Pyridoxal derivatives as anolytes for aqueous organic redox flow batteries: Computational screening

A Hamza, Á. Madarász, F. B. Németh, A. Nechaev, P. M. Pihko, P. Peljo, I. Pápai

CompBat WP1

CompBat: Developing tools for discovery of new prospective candidates for next generation RFBs

WP1: High-throughput screening and machine learning

Main objective: Development of a HTS methodology for identification of promising RFB compounds

Strategy: DFT molecular database → assessment of ML methods → iterative expansion of database

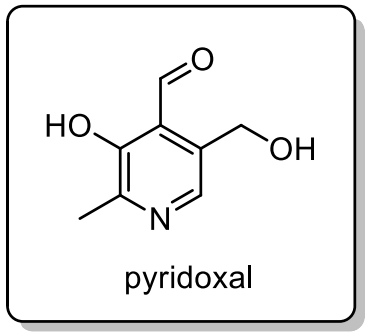
HTS requirements: large database – reliable data – automatization → efficient computational protocol

Partners involved: TTK (RCNS), Aalto, JYU, UTU

Pyridoxal database

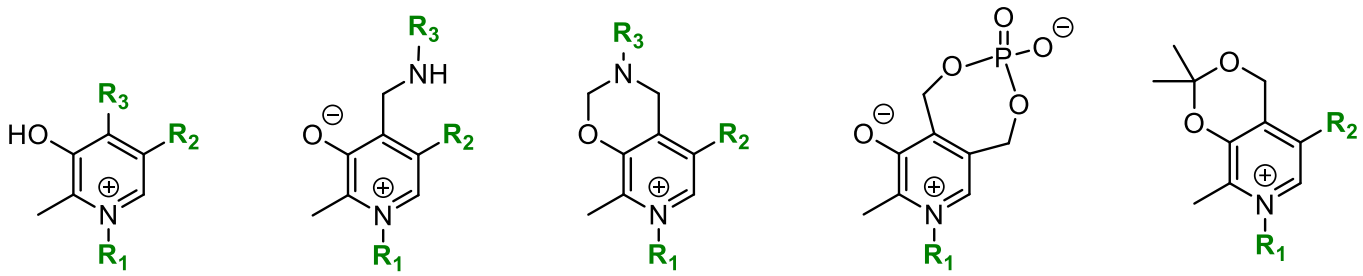
Pyridoxal database

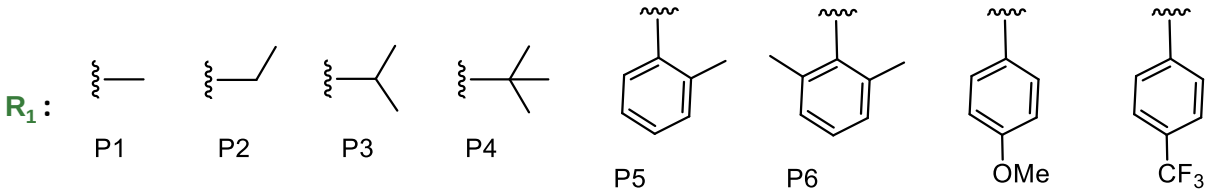
bioinspired AO RFBs



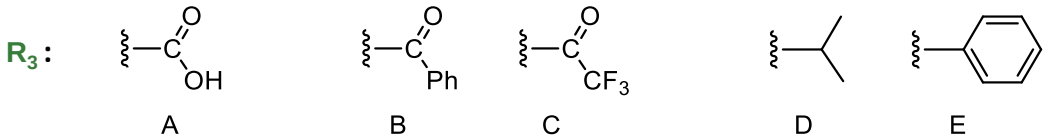
(vitamin B_6)

6712 molecules





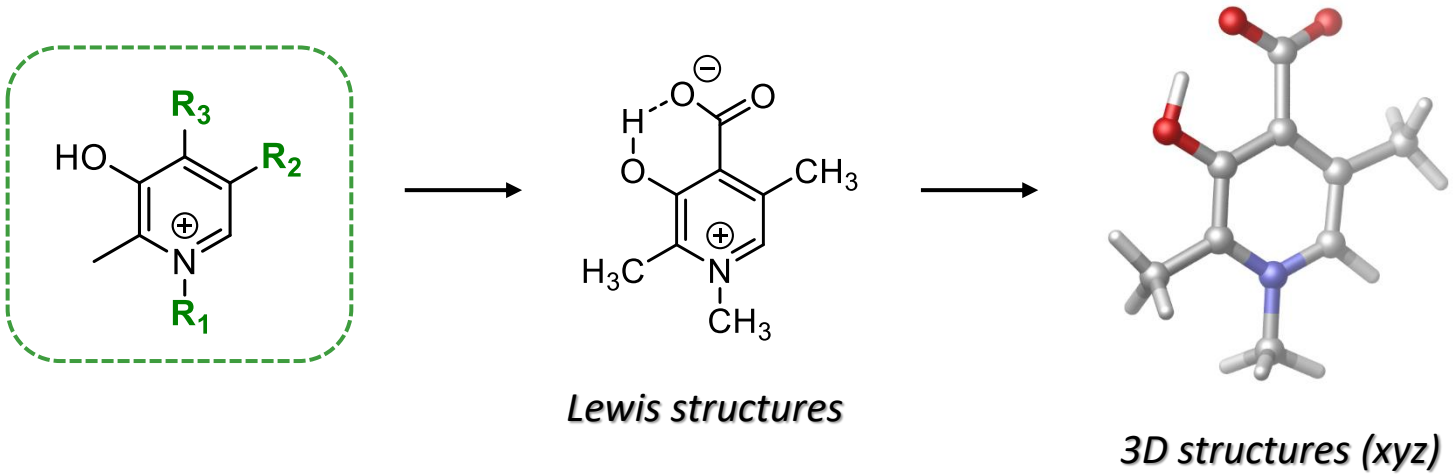
R_2 : 113 different substituents (a1, a2, ..., b1, b2, ..., etc)



(in pyr1) (in pyr1 and pyr2) (in pyr3)

Computational protocol

a) **<u>initial 3D structures</u>** – automatic generation of 2D (Lewis) structures (***CombiGlide***), conversion to 3D (***LigPrep***); protonation state at pH = 7 in water (***Epic***)



Lewis structures

3D structures (xyz)

Computational protocol

a) **initial 3D structures** – automatic generation of 2D (Lewis) structures (***CombiGlide***), conversion to 3D (***LigPrep***); protonation state at pH = 7 in water (***Epic***)

b) **conformational analysis** – geometry optimization at semiempirical GFN2-xTB level (***xtb*** package); finite temperature and implicit solvent effects (room temperature, water); CREST procedure for extensive conformational search (***crest*** utility) → most stable conformers

preoptimization

GFN-xTB method (semiempirical)

Grimme et al., *JCTC* **2017**, *13*, 1989;
Grimme et al., *JCTC* **2019**, *15*, 1652

comformational search

CREST method (Conformer–**R**otamer **E**nsemble **S**ampling **T**oo**)**

Grimme et al., *PCCP* **2020**, *22*, 7169

⟶ *the most stable structure*

Computational protocol

a) **initial 3D structures** – automatic generation of 2D (Lewis) structures (***CombiGlide***), conversion to 3D (***LigPrep***); protonation state at pH = 7 in water (***Epic***)

b) **conformational analysis** – geometry optimization at semiempirical GFN2-xTB level (***xtb*** package); finite temperature and implicit solvent effects (room temperature, water); CREST procedure for extensive conformational search (***crest*** utility) → most stable conformers

c) **electronic energies** – DFT calculations at M06-2X/6-311+G** level (***Gaussian***), single-point calculations in solution phase (SMD) → aqueous phase solvation free energies

Computational protocol

a) **initial 3D structures** – automatic generation of 2D (Lewis) structures (***CombiGlide***), conversion to 3D (***LigPrep***); protonation state at pH = 7 in water (***Epic***)

b) **conformational analysis** – geometry optimization at semiempirical GFN2-xTB level (***xtb*** package); finite temperature and implicit solvent effects (room temperature, water); CREST procedure for extensive conformational search (***crest*** utility) → most stable conformers

c) **electronic energies** – DFT calculations at M06-2X/6-311+G** level (***Gaussian***), single-point calculations in solution phase (SMD) → aqueous phase solvation free energies

d) **Gibbs free energies** – $G^{aq}_{red/ox} = E^{DFT}_{red/ox} + \Delta G^{xtb}$ → redox potentials according to the Nernst formula

Computational protocol

a) **initial 3D structures** – automatic generation of 2D (Lewis) structures (***CombiGlide***), conversion to 3D (***LigPrep***); protonation state at pH = 7 in water (***Epic***)

b) **conformational analysis** – geometry optimization at semiempirical GFN2-xTB level (***xtb*** package); finite temperature and implicit solvent effects (room temperature, water); CREST procedure for extensive conformational search (***crest*** utility) → most stable conformers

c) **electronic energies** – DFT calculations at M06-2X/6-311+G** level (***Gaussian***), single-point calculations in solution phase (SMD) → aqueous phase solvation free energies

d) **Gibbs free energies** – $G^{aq}_{red/ox} = E^{DFT}_{red/ox} + \Delta G^{xtb}$ → redox potentials according to the Nernst formula

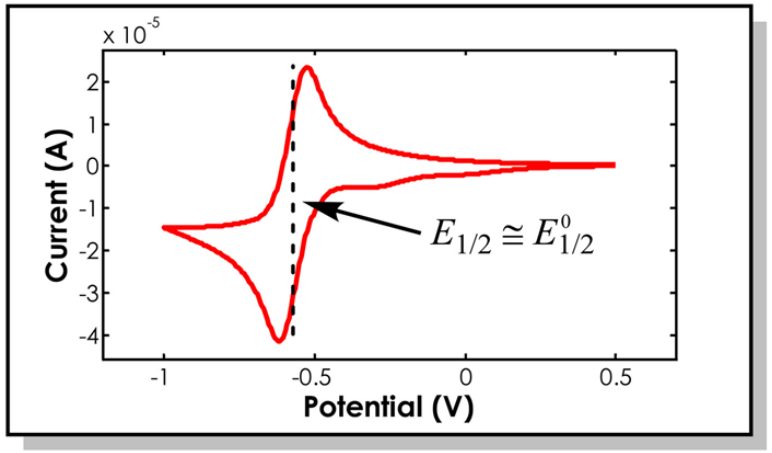
Composite method – force field + semiempirical + accurate QC

