Kazan Federal UNIVERSITY

Chemoinformatics and molecular modeling lab.

CONDENSED GRAPH OF REACTION - SWISS-KNIFE TOOL FOR REACTION INFORMATICS

Dr. Timur I. Madzhidov

Kazan Federal University, Department of Organic and Medicinal Chemistry

tmadzhidov@gmail.com

A dream…

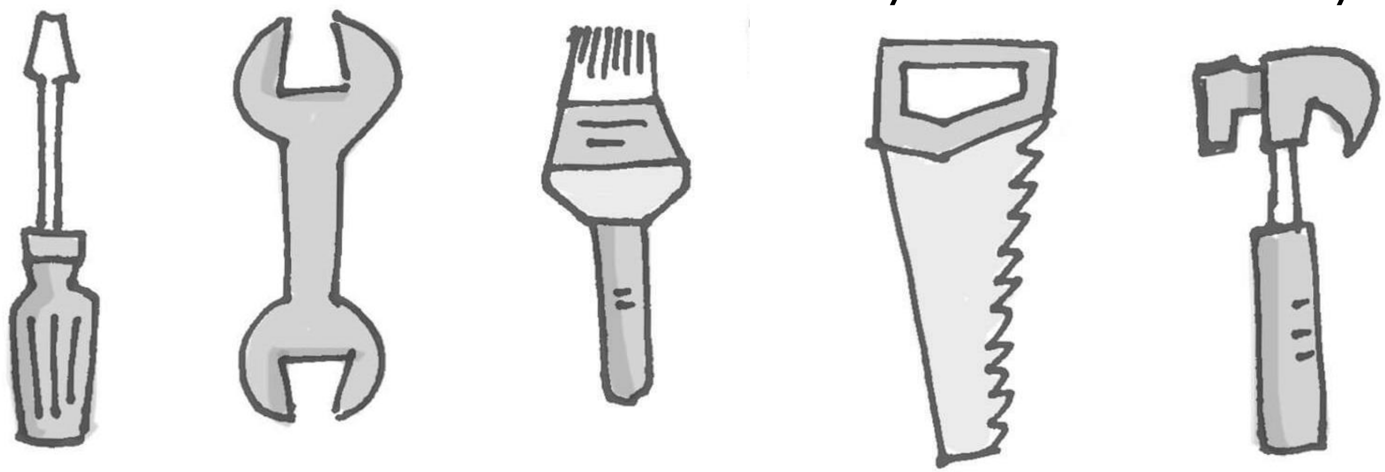
QSAR

SBDD

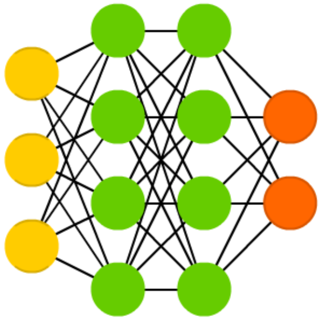
Similarity

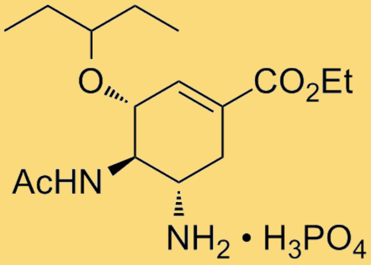
Molecular dynamics

Quantum chemistry

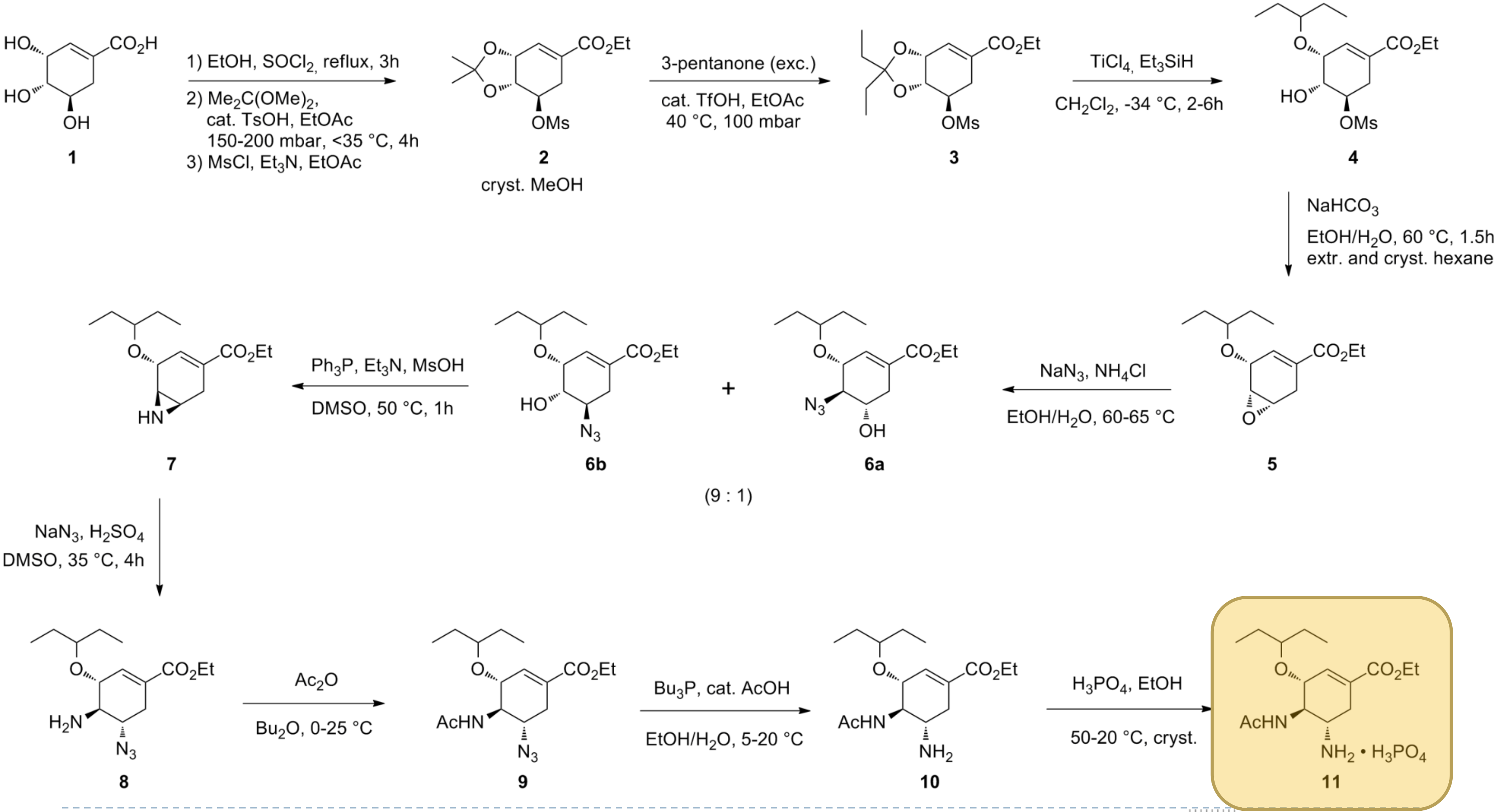


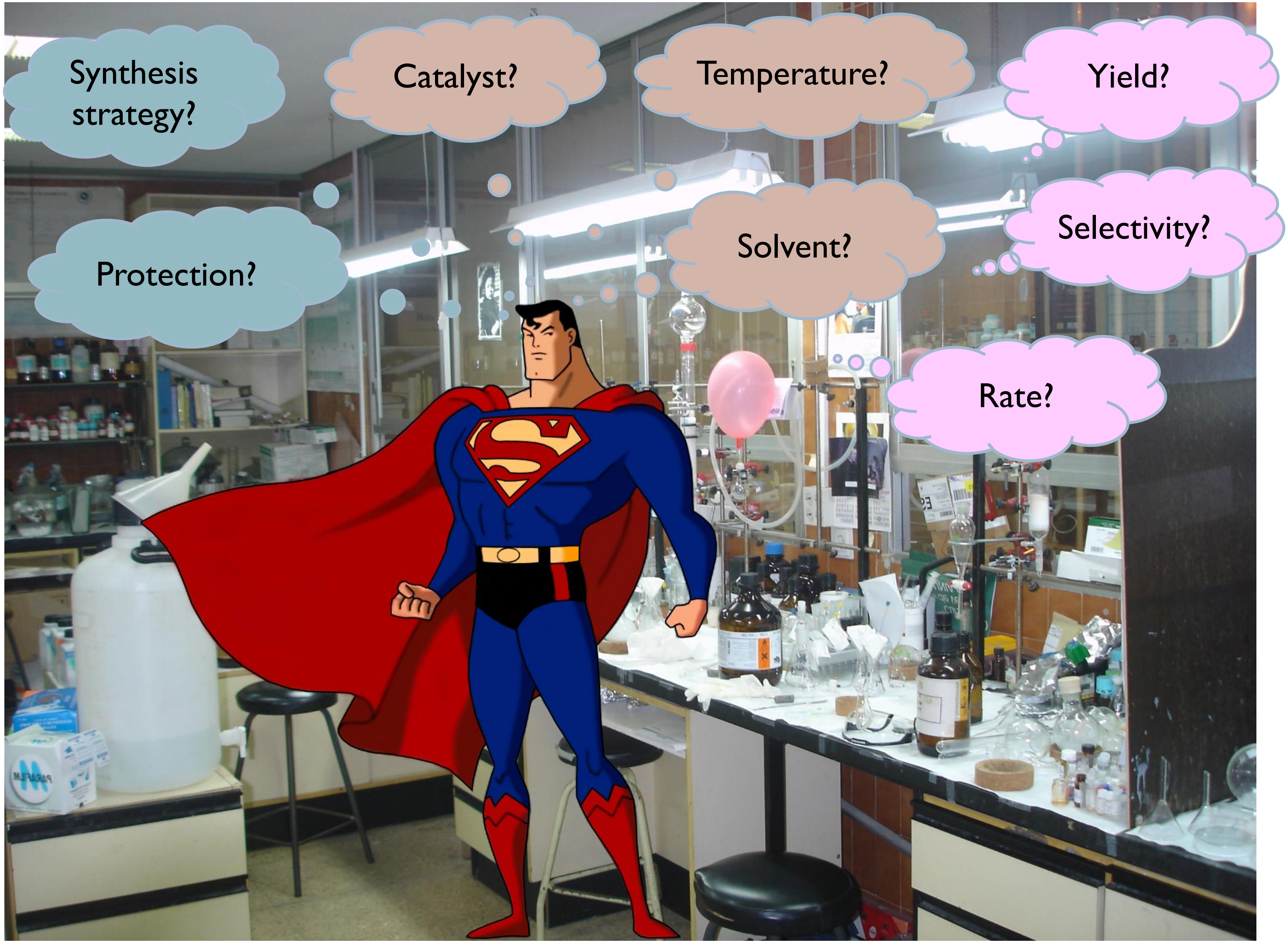
Generative Neural Nets



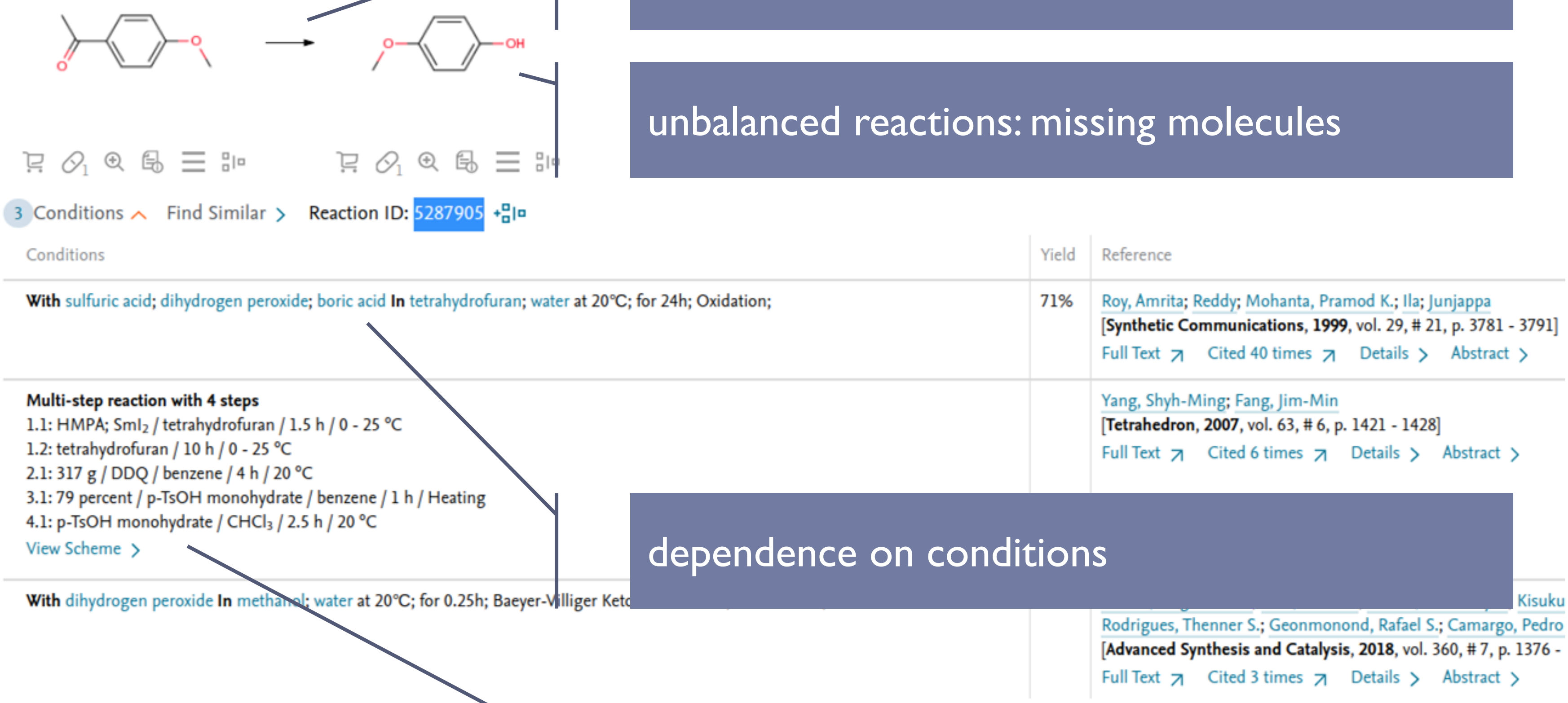


… and reality


1
1) EtOH, $SOCl_2$, reflux, 3h
2) $Me_2C(OMe)_2$,
cat. TsOH, EtOAc
150-200 mbar, <35 °C, 4h
3) MsCl, Et_3N, EtOAc
2
cryst. MeOH
3-pentanone (exc.)
cat. TfOH, EtOAc
40 °C, 100 mbar
3
$TiCl_4$, Et_3SiH
CH_2Cl_2, -34 °C, 2-6h
4
$NaHCO_3$
$EtOH/H_2O$, 60 °C, 1.5h
extr. and cryst. hexane
5
NaN_3, NH_4Cl
$EtOH/H_2O$, 60-65 °C
6a
+
6b
(9 : 1)
Ph_3P, Et_3N, MsOH
DMSO, 50 °C, 1h
7
NaN_3, H_2SO_4
DMSO, 35 °C, 4h
8
Ac_2O
Bu_2O, 0-25 °C
9
Bu_3P, cat. AcOH
$EtOH/H_2O$, 5-20 °C
10
H_3PO_4, EtOH
50-20 °C, cryst.
11
$NH_2 \cdot H_3PO_4$


Synthesis strategy?
Catalyst?
Temperature?
Yield?
Protection?
Solvent?
Selectivity?
Rate?

Reaction is complex



