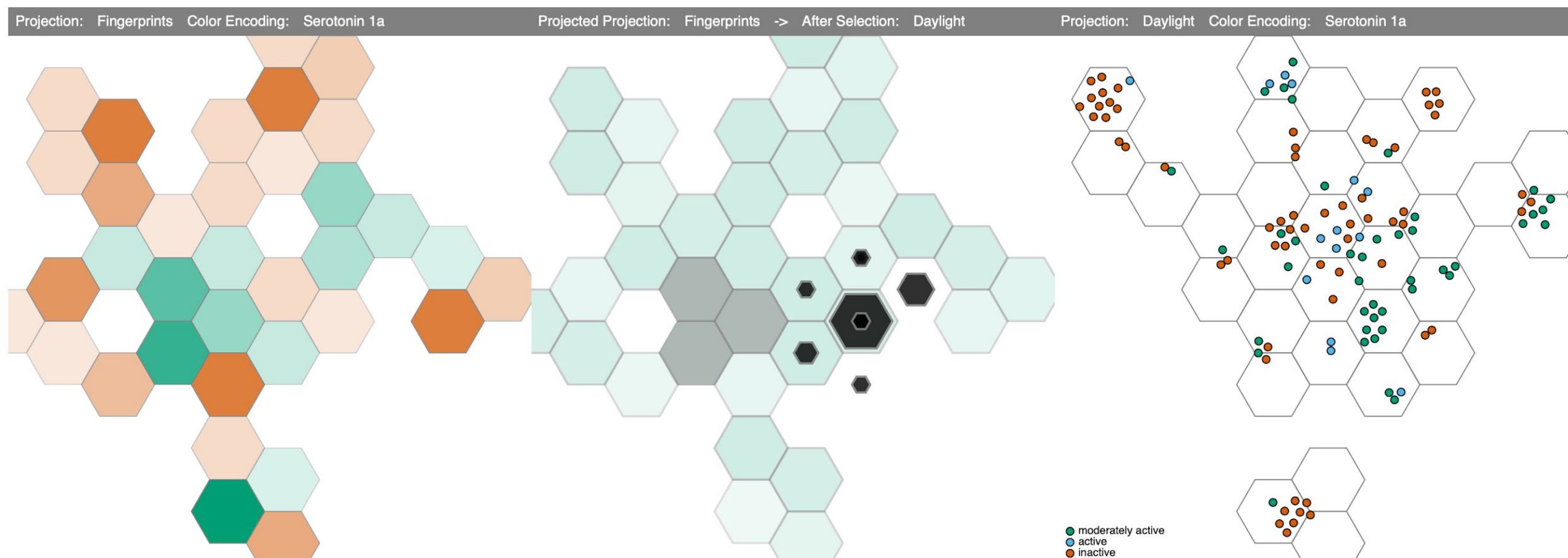
- Data from multiple input sources viewed from different perspectives
 - Variable color and opacity encoding