Software packages: *Schrödinger*, *xtb*, *Gaussian* + own scripts

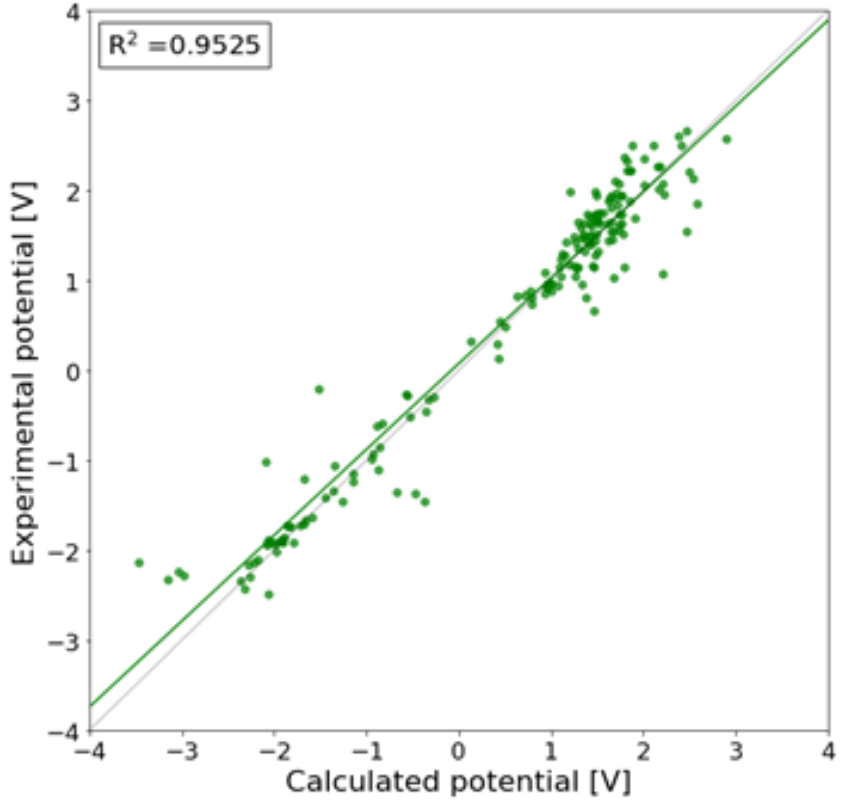
Test and benchmark calculations

Tests wrt experimental data

188 organic molecules (consistent CV data)



Roth et al., *Synlett*. **2016**, *27*, 714



MAE = 0.23 V, R^2 = 0.95

Test and benchmark calculations

Benchmarks wrt full DFT data

50/6712 molecules (diversity based selection)

full DFT calculations

opt: M06-2X/6-311G**

SP: M06-2X/6-311^{+}G** and M06-2X/6-311^{++}G(3df, 3pd)

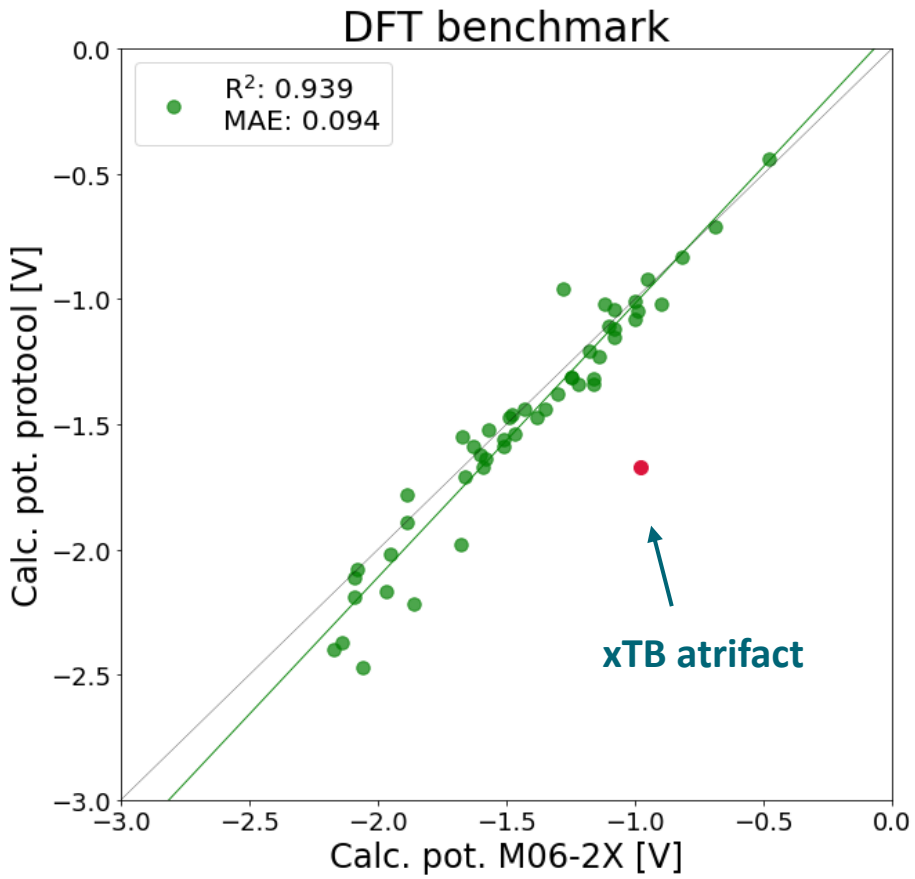
smaller basis set is sufficient

choosing the most stable conformer is reasonable

in some cases xTB results in ring closure

→ 6558 molecules

sufficient level of correlation



MAE = 0.094 V, R^2 = 0.94

Pyridoxal database

Histplotly utility

Database in tabulated form

NR	POT	R3	R1	R2	CHG	SYS
1	-1.29	COOH	P1	a1	0	pyr1
2	-1.26	COOH	P1	a2	0	pyr1
3	-1.34	COOH	P1	a3	0	pyr1
4	-1.37	COOH	P1	a4	0	pyr1
5	-1.27	COOH	P1	a5	0	pyr1
6	-1.28	COOH	P1	a6	0	pyr1
7	-1.35	COOH	P1	a7	0	pyr1
8	-1.25	COOH	P1	a8	0	pyr1
9	-1.2	COOH	P1	b1	0	pyr1
10	-1.21	COOH	P1	b2	0	pyr1
11	-1.22	COOH	P1	b3	0	pyr1
12	-1.22	COOH	P1	b4	0	pyr1