3 Conditions ^ Find Similar > Reaction ID: 5287905

Conditions	Yield	Reference
With sulfuric acid; dihydrogen peroxide; boric acid **In** tetrahydrofuran; water at 20°C; for 24h; Oxidation;	71%	Roy, Amrita; Reddy; Mohanta, Pramod K.; Ila; Junjappa [**Synthetic Communications**, **1999**, vol. 29, # 21, p. 3781 - 3791] Full Text ↗ Cited 40 times ↗ Details > Abstract >
Multi-step reaction with 4 steps 1.1: HMPA; SmI_2 / tetrahydrofuran / 1.5 h / 0 - 25 °C 1.2: tetrahydrofuran / 10 h / 0 - 25 °C 2.1: 317 g / DDQ / benzene / 4 h / 20 °C 3.1: 79 percent / p-TsOH monohydrate / benzene / 1 h / Heating 4.1: p-TsOH monohydrate / $CHCl_3$ / 2.5 h / 20 °C View Scheme >		Yang, Shyh-Ming; Fang, Jim-Min [**Tetrahedron**, **2007**, vol. 63, # 6, p. 1421 - 1428] Full Text ↗ Cited 6 times ↗ Details > Abstract >
With dihydrogen peroxide **In** methanol; water at 20°C; for 0.25h; Baeyer-Villiger Keto		Kisuku Rodrigues, Thenner S.; Geonmonond, Rafael S.; Camargo, Pedro [**Advanced Synthesis and Catalysis**, **2018**, vol. 360, # 7, p. 1376 - Full Text ↗ Cited 3 times ↗ Details > Abstract >

two types of species: reactants and products;

unbalanced reactions: missing molecules

dependence on conditions

multi-step reactions

Condensed Graph of Reaction: why?

