CompBat Deliverable

D1.1 Report on electrochemical experiments

Grant Agreement number	875565
Action Acronym	CompBat
Action Title	Computer aided desing for next generation flow batteries
Funding Scheme	H2020-LC-BAT-2019
Duration of the project	36 months, 1 February 2020 – 31 January 2023
Work package	WP1 - High-throughput screening and machine learning
Due date of the deliverable	31 January 2021
Actual date of submission	29 January 2021
Lead beneficiary for the deliverable	TTK
Dissemination level of the deliverable	Public

Project coordinator's scientific representative

Dr. Pekka Peljo

Aalto korkeakoulusäätiö, Aalto University (Aalto), School of Chemical Engineering

Department of Chemistry and Materials Science

pekka.peljo@aalto.fi

CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

compbat

Authors		
Name	**Beneficiary**	**E-mail**
Imre Pápai	TTK	papai.imre@ttk.hu
Andrea Hamza	TTK	hamza.andrea@ttk.hu
Ádám Madarász	TTK	madarasz.adam@ttk.hu
Németh Flóra Boróka	TTK	nemeth.flora@ttk.hu

Internal QA			
Reviewer	**Date of review**	**Comments**	**Date of revision**
Pekka Peljo Kari Laasonen	27.1.2021 29.1.2021	OK OK	

Abstract/Executive summary (of the deliverable)

This report presents the first molecular database of redox potentials and aqueous solubilities derived from DFT calculations. The virtual molecular library has been generated within Task 1.1, and it represents deliverable D1.1 of the CompBat project.

The database includes over 6200 molecules derived from pyridoxal via systematic variation of three different substituents of the framework. A computational protocol has been developed, tested and applied to provide high quality predictions for reduction potentials and estimates for solvation free energies. The protocol involves automated generation of 3D structures, extensive conformational screening, and utilization of a composite electronic structure method to predict the desired properties. The obtained data are analyzed using a histogram analyzer software developed within the framework of the project.

The developed molecular database will be used as a training/validation set in assessing various machine learning approaches in the next phase of the project (within Task 1.2). A variety of descriptors including electronic and steric parameters, as well as different molecular representations (fingerprints and graphs) are considered for such analysis.

CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

compbat

Contents

1 Introduction

The CompBat project aims at developing various tools for discovery of new prospective candidates for next generation of redox flow batteries (RFB). High-throughput screening (HTS) that enables the identification of promising candidates of water-soluble redox-active compounds for experimental synthesis and electrochemical characterization is one target tool, and these developments are carried out within work package WP1.

The first step of our general strategy along these lines is to build an initial database for two basic properties (redox potentials and aqueous solubilities), and utilize various machine learning techniques to provide efficient screening methodology applicable for a large and diverse set of molecules. The molecular database is built in a combinatorial manner. We define bioinspired frameworks (vitamins and amino acids) and introduce substituents with broadly varying electronic and steric properties. The tools of computational chemistry are used to predict high quality data for redox potentials and reasonable estimates for aqueous solvation free energies. The generated molecular library forms the training-validating set for subsequent analysis using different machine learning algorithms. The molecular library is then expanded iteratively in terms of the number of molecules, as well as the diversity of molecular frameworks.

Herein, we report on a database of over 6200 molecules derived from the pyridoxal framework. We present a computational protocol that enables high-quality predictions for the target properties at a reasonable computational cost. In addition, we offer a Python utility to assist the analysis of the present database.

CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

2 Molecular Sets

Four different molecular sets, denoted as **pyr1** – **pyr4**, were chosen to generate the present database (Figure 1). Label "pyr" refers to the pyridoxal framework, which is one form of vitamin B6.