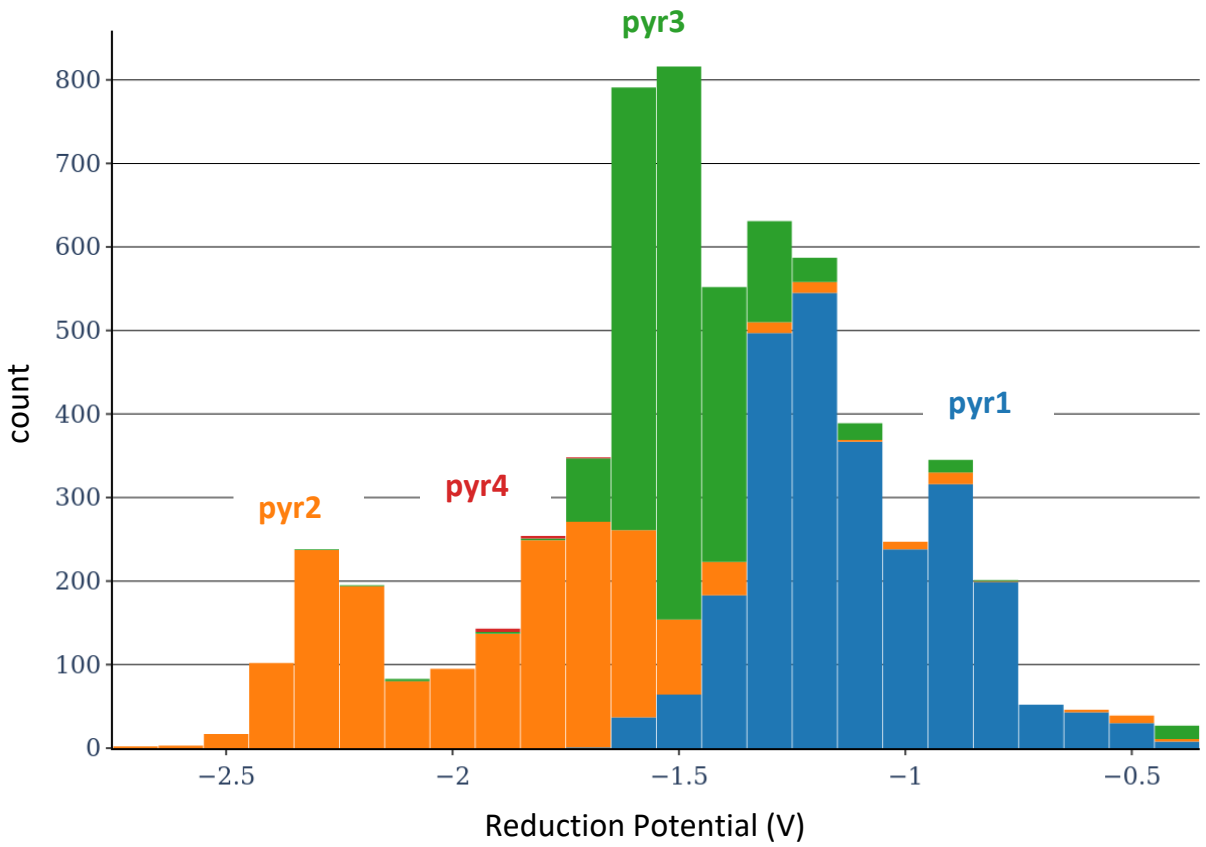
→

histogram analyzer

https://histplotly.herokuapp.com/

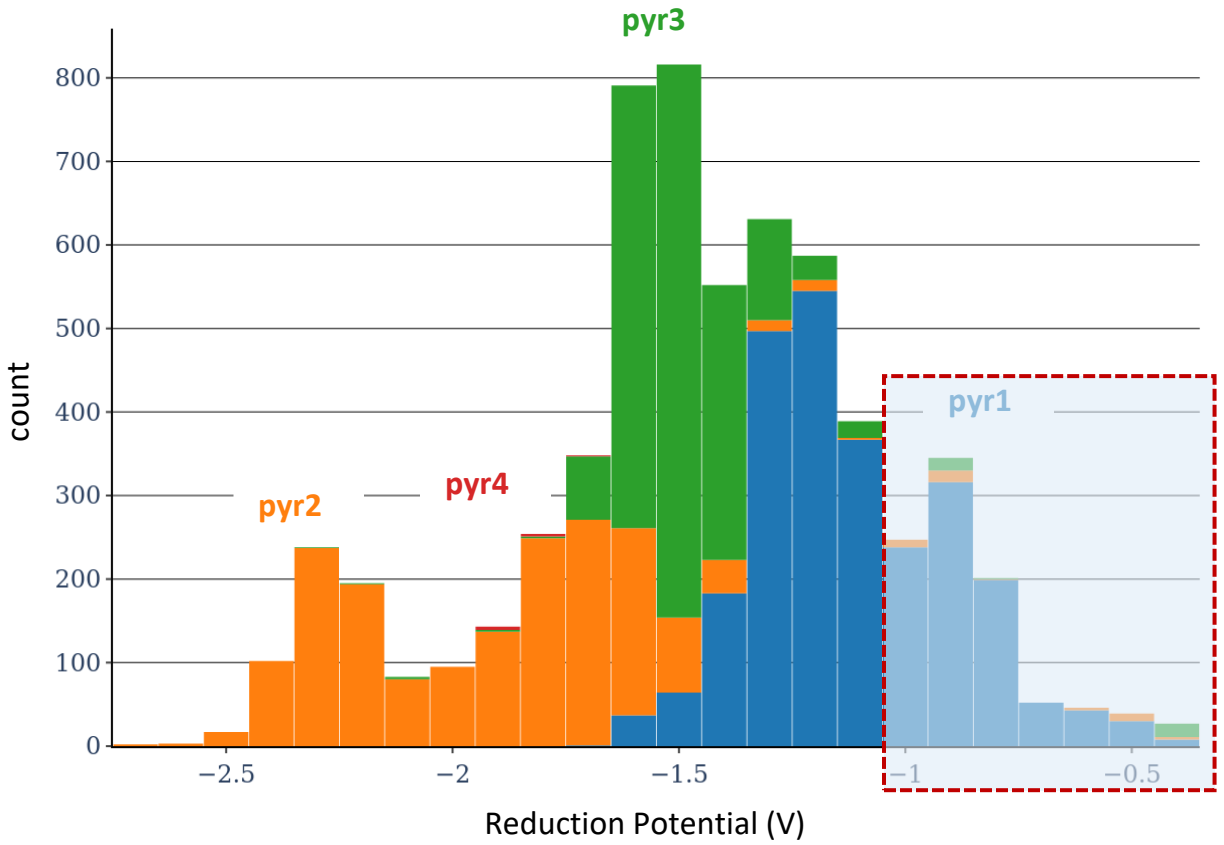
Pyridoxal database

Reduction potentials



Pyridoxal database

Reduction potentials

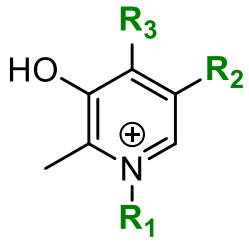


electrochemically relevant range (above -1 V) has lower populations, where

almost all compounds are from pyr1

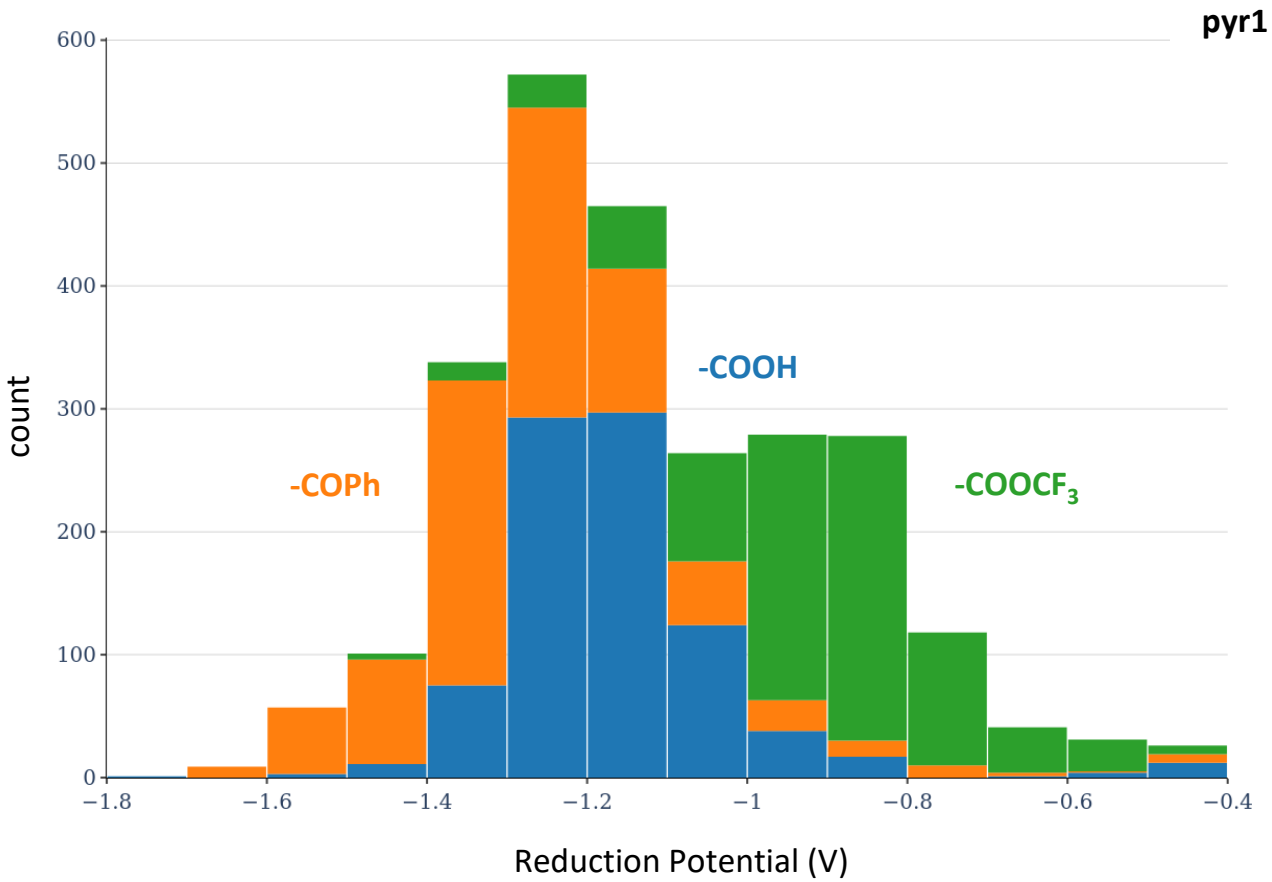
Pyridoxal database

Reduction potentials



pyr1

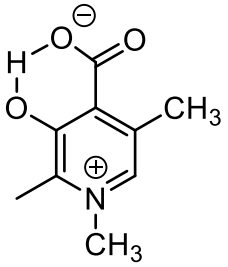
effect of R_3 substitution



Machine learning analysis

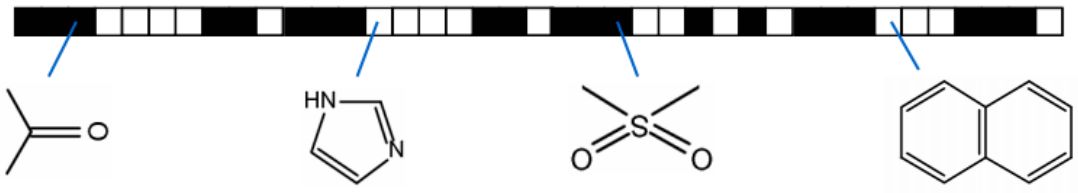
Molecular representations

a) SMILES strings



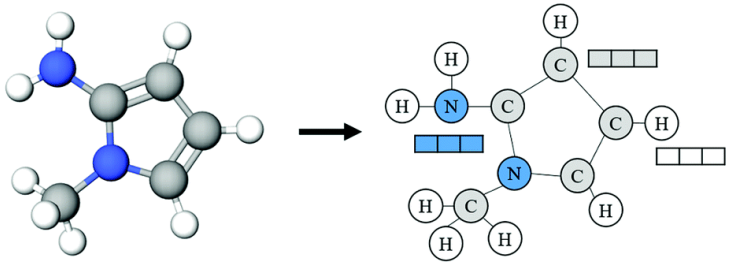
Cc1c[n+](C)c(C)c(O)c1C(=O)[O-] (atoms, bonds, connectivity)

b) Molecular fingerprints



(presence or absence of substructures)

c) Graph representation



(atoms = nodes, bonds = edges
each node has a feature vector)

Machine learning analysis

Machine learning methods

a) classic ML methods (Aalto)

- Random Forest analysis followed by SHAP feature attribution analysis

b) deep-learning methods (TTK)

- convolutional neural networks (CNN)
- extract higher-level features from the raw input
- ***DeepChem*** open source tool
- collaboration with **Budapest University of Technology**
- 3D graph convolutional network (3DGCN)

Machine learning analysis

Results

Performance of deep learning methods for reduction potentials (in V)

	Test MAE / V	Test RMSD / V	Test R^2
GraphConvModel	0.113	0.144	0.936
GCNModel	0.084	0.109	0.957
GATModel	0.093	0.124	0.929
DAGModel	0.076	0.106	0.943
AttentiveFPModel	0.061	0.086	0.963
Smiles2Vec	0.066	0.095	0.955
3DGCN	0.068	0.099	0.952

(**MAE**: mean absolute error; **RMSE**: root mean square error; **R^2**: coefficient of determination;