Reaction space visualization

Novel reaction prediction

Reaction classification

Data curation

Condition prediction

Reaction search

Rate modeling

Equilibrium constant prediction

CGR: history

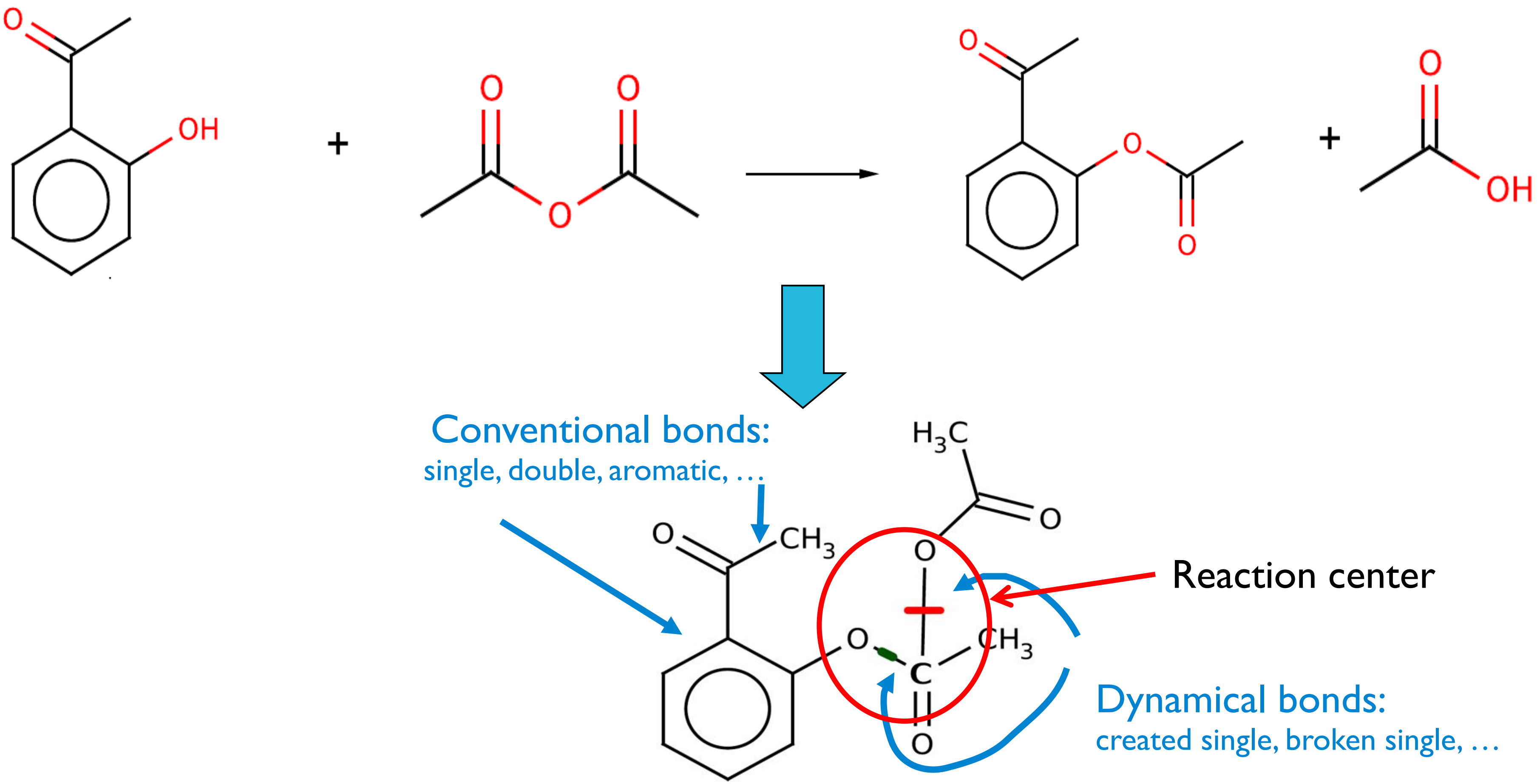
- Yuri KIHO (1972)
- George VLADUTZ (1974) - *Superimposed Reaction Skeleton Graph*
- Shinsaku FUJITA (1986) - *Imaginary Transition Structures*
- Gérard KAUFFMAN (1990) - *Condenced Graph of Reaction*

Reactions in graph-based chemical space → *Reactions classification*
Reaction rules
Synthesis design

- Alexandre VARNEK (2005) - *Condensed Graph of Reaction*

Reactions in descriptors-based chemical space → *Machine-learning models*

Condensed Graph of Reaction



Conventional bonds:
single, double, aromatic, …

Reaction center

Dynamical bonds:
created single, broken single, …

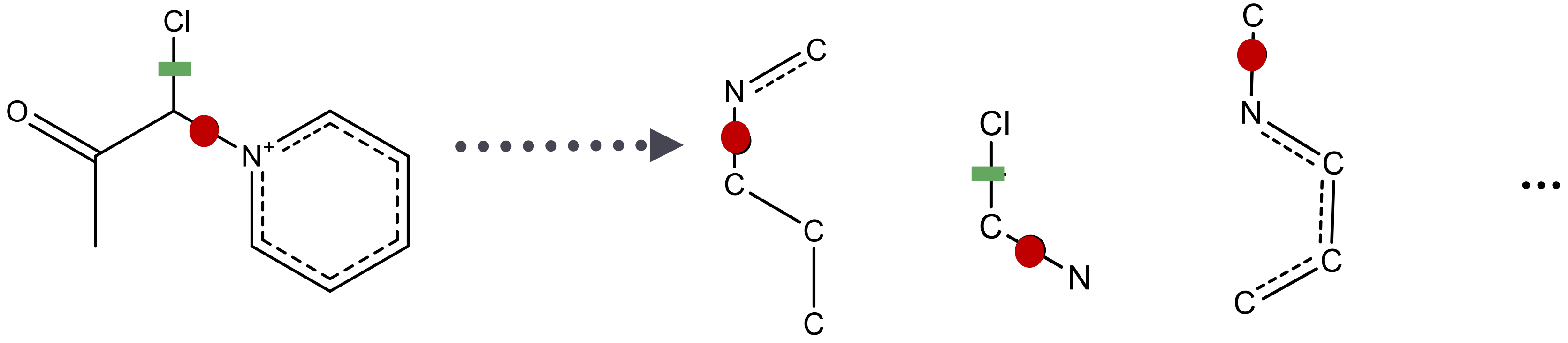
https://github.com/cimm-kzn/CGRtools

Varnek A., et al. (2005) J Comput Aided Mol Des 19:693
Nugmanov, R.I. et al. (2019) JCIM 59: 2516

ISIDA/CGR fragment descriptors

Condensed graph of reaction

ISIDA fragment descriptors



1	1	2	...

Reaction can be encoded by a descriptors vector which can be used in data analysis or in structure-reactivity modeling

A. Varnek In: "Chemoinformatics and Computational Chemical Biology", J. Bajorath, Ed., Springer, 2010

Condensed Graph of Reaction: why?

CGR as graph object

Reaction balancing

Data curation

Reaction search

Novel reaction prediction

Reaction classification

CGR represented by descriptors

Condition prediction

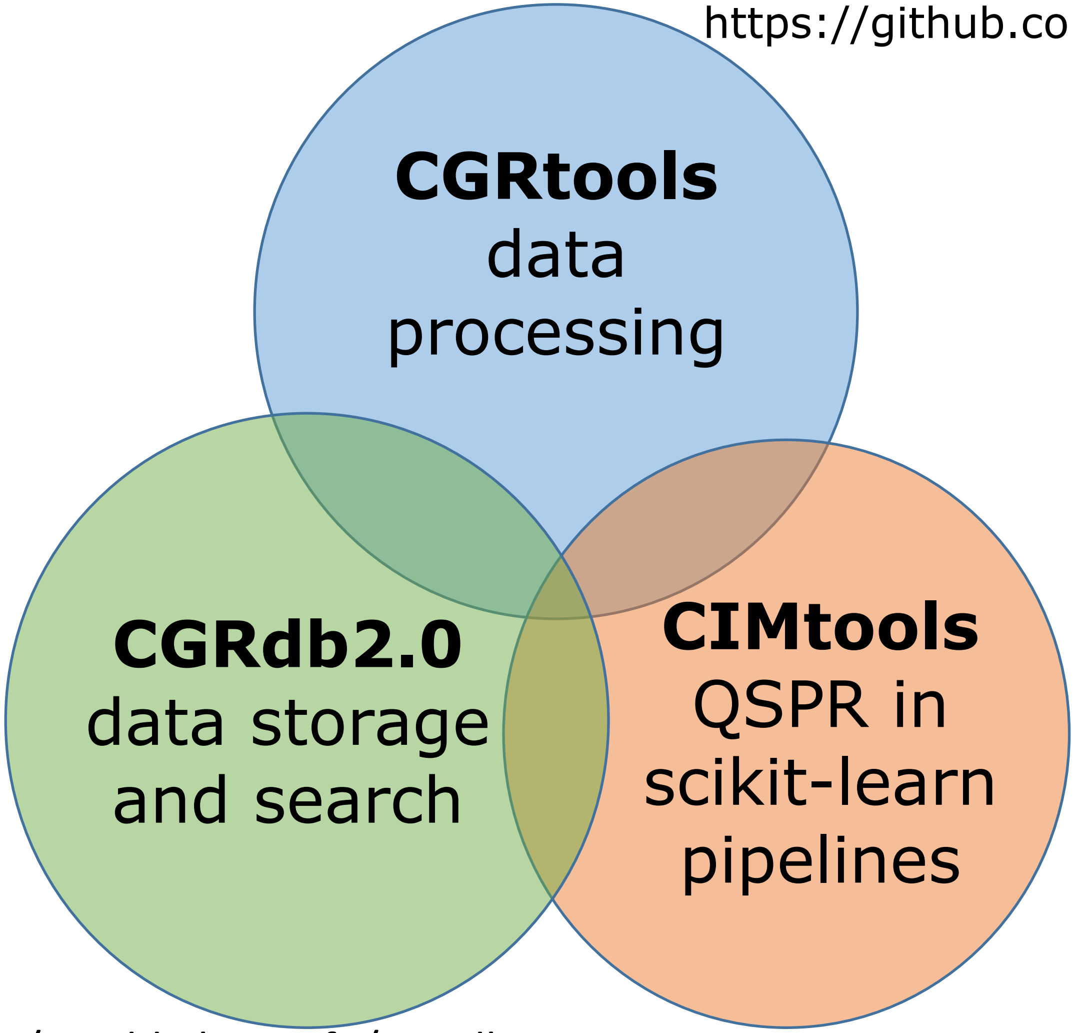
Reaction space visualization

Equilibrium constant prediction

Rate modeling

Tools

https://github.com/cimm-kzn/CGRtools



CGRtools
data
processing

CGRdb2.0
data storage
and search

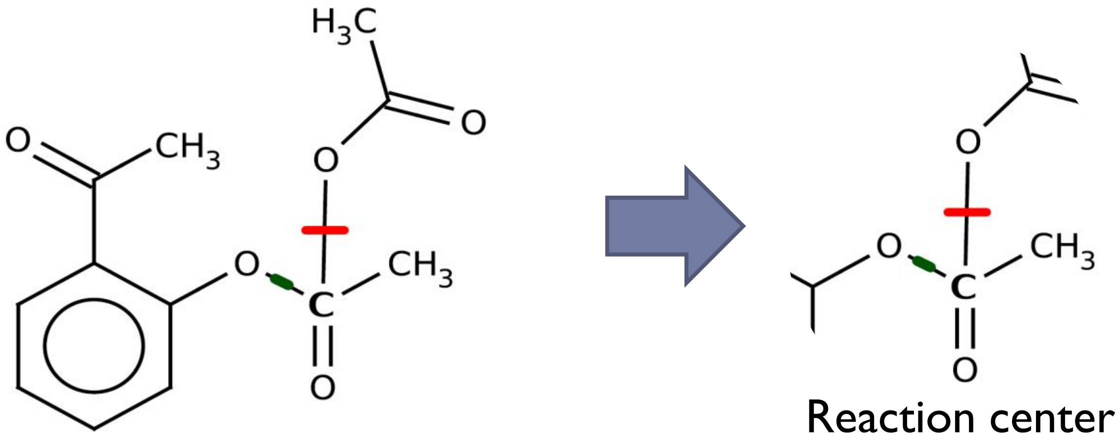
CIMtools
QSPR in
scikit-learn
pipelines

https://github.com/icredd-cheminfo/CGRdb2

https://github.com/cimm-kzn/CIMtools

CGR as graph object

Reaction centers as reaction type markers



Reaction center

Signatures for reaction classification

Baskin I.I. et al. Russ. Chem. Rev. 86, 1127 (2017)

Delannée, V. et al. J Cheminform 12, 72 (2020).