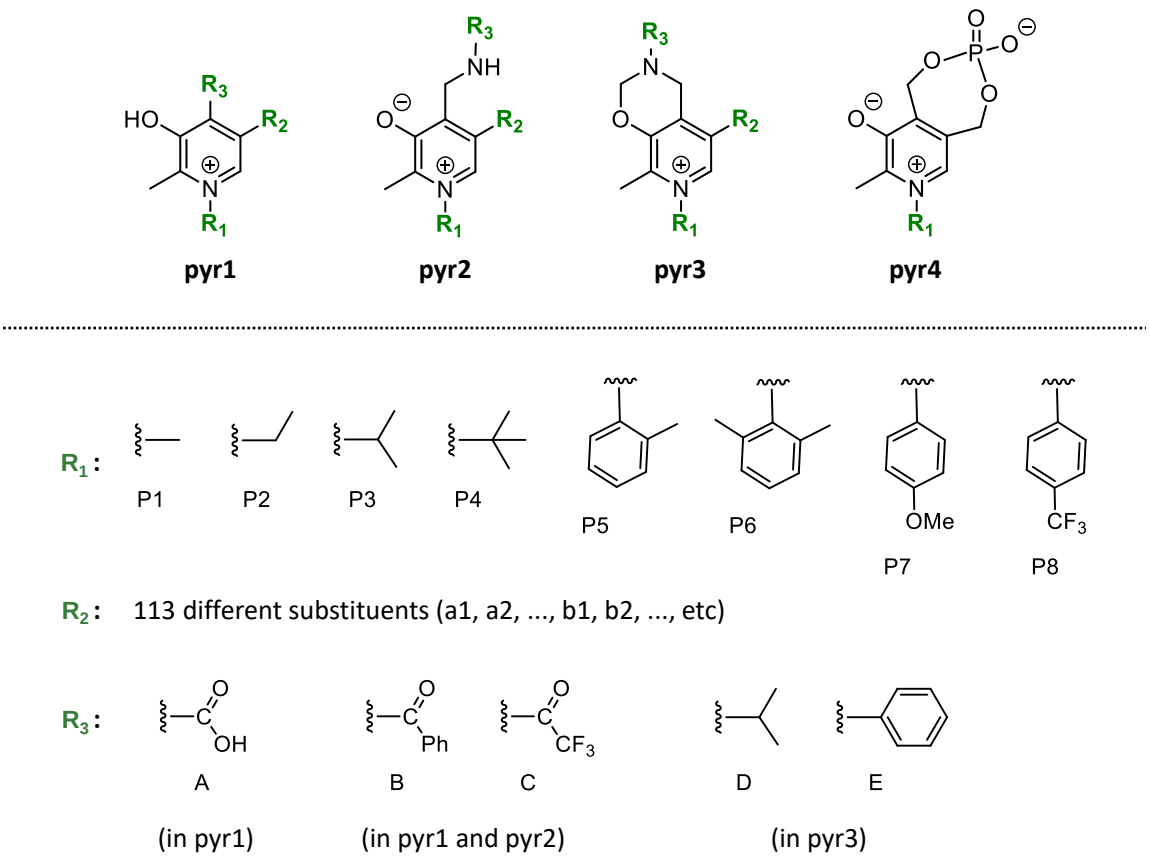


Figure 1: Molecular sets derived from the pyridoxal framework. The full set of R_2 substituents is provided in Appendix 2.

The combination of R_1, R_2 and R_3 substituents gives rise to 6336 molecules (2712 for **pyr1**, 1808 for **pyr2** and **pyr3**, 8 for **pyr4** sets). All these molecules were considered for quantum chemical calculations using the protocol described below. The molecules are labeled according to the sets they belong to and their substituents. Selected examples with the introduced labeling are shown in Figure 2. The full list of R_2 substituents is given in Appendix 2.

CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

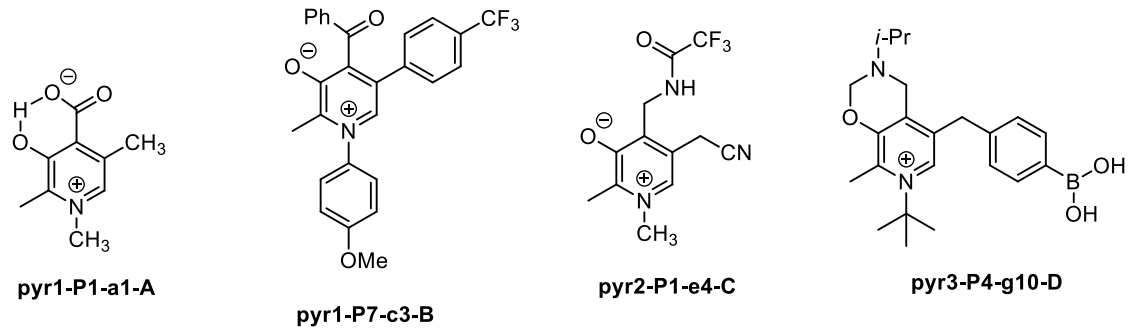


Figure 2: Selected examples from the molecular database.

3 Computational Methodology

Computed properties

The reduction potentials of the molecules were calculated by Nernst equation as

$$E_{\mathrm{red}}^{\mathrm{comp}} = -\frac{\left(G_{\mathrm{red}}^{\mathrm{aq}} - G_{\mathrm{oxid}}^{\mathrm{aq}}\right)}{nF} + \Delta E_{\mathrm{red}}^{\mathrm{ref}}$$

where $G_{\mathrm{red}}^{\mathrm{aq}}$ and $G_{\mathrm{oxid}}^{\mathrm{aq}}$ are computed aqueous phase Gibbs free energies of the reduced and oxidized forms of the molecules, *n* is the number of electrons involved in the redox process (*n* = 1 in our case), *F* is the Faraday constant, and $\Delta E_{\mathrm{red}}^{\mathrm{ref}}$ is a reference value set to –4.290 V, so as the computed reduction potentials are given relative to that of methyl viologen.