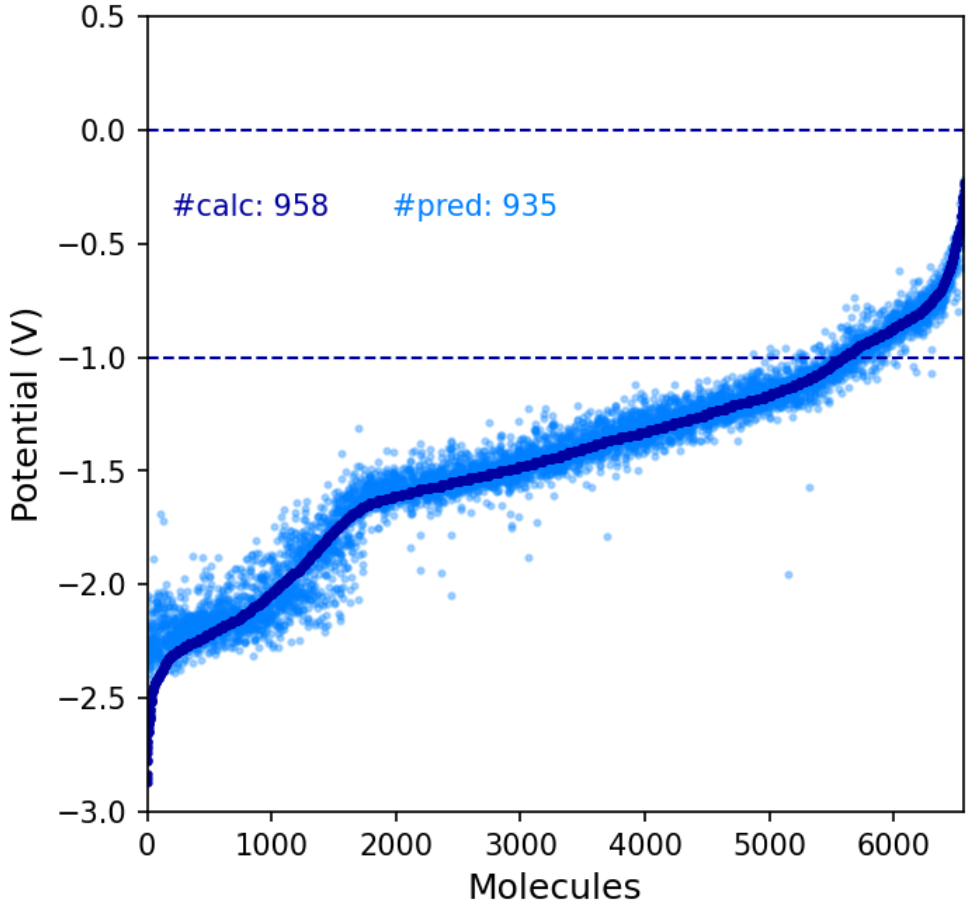
⟶ **remarkable performance exceeding the accuracy of the QC protocol**