AAM Fixing

Lin A. et al. Mol. Inform. 2100138 (2021)

Retrosynthetic rule extraction

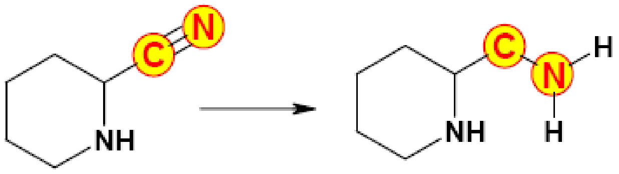
Segler M. et al.Nature. 555, 7698 (2018)

Applicability domain - RTC

Rakhimbekova A. et al. Int. J. Mol. Sci. 21, 5542 (2020)

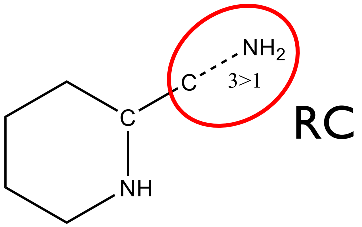
Different levels of reaction centers

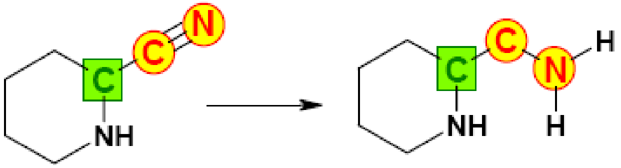
ICClassify (*InfoChem*)



0-Sphere (BROAD)

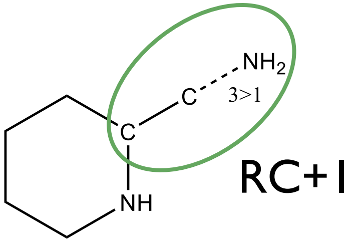
Reaction centers only

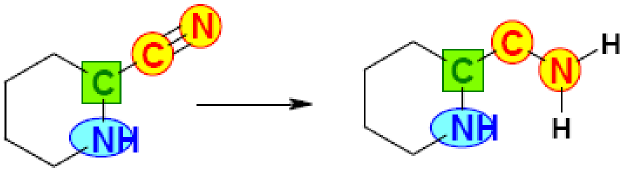




1-Sphere (MEDIUM)

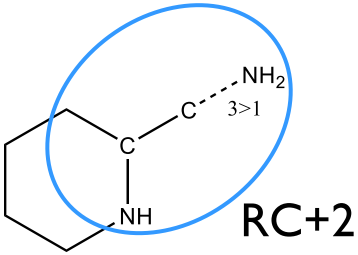
Reaction centers plus alpha atoms, excluding hydrogens





2-Sphere (NARROW)

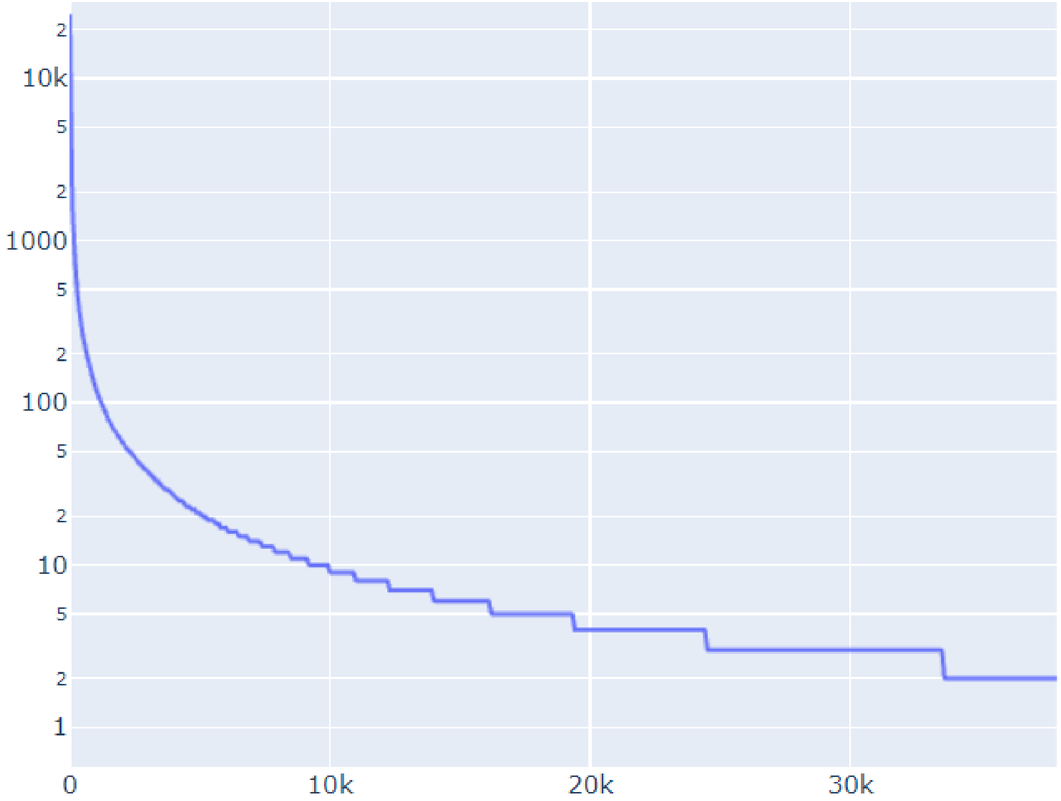
Reaction centers plus beta atoms, excluding hydrogens and consecutive sp^3-atoms



H. Krout et al. J. Chem. Inf. Model. 2013, 53 (11), 2884–2895

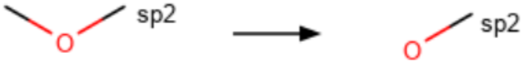
Baskin I.I. et al. Russ. Chem. Rev. 86, 1127 (2017)

Reaction types in USPTO database



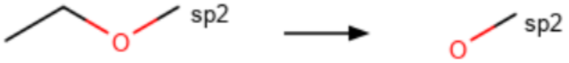
Top-3 (RC+1)

Count



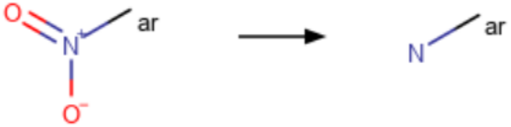
24361

Hydrolysis of methyl ester



20857

Hydrolysis of ethyl ester

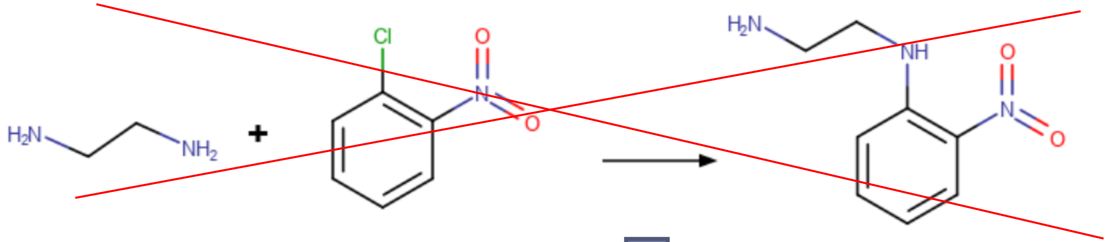


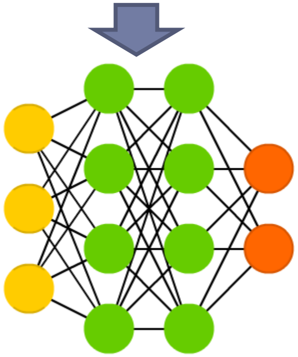
20240

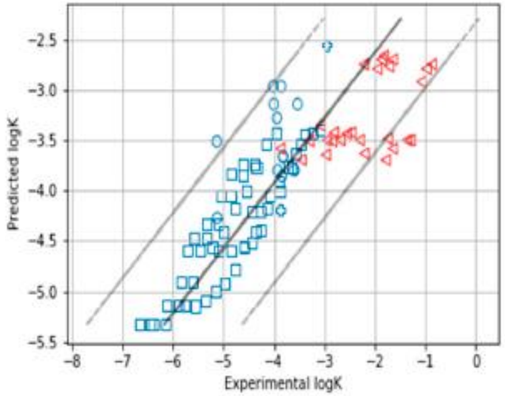
Aromatic nitro reduction

- 219K "RC+1" motifs were found in 1,36M reactions
- 1063 motifs occur in ≥100 reactions