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

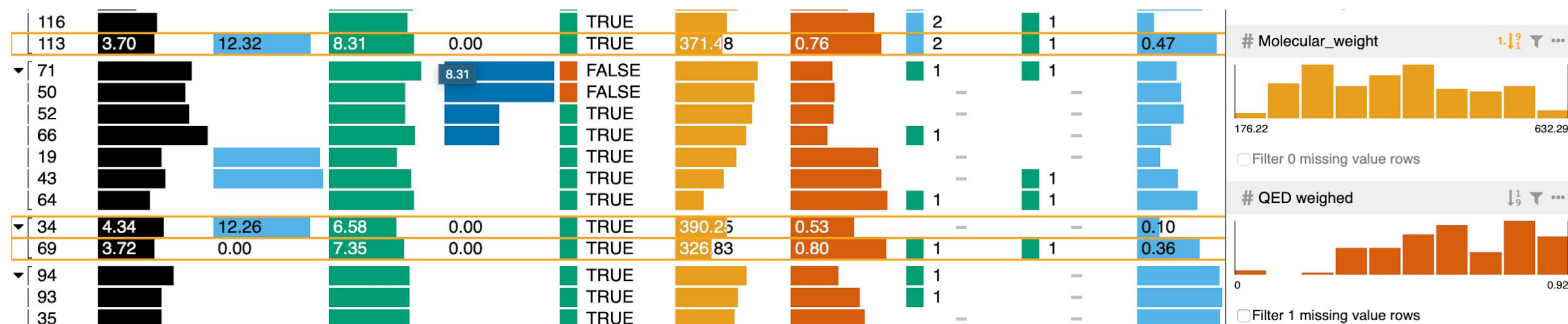
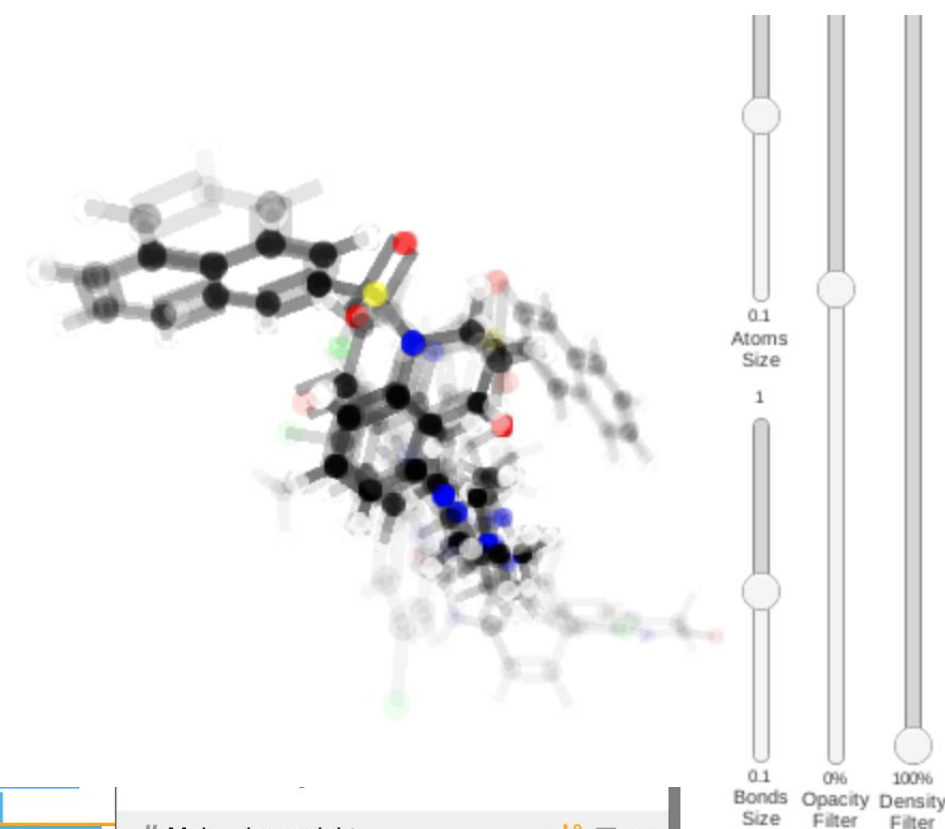
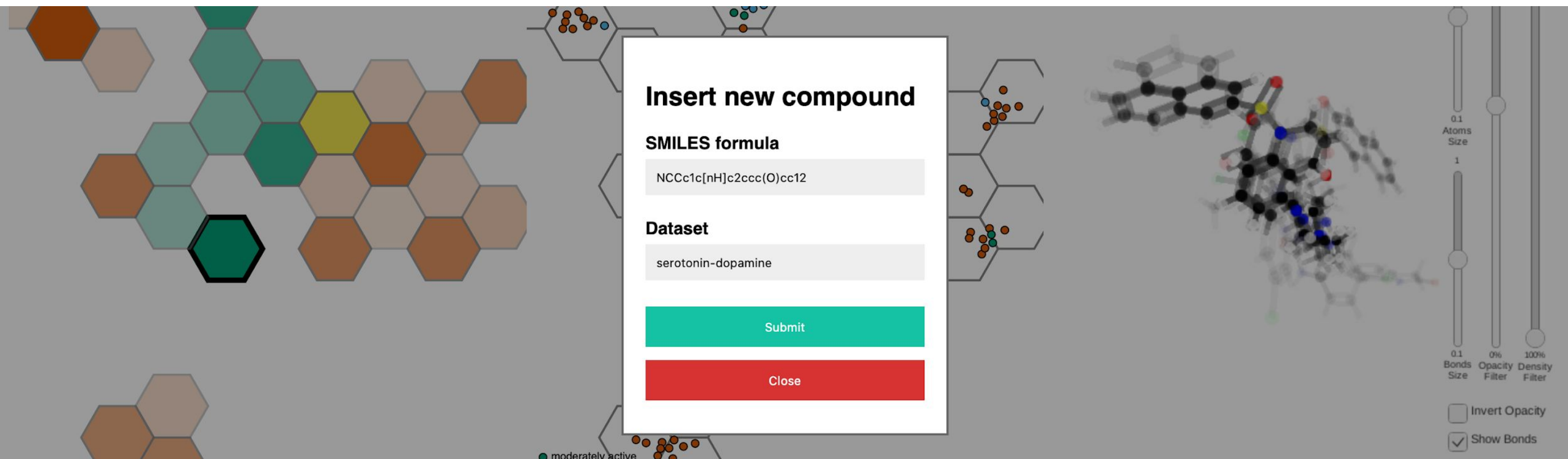


Table View & 3D View

Application Scenario: Adding New Compounds

- Highlighted with yellow color in the dataset

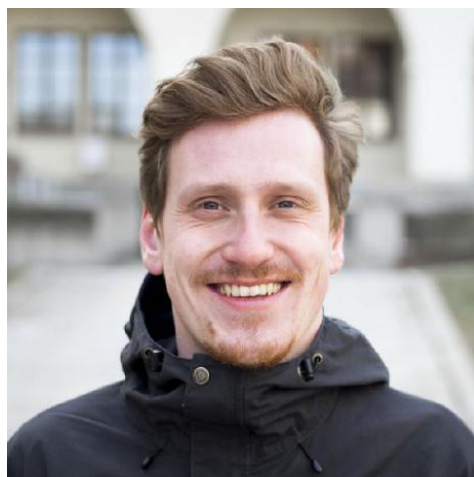


Case Study:

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

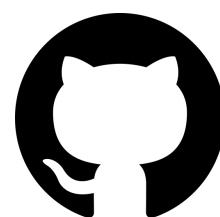


Thank you!



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

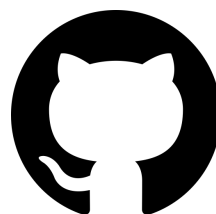


Thank you!



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*