Machine learning analysis

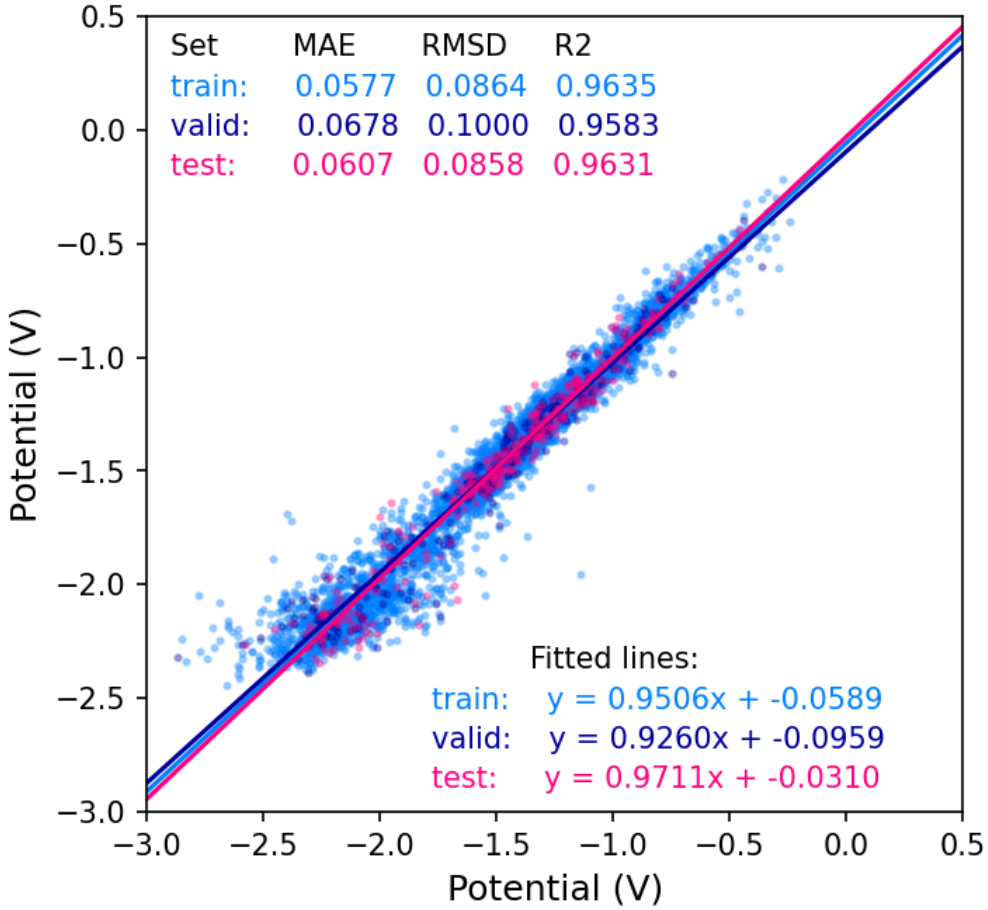
Results

AttentiveFPModel

a) distribution



b) correlation

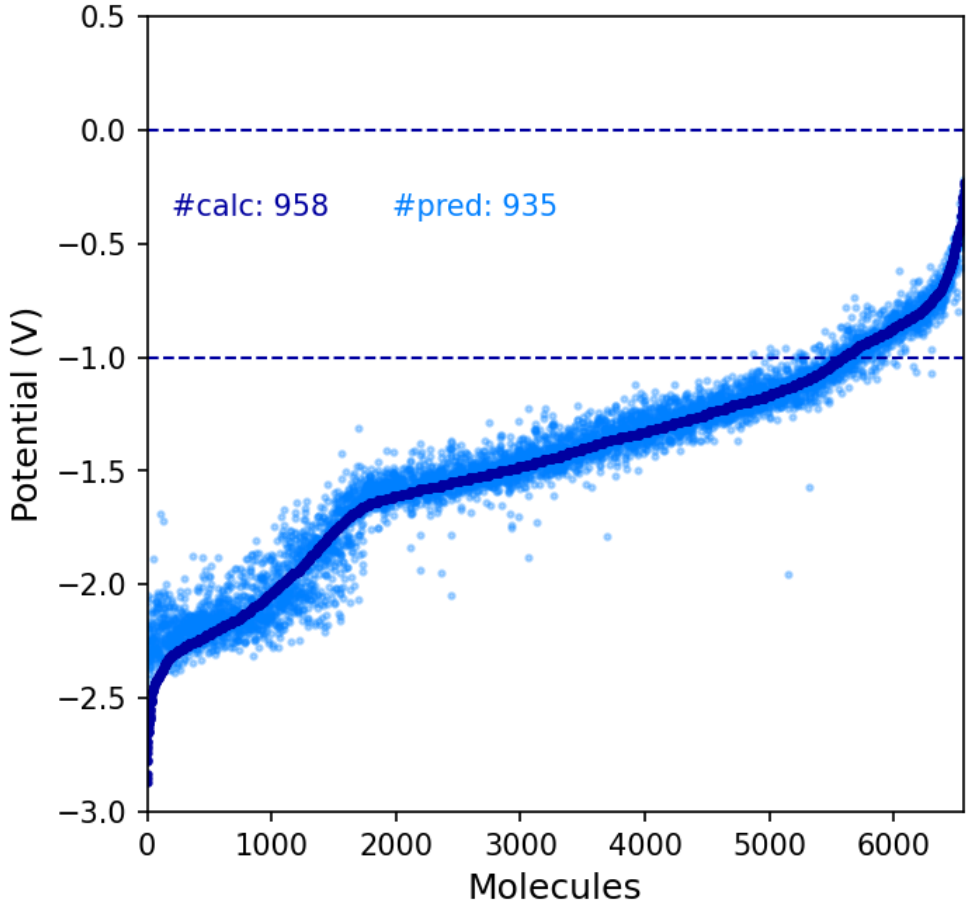


Machine learning analysis

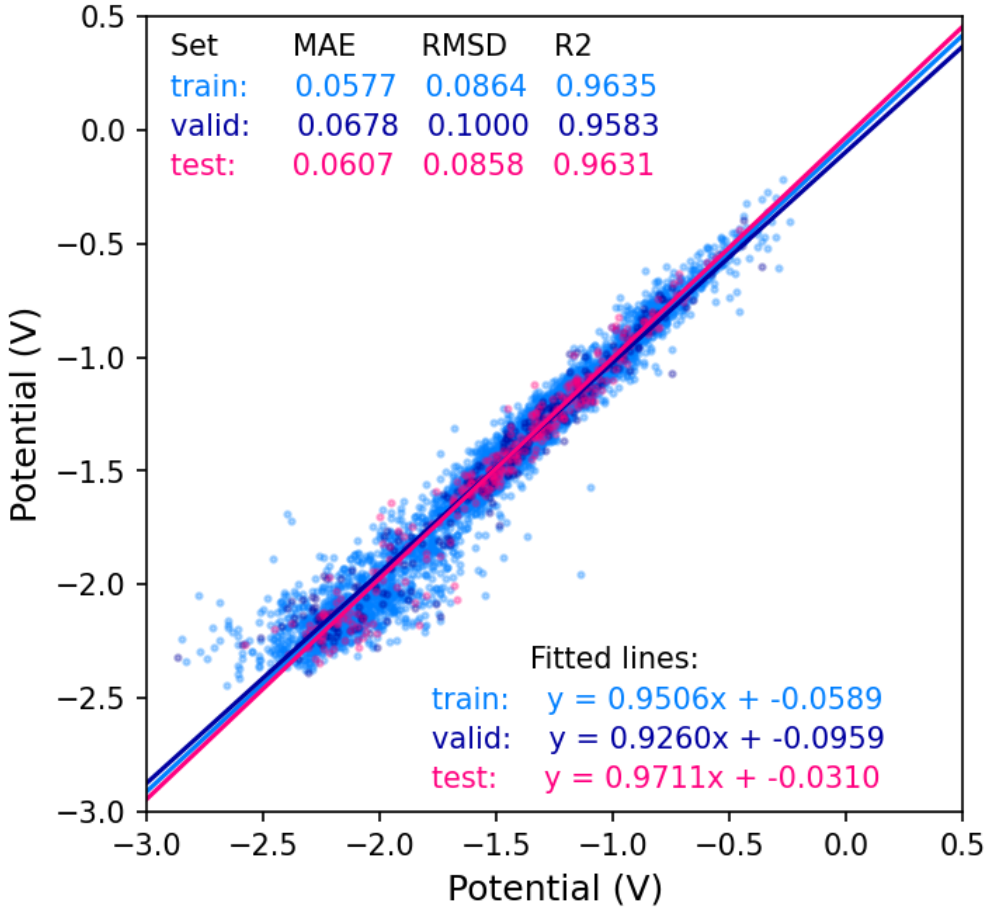
Results

AttentiveFPModel

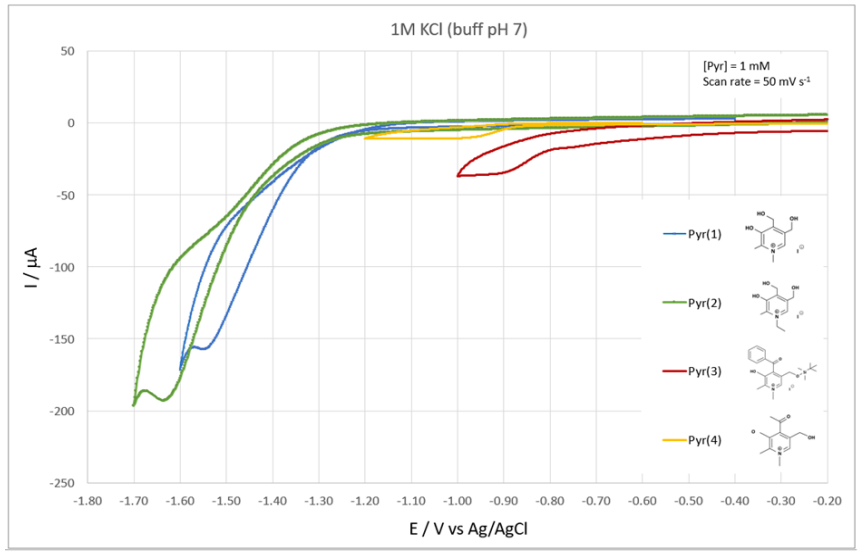
a) distribution

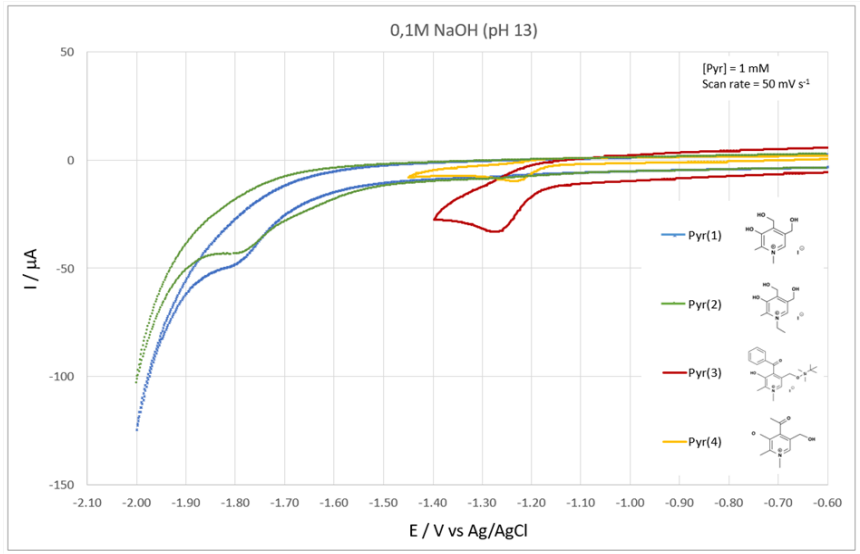


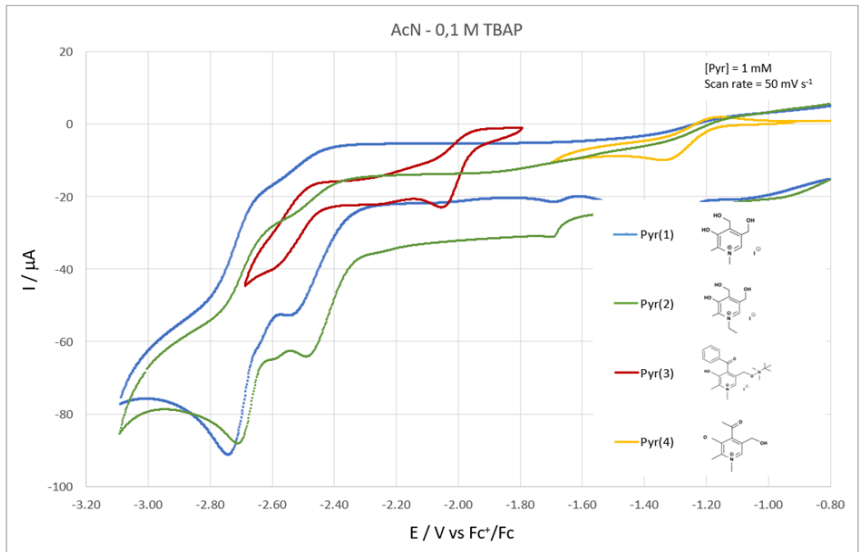
b) correlation

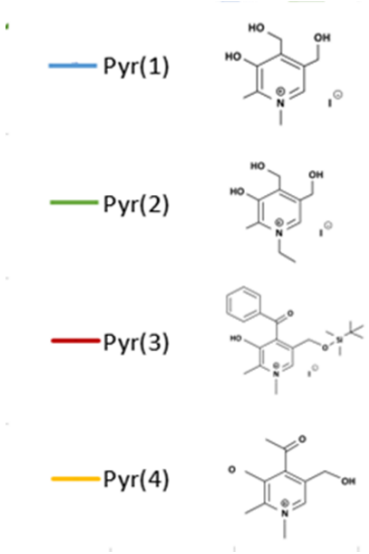


Experimental verification





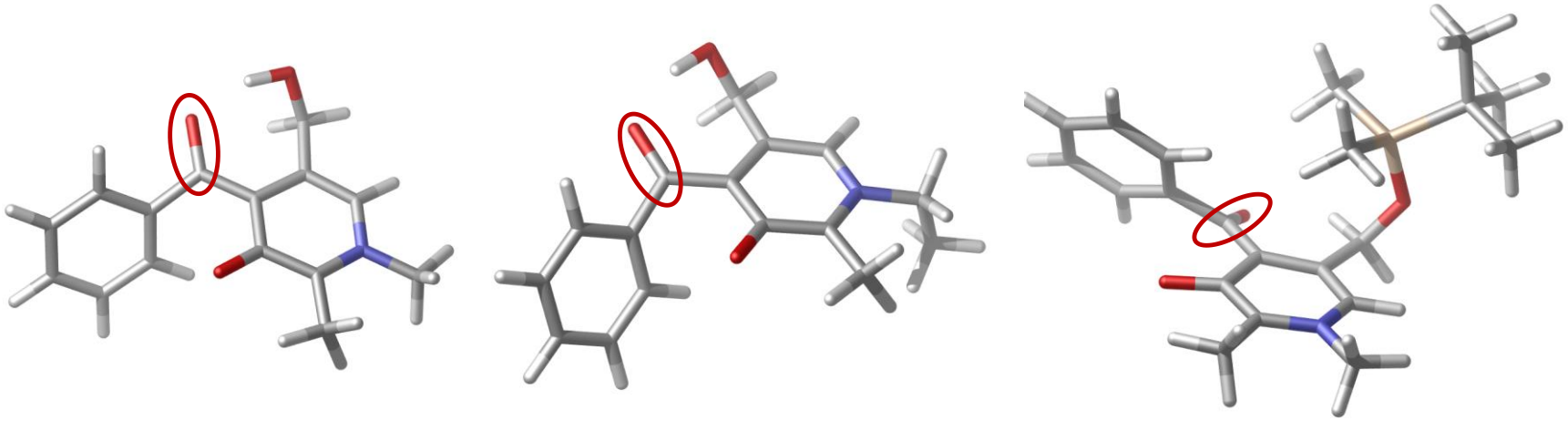




Gabriel Gonzalez
University of Turku

Anton Nechaev
Petri Pihko
University of Jyväskylä

Computational evaluation of stability

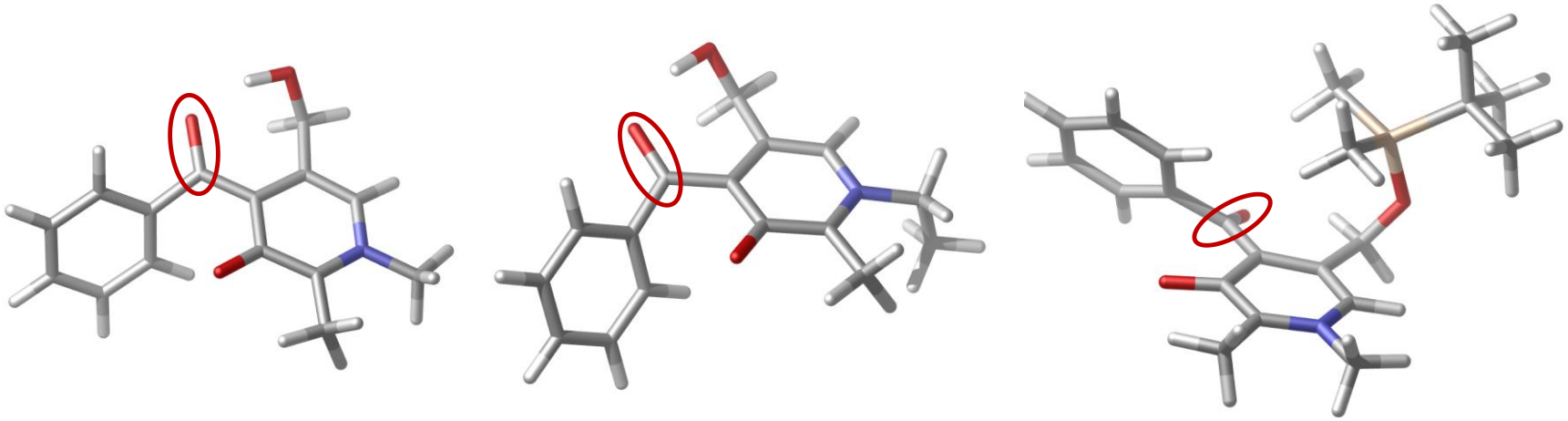


0.053 (0.060) 0.053 (0.060) 0.049 (0.061)

The bond elongations (in Å) for the synthesized molecules. Values obtained from DFT geometry optimizations are shown in parenthesis.