Reaction centers as applicability domains







S_N2 reaction rate dataset

Reaction centers

Model for S_N2 reaction rate prediction

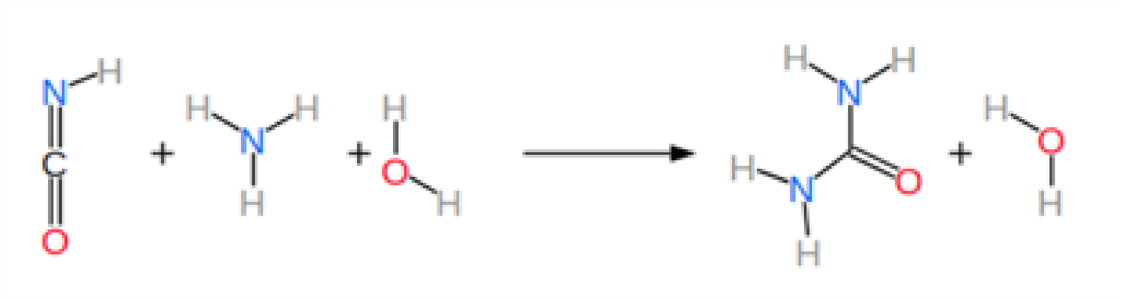
Reaction type control AD

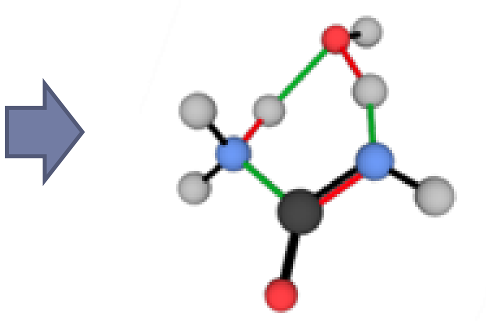
Model = RF
AD = RF variance + RTC

Rakhimbekova A. et al. Int. J. Mol. Sci. 21, 5542 (2020)

CGR can be used for…

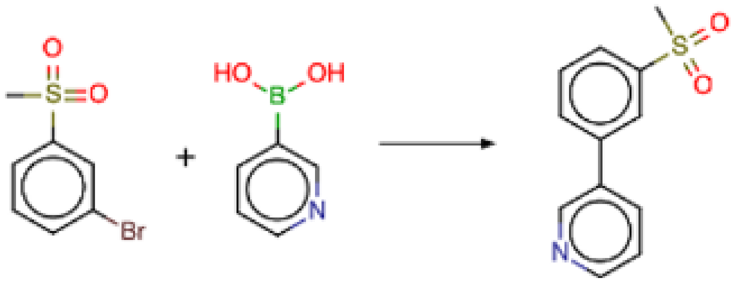
3D CGR proposed for Transition State storage and visualization

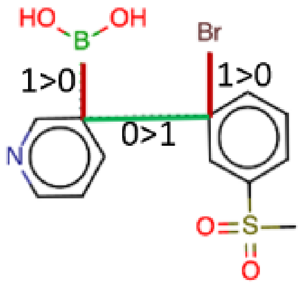




T. Gimadiev, et al. J. Chem. Inf. Model., 2021, **61**, 554.

CGR SMILES as reaction representation

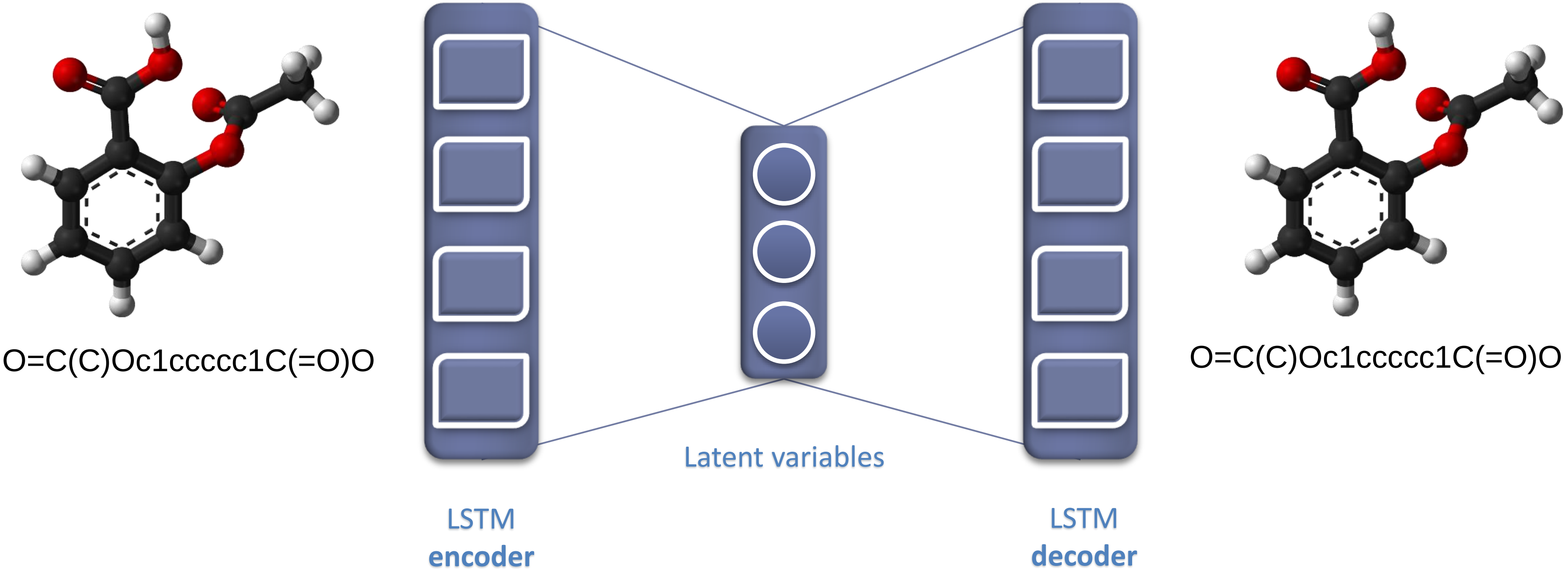




OB(O)[->.]C1(:C:N:C:C:C1)[.>-]C2([->.]Br):C:C:C:C(:C2)S(=O)(C)=O

W. Bort, et al.. Sci. Rep. 2021, **11**, 3178.

Autoencoder performing SMILES reconstruction


O=C(C)Oc1ccccc1C(=O)O

LSTM
encoder

Latent variables

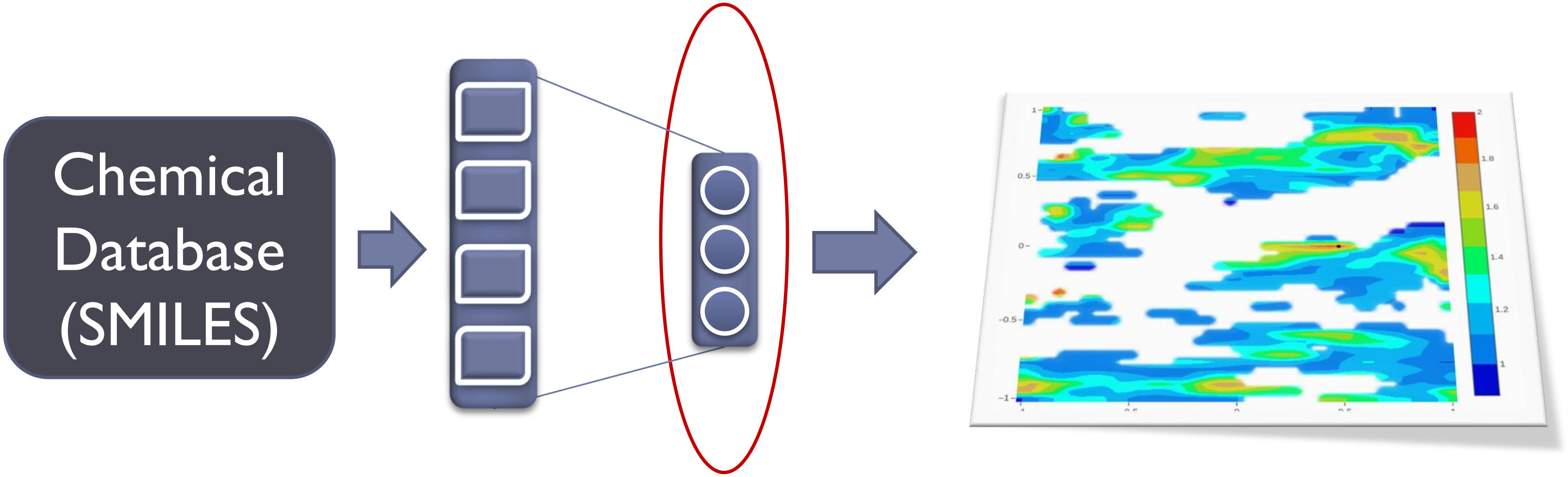
LSTM
decoder

O=C(C)Oc1ccccc1C(=O)O

Chemical structure → **Real numbers encoding** → **Chemical structure**

Building GTM on latent variables of autoencoder

Latent variables
(*vector on real numbers*)