The aqueous solubilities were related to Gibbs free energies of aqueous solvation, which were estimated via an implicit polarizable continuum model. This approach cannot be used to provide accurate absolute solubilities, but the free energy data could be useful to approximate trends in the screening procedure.

Computational protocol

The solution phase Gibbs free energies were computed via a composite method that employs different modelling tools. The applied protocol involves the following steps.

1) Generation of 3D structures

The 2D (Lewis) structures of the molecular database were generated from the four molecular frameworks in a combinatorial manner using the *CombiGlide* module of the *Schrödinger* software package [1]. The conversion of 2D to 3D structures was carried out with the *LigPrep* module. The protonation state of each molecule was established by the *Epic* module according to pH = 7 in water. A preliminary conformational search in water as a solvent was performed at this stage by using the OPLS

CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

force field as implemented in *Macromodel*. The energetically lowest lying conformers were subject to subsequent calculations.

2) Conformational analysis

The semiempirical extended tight-binding method GFN2-xTB [2,3] developed by the Grimme group was used in the next step for geometry optimizations in aqueous phase. This approach was reported to be a robust and accurate method for molecular structures and vibrational frequencies, which is exploited in our present project as well. The solvent effects are incorporated implicitly via the generalized Born model with surface area contributions (GBSA). The Conformer–Rotamer Ensemble Sampling Tool (CREST) procedure, via the *crest* utility of the *xtb* program, was then used for conformational search in water for all molecules and their reduced forms as well. The conformational search procedure involves extensive metadynamic sampling and additional genetic Z-matrix crossing algorithms. For the most favored conformers, the thermal and entropic contributions are computed using the rigid-rotor-harmonic-oscillator (RRHO) approximation at room temperature.

3) Electronic energies via DFT

Density functional theory (DFT) was applied to compute the electronic energy term of the Gibbs free energy. The DFT calculations were carried out for the most stable conformer of the molecule using the xTB optimized structure (single-point calculations). We used Truhlar's M06-2X functional [4] along with the 6-311+G** basis set in these calculations as implemented in *Gaussian 16* [5]. The electronic energies include aqueous phase solvation free energies as well estimated via the SMD implicit solvation model [6]. For each species, additional single-point gas-phase energy calculations were carried out to elucidate the aqueous phase solvation free energies explicitly.

4) Gibbs free energies

The solution phase Gibbs free energies $G^{\mathrm{aq}}_{\mathrm{red}}$ and $G^{\mathrm{aq}}_{\mathrm{oxid}}$ used in the Nernst formula are thus computed via a composite method as

$$G^{\mathrm{aq}}_{\mathrm{red/oxid}} = E^{\mathrm{DFT}}_{\mathrm{red/oxid}} + \Delta G^{\mathrm{xtb}}$$

where $E^{\mathrm{DFT}}_{\mathrm{red/oxid}}$ refers to the electronic energy term and ΔG^{xtb} involves all finite temperature contributions, as described above. We think this combined approach represents an efficient method to provide reliable (DFT quality) redox potentials and aqueous solubilities for the target molecular database.

We note at this point that geometry optimizations carried out with the semiempirical GFN2-xTB method, or the CREST conformational screening procedure, gave rise to unexpected chemical 6rtefact6ments (cyclization, water elimination, for instance) for some molecules. Although these transformations may have chemical relevance, we decided to exclude these cases from our present molecular library. For the **pyr2** set, C-O bond formation (ring closure) between the electrophilic carbon of the R_3 = $C(O)CF_3$ group and the nucleophic oxygen of the pyridoxal 6rtefact systematically. In these cases, we performed structure filtering to eliminate this 6rtefact. Investigation of stability issues arising

CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

from the decomposition of the electrolytes is within the scope of WP1 and WP4, but in Task 1.1, we have not addressed this otherwise important issue. For the sake of simplicity, the current database includes molecules only that do not undergo chemical changes upon the computational procedure. This requirement is monitored by checking the connectivity patterns of molecules at various phases of the computational protocol. As a result, we have 6204 compounds with computed reduction potential and aqueous solubilities in the present database.

4 Results and Discussion

4.1 Test and benchmark calculations

To assess the accuracy of our computational protocol, two types of test calculations were carried out.