.

Computational evaluation of stability



0.053 (0.060) 0.053 (0.060) 0.049 (0.061)

The bond elongations (in Å) for the synthesized molecules. Values obtained from DFT geometry optimizations are shown in parenthesis.

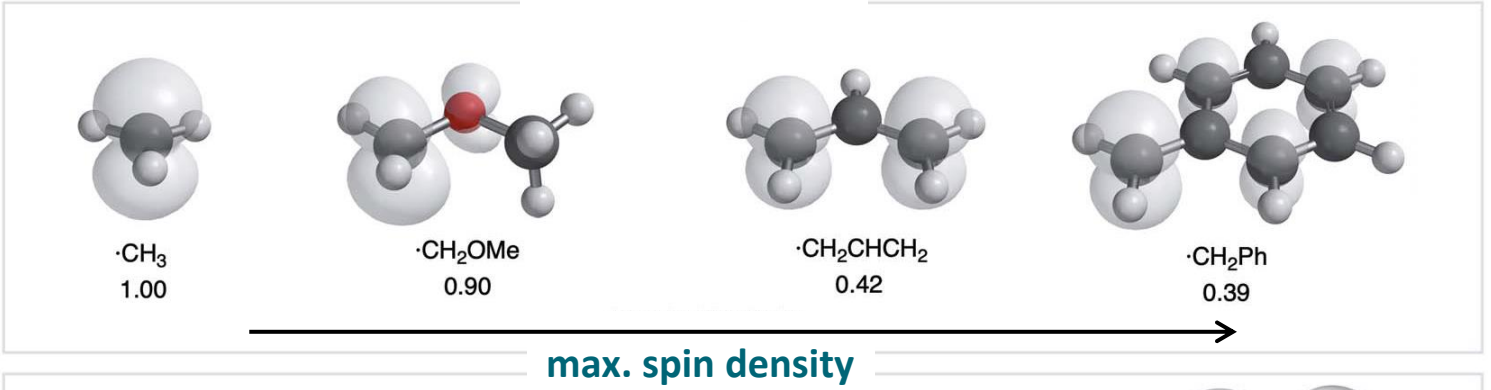
Elongation unlikely to explain instability upon reduction.
Experimental analysis suggests dimerization.

Stability studies

Two descriptors

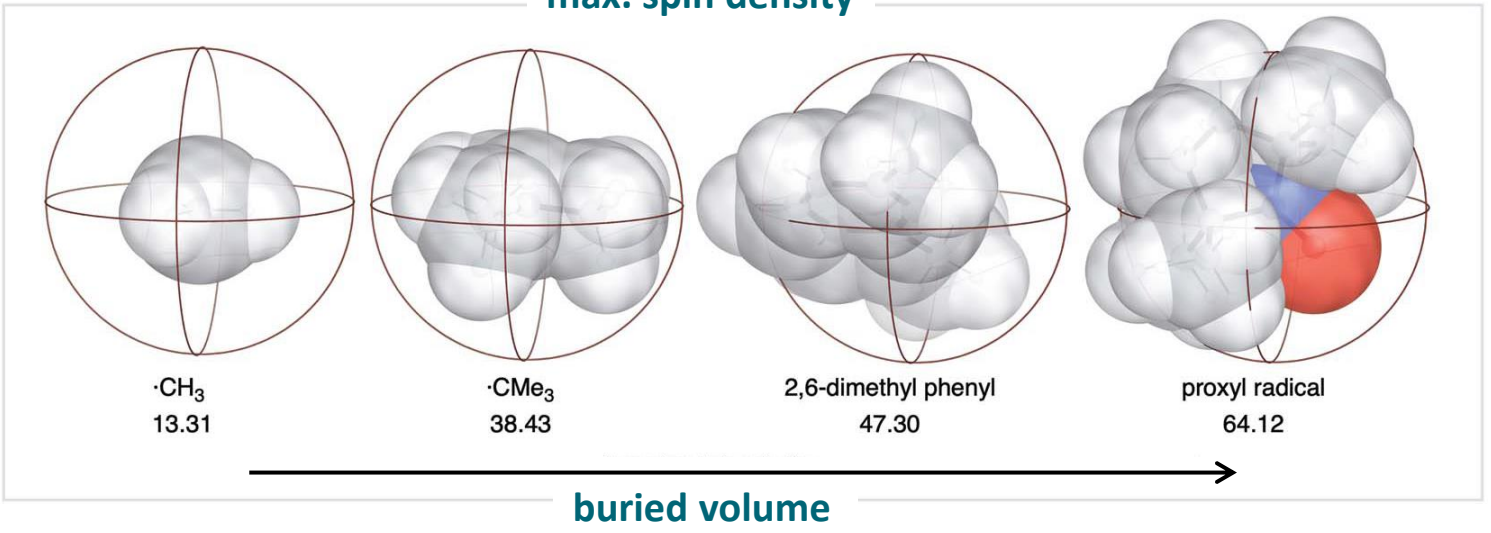
electronic

max. spin density



steric

percent buried volume



Stability studies

Two descriptors

electronic

max. spin density

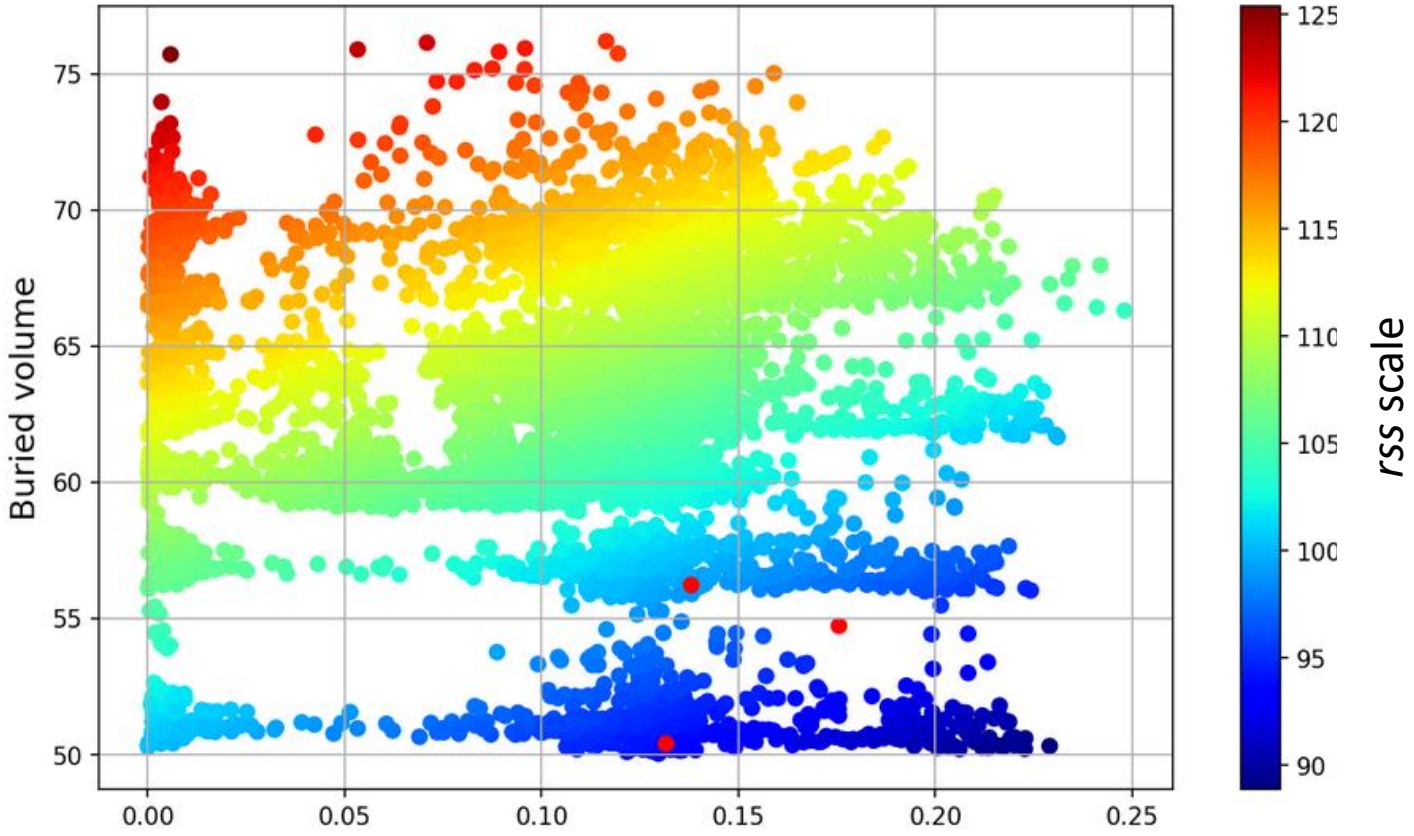
steric

percent buried volume

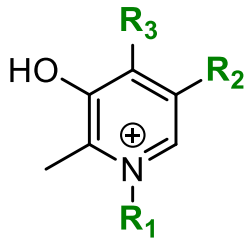
single metric: $rss = V_{bur} + 50 \times (1 - max.\ spin)$