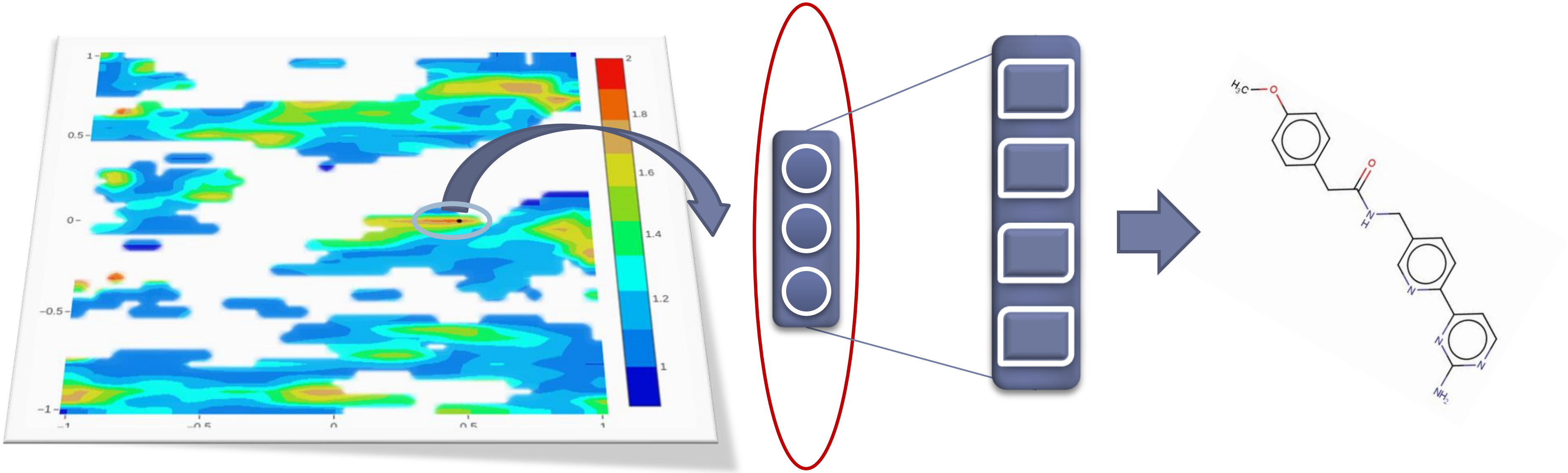
Trained Encoder

GTM

B. Sattarov et al. J. Chem. Inf. Model., 2019, 59(3), 1182-1196

Generation of novel structures from specific areas of the map

Latent variables



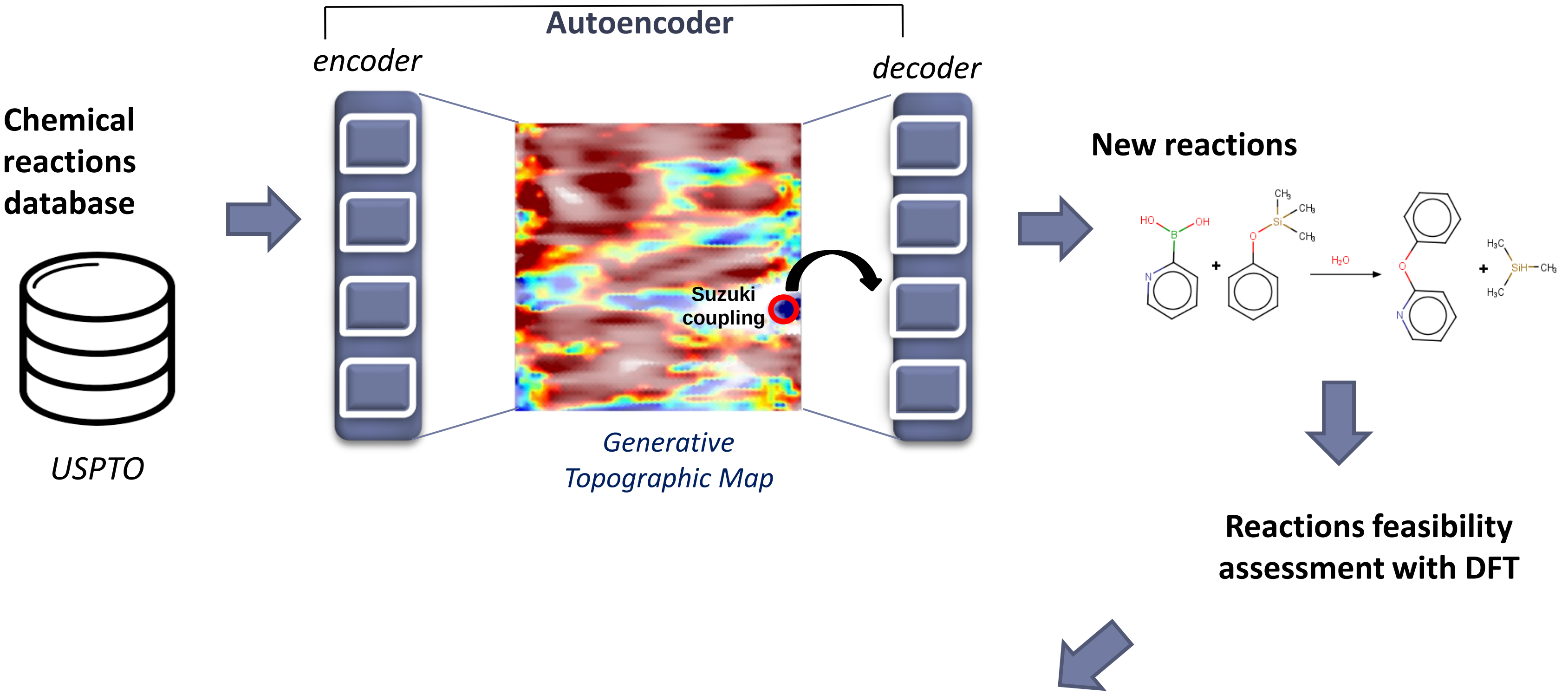
GTM activity landscape

Trained Decoder

SMILES

B. *Sattarov et al. J. Chem. Inf. Model.*, 2019, 59(3), 1182-1196

AI-driven design of new Suzuki-like reactions



Chemical reactions database

USPTO

Autoencoder

encoder

decoder

Suzuki coupling

Generative Topographic Map

New reactions

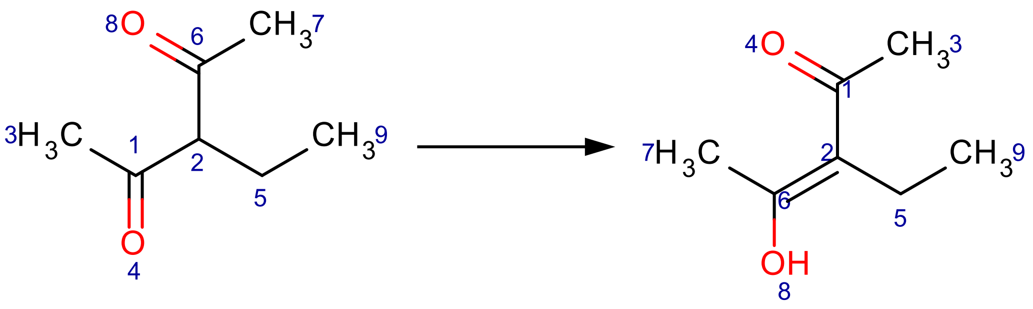
Reactions feasibility assessment with DFT

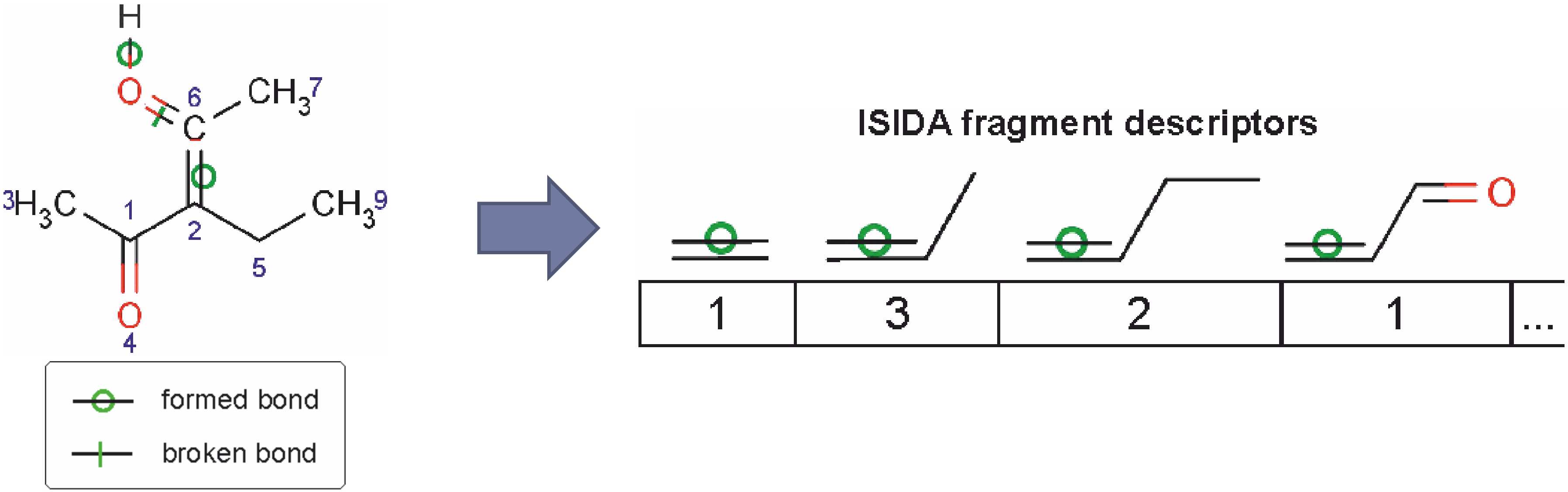
- **13 new (with respect to the training data) Suzuki-like reactions have been detected**
- **5 of them have been found in recent publications**