Stability studies

Analysis of PYR database



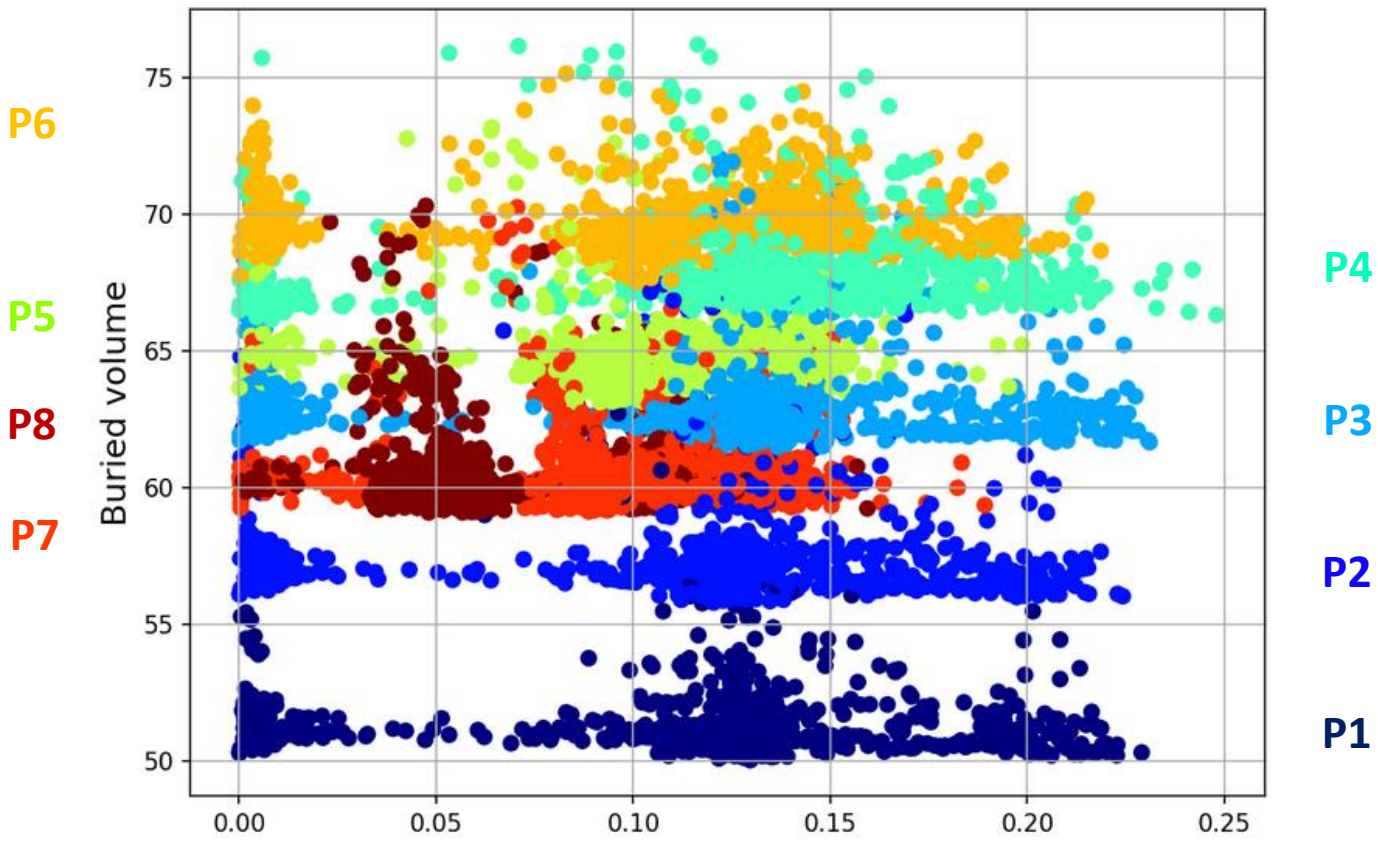
Fractional spin density at the N atom of the central ring



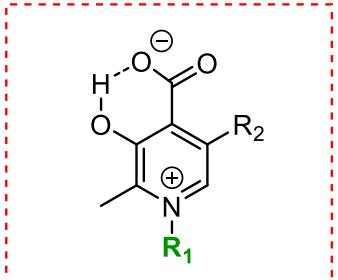
pyr1

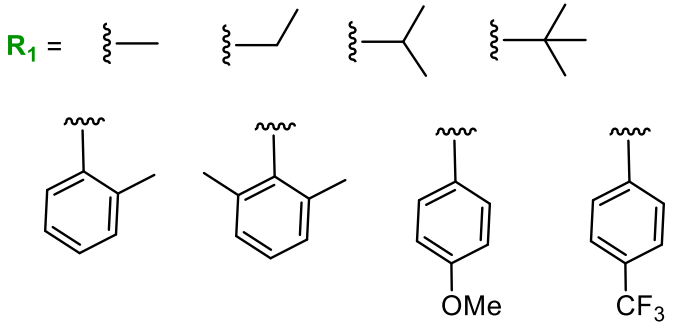
Stability studies

Analysis of PYR database



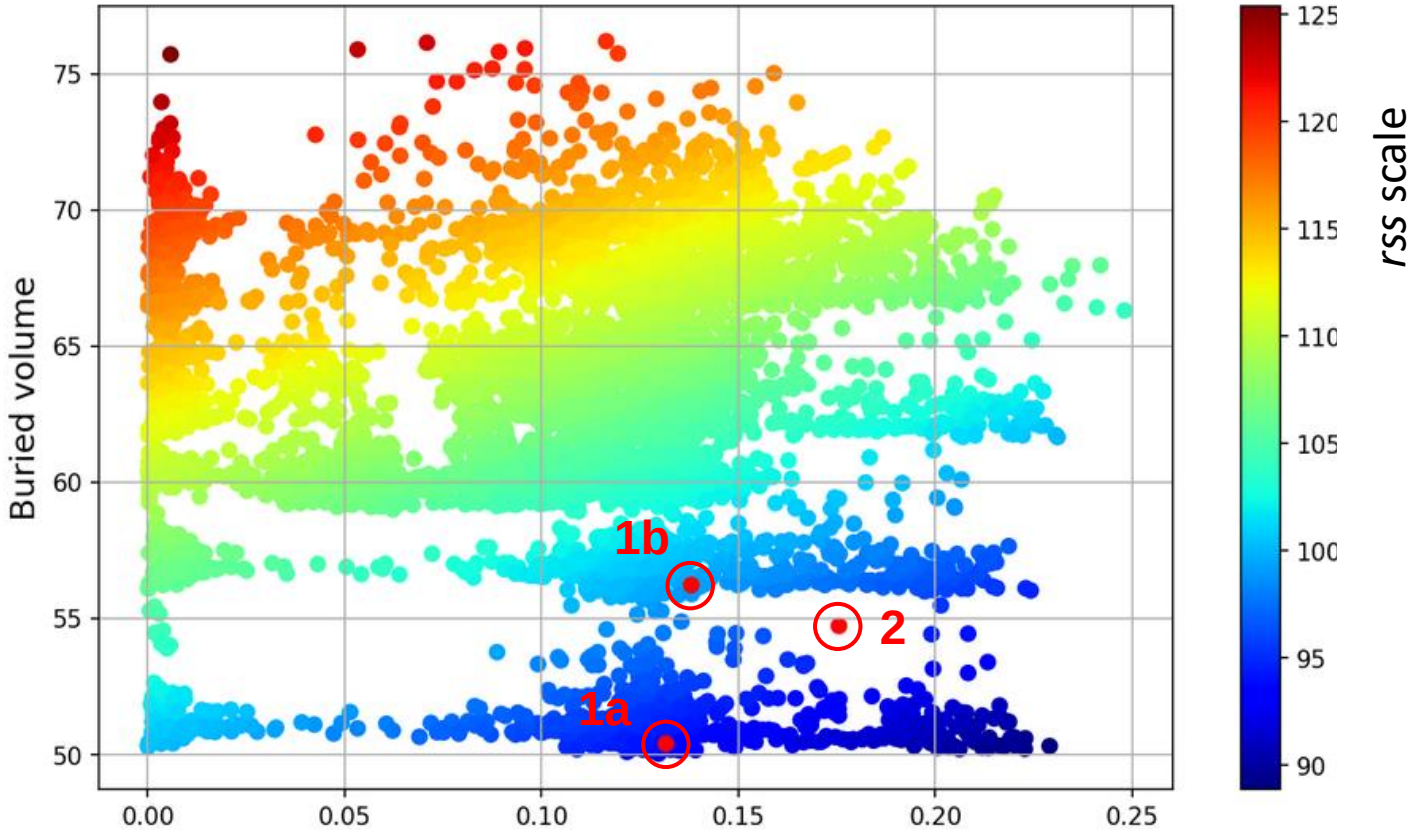
Fractional spin density at the N atom of the central ring



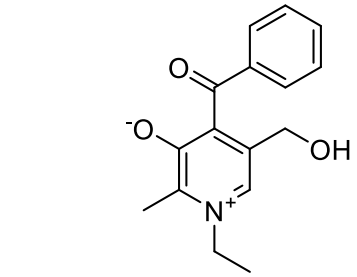


Stability studies

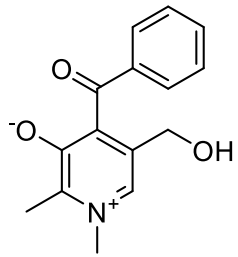
Analysis of PYR database



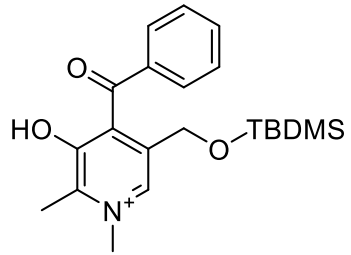
Fractional spin density at the N atom of the central ring