W. Bort *et al.*, *Nature Scientific Reports*, 2021, 11, 3178

CGR encoded by descriptors

General concept



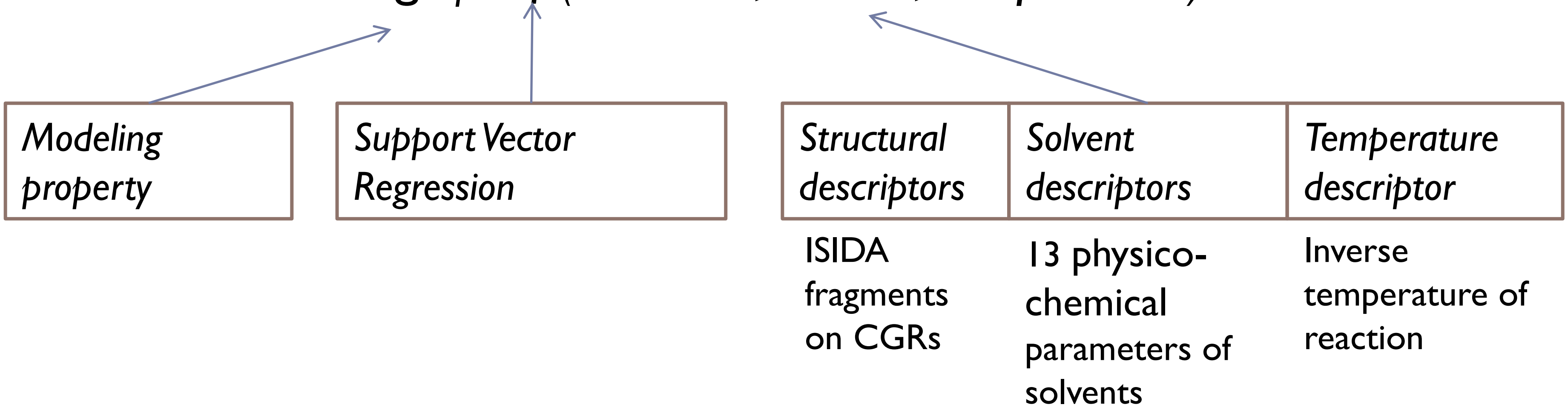


A. Varnek, D. Fourches, F. Hoonakker, V. P. Solov'ev. *J. Computer-Aided Molecular Design,* 2005, 19, 693-703.

General workflow of “reaction QSAR”

Quantitative Structure-Reactivity Relationships

$$\log K_T = f\,(\mathit{structure, solvent, temperature})$$



Modeling property

Support Vector Regression

Structural descriptors

ISIDA fragments on CGRs

Solvent descriptors

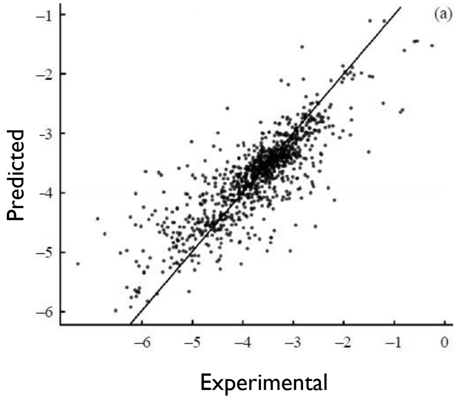
13 physico-chemical parameters of solvents

Temperature descriptor

Inverse temperature of reaction

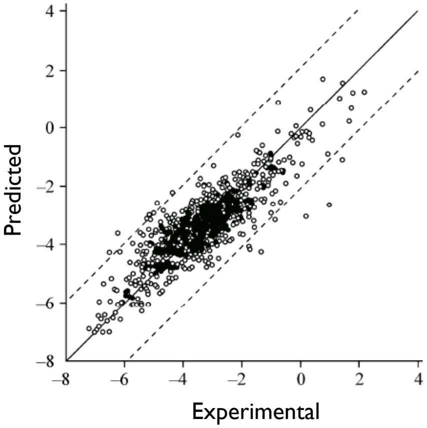
Reaction rate and equilibrium constant prediction

S_N2 reaction rate constant



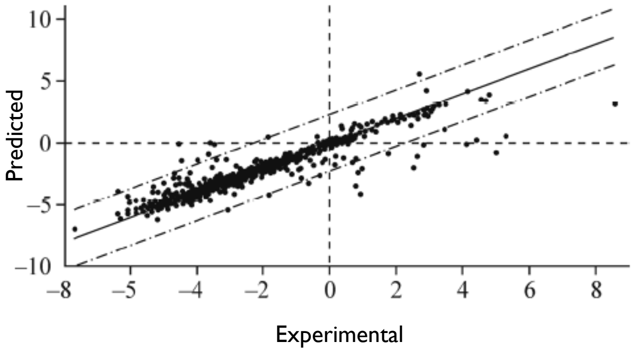
Russ. J. Org. Chem, 2014, 50, 47

E2 reaction rate constant



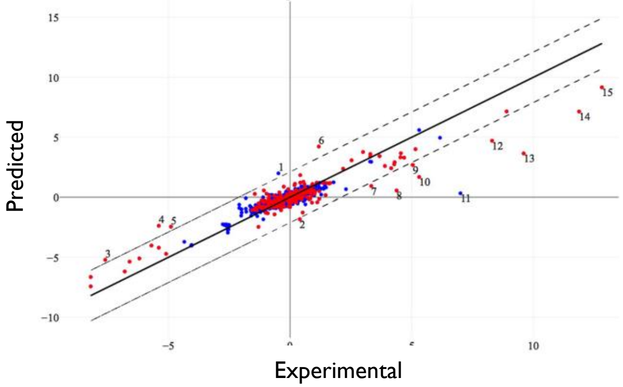
Russ. J. Struct. Chem., 2015, 56, 1080

Cycloaddition rate constant



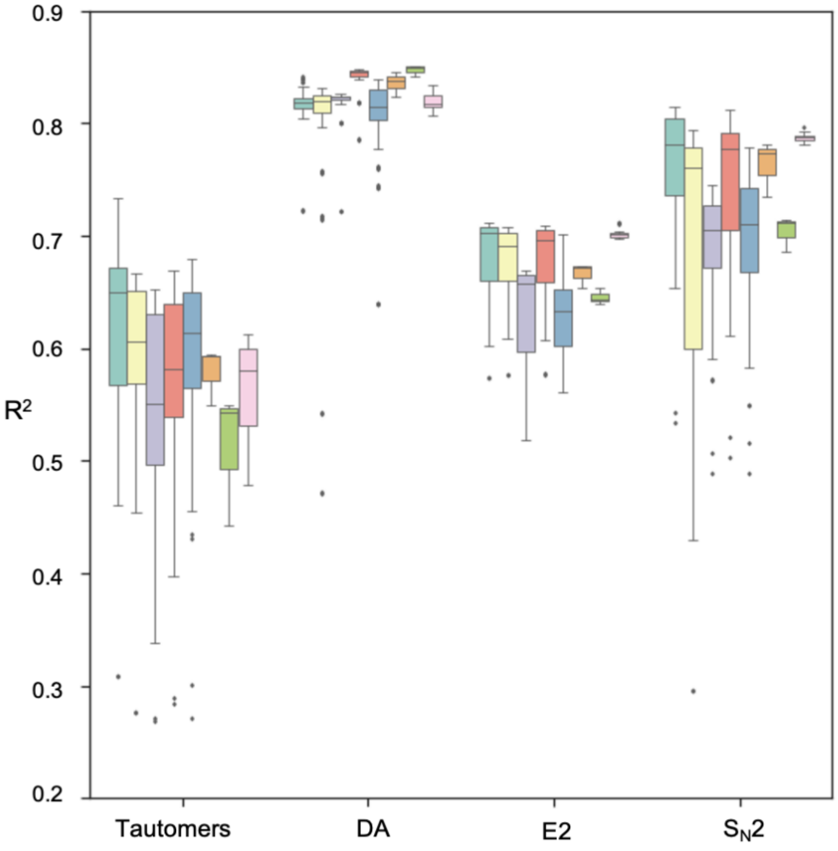
Russ. J. Struct. Chem., 2017, 58, 650

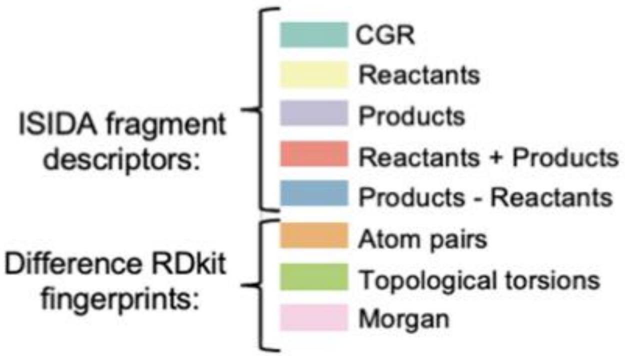
Tautomerisation equilibrium constants



T. Gimadiev, **T. Madzhidov**, R. Nugmanov, I.I. Baskin, I.S. Antipin, A. Varnek.
Journal of Computer-Aided Molecular Design, 2018, 32, 401

Benchmark of reaction descriptors





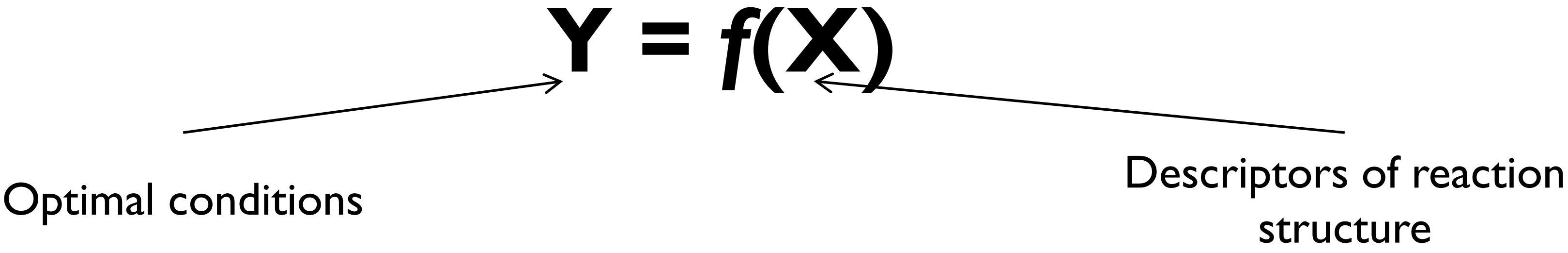
Models for

- reaction rate constant of Diels-Alder, S_N2 and E2 reactions
- tautomeric equilibrium constant

CGR descriptors were used in top ranked models

Rakhimbekova A. et al Mendeleev Commun., 2021, 31, 769–780

Reaction condition prediction complexity



$$\mathbf{Y} = \boldsymbol{f}(\mathbf{X})$$

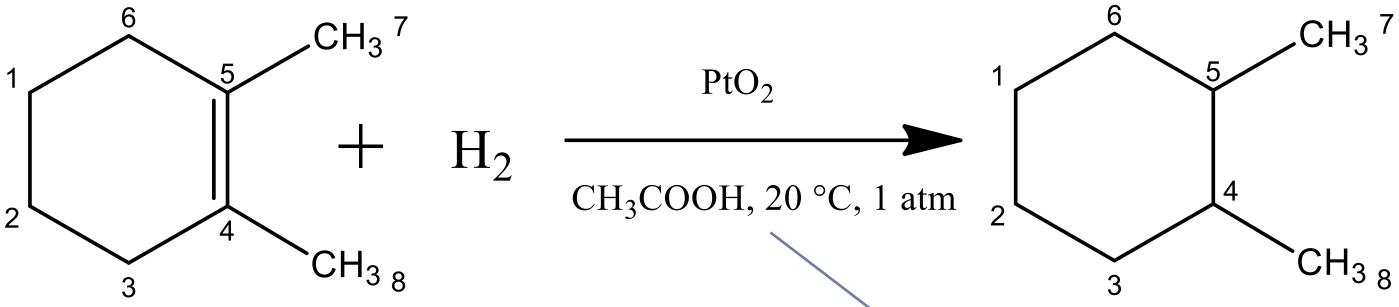
Optimal conditions

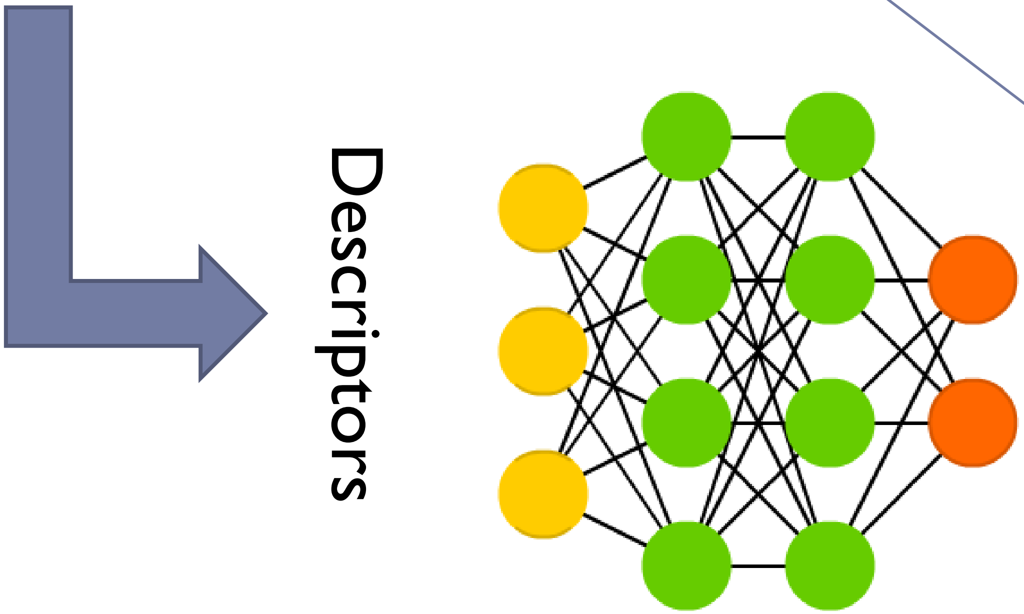
Descriptors of reaction structure

The same reaction could go at several conditions!

No knowledge which conditions are not good for particular reaction!

Condition modelling as ranking





1) (PtO$_2$, acid, 20 °C, 1atm)
2) (Pd/C, 20 °C, 1 atm)
3) (Pt, 20 °C, 1 atm)

....

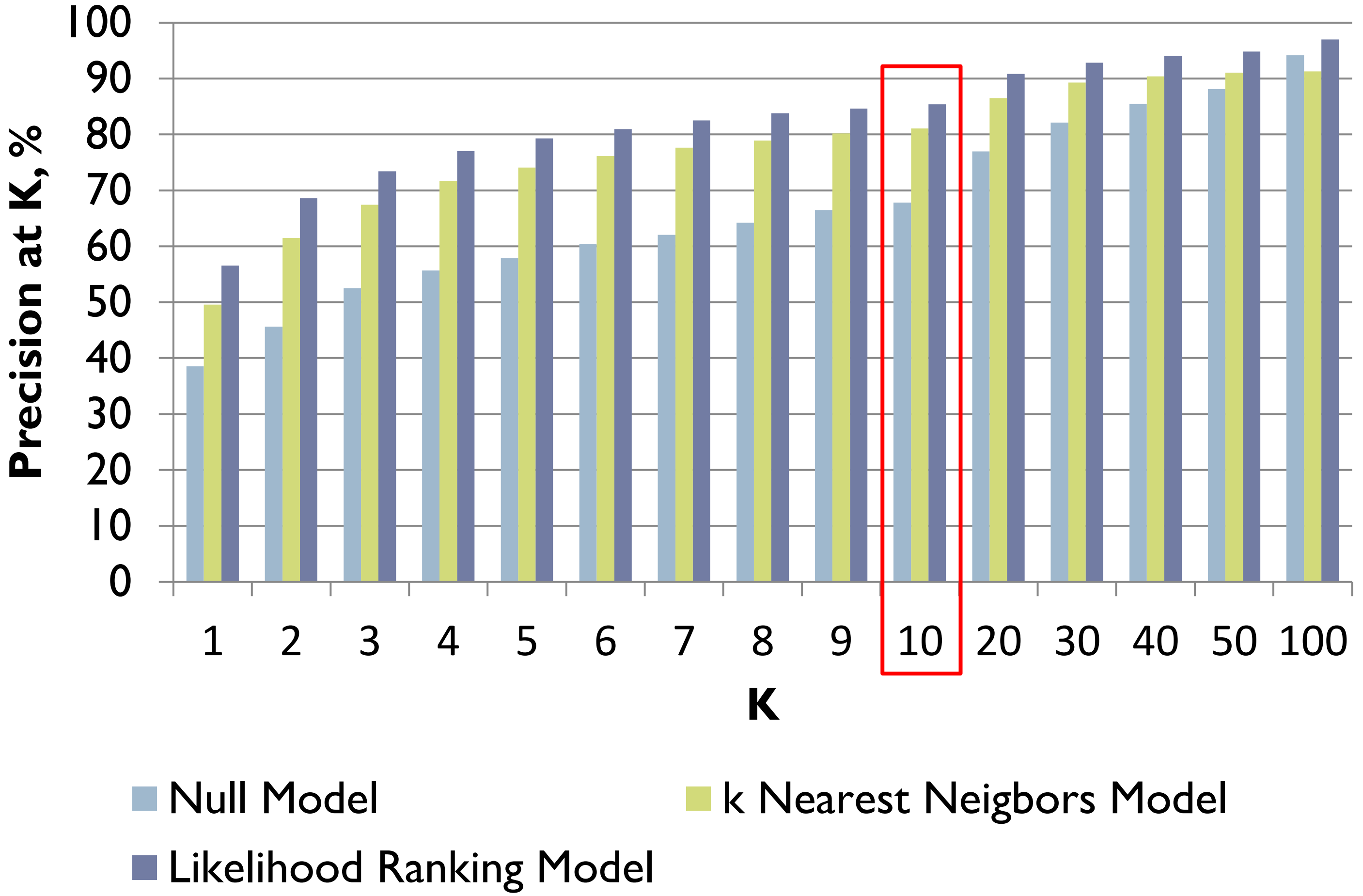
....

Afonina, V.A.; et al *Int. J. Mol. Sci.* **2022**, *23*, 248.
https://doi.org/10.3390/ijms23010248

Model performance

Dataset: ~90 000 hydrogenation reactions

Reaxys®



Afonina, V.A.; et al *Int. J. Mol. Sci.* **2022**, *23*, 248.
https://doi.org/10.3390/ijms23010248

Conclusions

- CGR is universal approach for reaction representation solving most of their complexity
- CGR can be manipulated as graphs or can be used for descriptor calculations
- CGRs as graph can be utilized for AAM check or correction, missing molecules identification, data curation, and effective applicability domain for reaction characteristics prediction
- CGRs can be encoded by SMILES and be coupled with generative neural networks for novel reaction generation
- CGR-based fragment descriptors can be applied for reaction characteristics modeling, condition recommendation, reaction space visualization

Acknowledgements

Kazan Federal University
Prof. Igor Antipin
Dr. Ramil Nugmanov (Janssen)
Dr. Timur Gimadiev (BIOCAD)
Dr. Marta Glavatskikh (UNC)
Valentina Afonina
Adelia Fatykhova
Asima Rakhimbekova
Dmitrii Zankov (UniStra)
Aigul Khakimova (BIOCAD)
Artem Kokorin (UniLux)
Ravil Mukhametgaliev
Tagir Akhmetshin (UniStra)
Etc…

University of Strasbourg
Prof. Alexandre Varnek
Dr. Gilles Marcou
Dr. Dragos Horvath
Dr. Olga Klimchuk
Dr. Fanny Bonachera
Dr. Arkadii Lin (InSilico)
William Bort
Iuri Casciuc (Syngenta)
Boris Sattarov (Qubit Pharma)

Hokkaido University
Dr. Pavel Sidorov

University of Olomouc
Dr. Pavel Polishchuk
Mariia Matveyeva
Alexandra Nikonenko

Technion, Israel
Dr. Igor Baskin

Janssen Pharmaceutica
Dr. Natalia Dyubankova
Dr. Jonas Verhoeven
Dr. Joerg Wegner

Elsevier (Reaxys)
Dr. Elena Herzog
Dr. Marcus Fischer

РНФ | Российский научный фонд
Projects 19-73-10137, 14-43-00024

Министерство науки и высшего образования РФ, соглашение 14.587.21.0049 (проект RFMEFI58718X0049)