1b



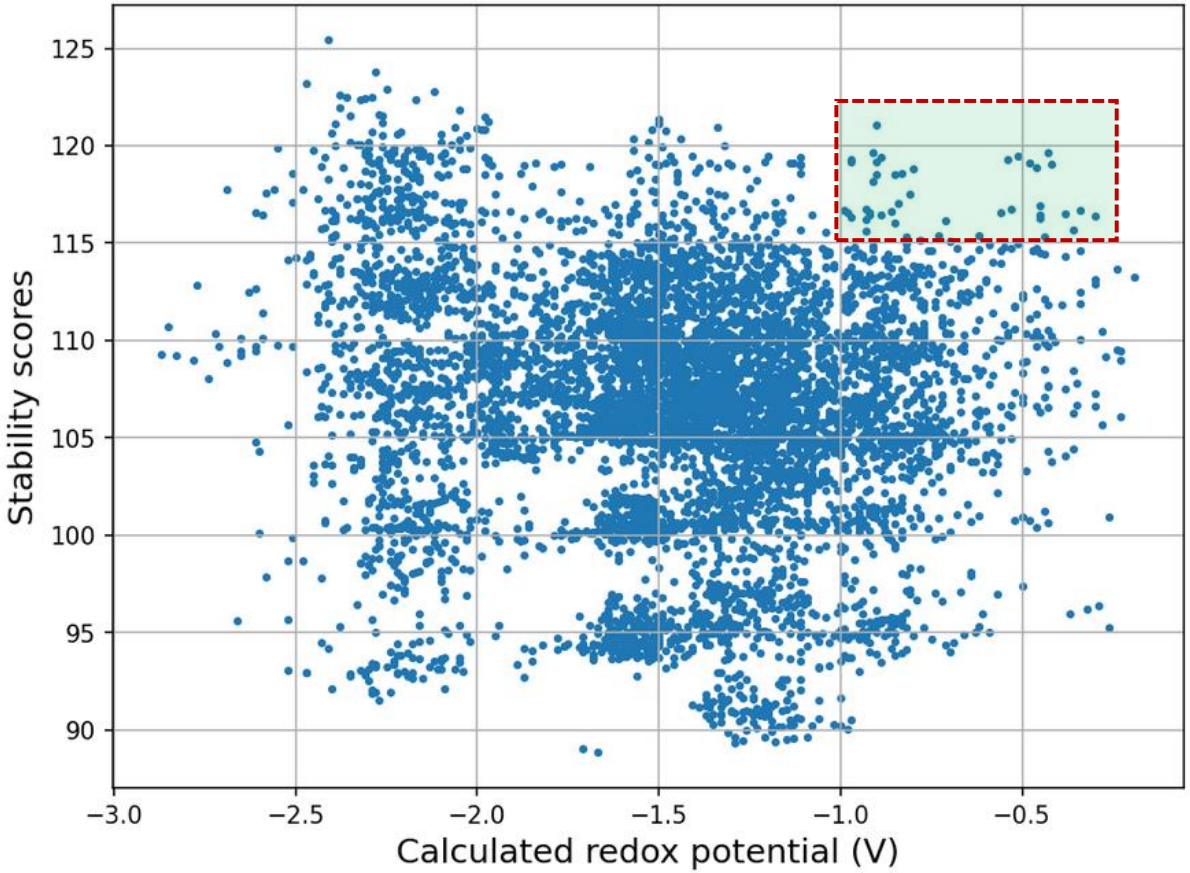
1a

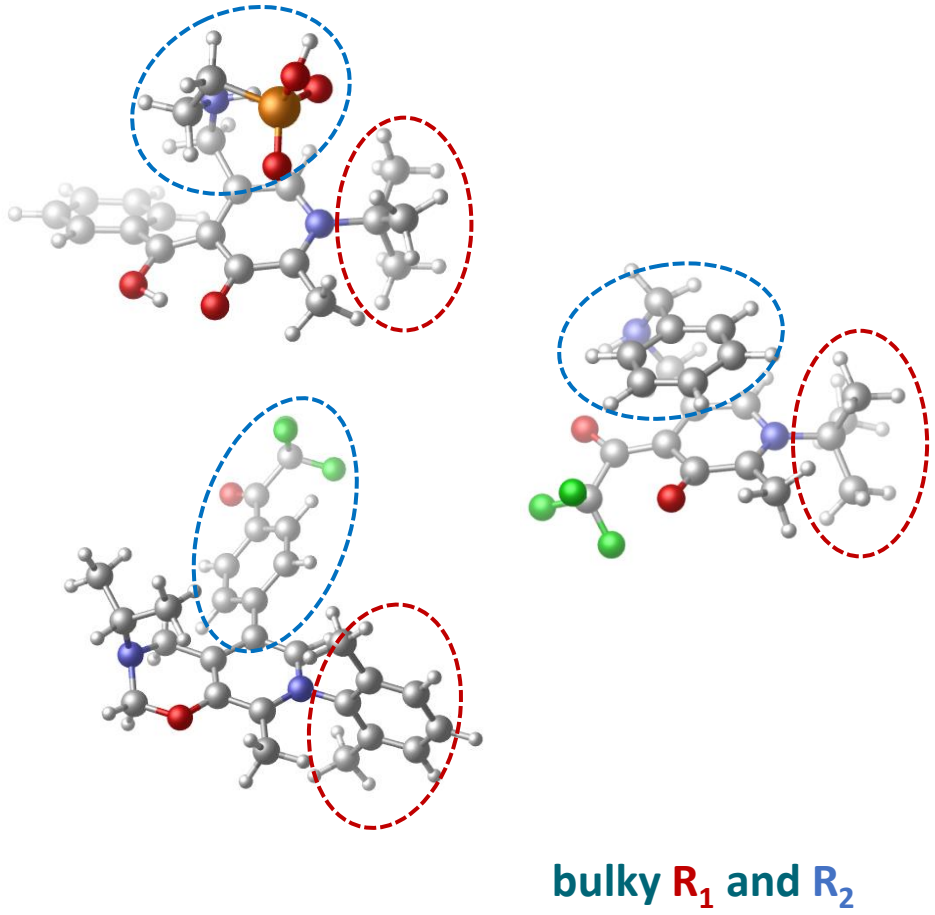


2

Stability studies

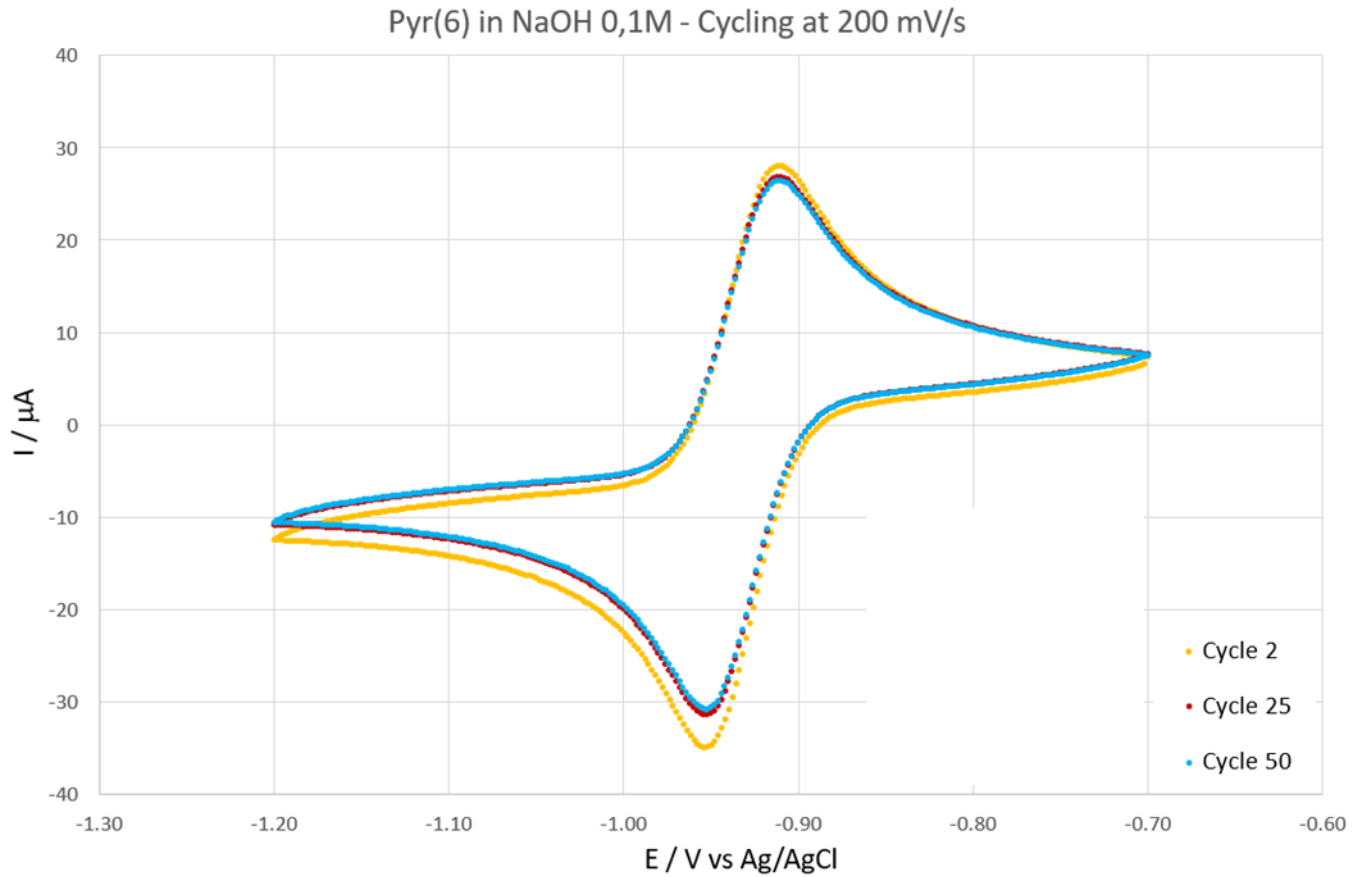
Analysis of PYR database





bulky R_1 and R_2

Stable molecule synthesized



Summary

WP1 results

- efficient computational protocol for building the molecular database
- molecular database of pyridoxal compounds,
- histogram analyzer (*histplotly*)
- assessment of machine learning models, all tested models work well
- application of a new approach to radical stability

work in progress

- more diverse molecular database (ChEMBL)
- ML redox potential predictions for 100.000 molecules

Participants

People involved from TTK and BME

Andrea Hamza
Ádám Madarász
Flóra Németh
Gergely Laczkó
Ákos Galvács

Imre Pápai

Luca Szegletes

Mátyás Pelle
Marcell Mészáros
Marcell Csikós
Tugyi Beatrix
El-Ali Maya

Thanks to