1) Tests with respect to experimental data

We first considered a molecular set (183 organic molecules) with reliable experimental electrochemical data available [7] and computed the redox potentials using our protocol. The experimental reference data were obtained via cyclic voltammetry at the same conditions. The molecular set includes a broad variety of organic functionalities having their measured redox potentials in the -2.5 – +2.7 V range (130 data for one-electron oxidation potentials in the 0.26 – 2.66 V and 53 reduction potential values in the -2.48 – -0.20 V range). We find that the mean absolute error of predictions (with respect to the experimental data) is 0.24 V (0.23 and 0.25 V for the oxidation- and reduction potentials, respectively), which is encouraging considering that full DFT methods provide results with similar quality [7]. Correlation of calculated redox potentials versus experimental data is shown in Figure 3.

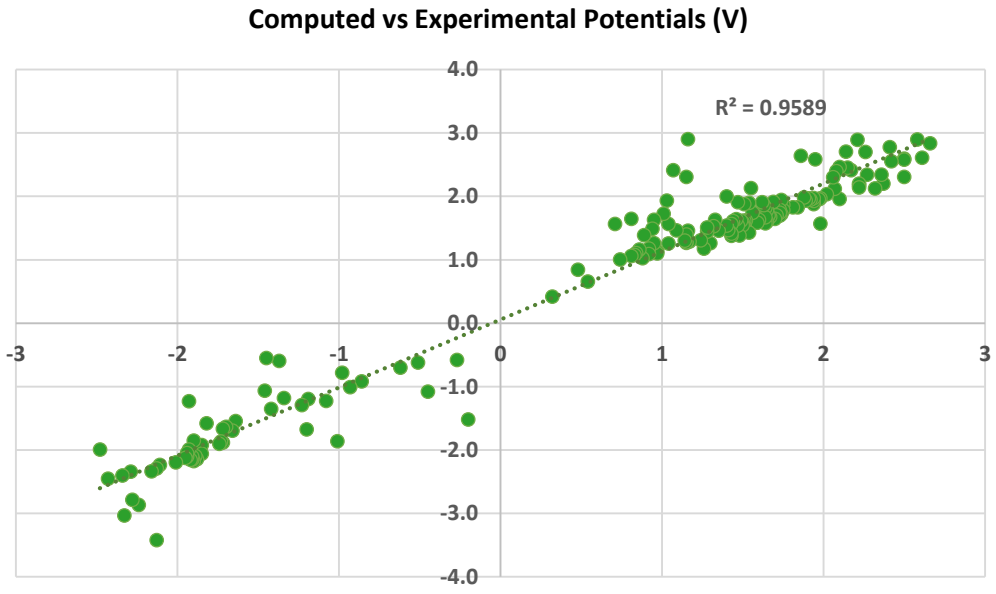
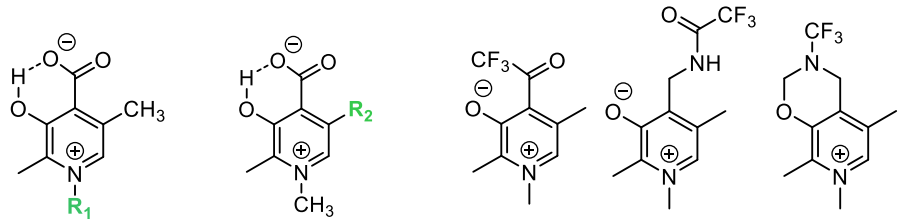


Figure 3: Correlation of calculated redox potentials vs experimental values.

CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

b) Benchmark calculations with respect to full DFT results

Our combined protocol was then assessed with reference to full DFT results. For a selected molecular set (18 molecules from the **pyr1** set; see Figure 4), benchmark calculations were performed using standard DFT calculations for all free energy contributions (full DFT). In these calculations, the geometries of the 5 lowest lying molecular conformers of the *crest* analysis were reoptimized at the M06-2X/6-311G** level of theory using water as a solvent (SMD solvent model). Thermal and entropic corrections to the Gibbs free energies were estimated at the same level using the RRHO approximation. For each optimized geometry additional single point calculations were carried out with the larger basis set 6-311+G**. The most stable conformer for each molecule was chosen for the final solvent phase Gibbs free energies. The DFT redox potentials were calculated according to the Nernst formula and they range between -0.9 and -1.8 V. The analysis gives mean absolute error (MAE) less than 0.1 V with a maximum deviation of 0.27 V.



R_1 = P1-P8 **R_2 = b1,c1,d1, e1,f1,g1,h1**

MAE: 0.07 V 0.07 V 0.09 V

Figure 4: Test set for benchmark calculations. The MAE with the reference to DFT calculations is shown in red.

These test calculations indicate that the present protocol approaches DFT-level accuracy for redox potentials at a significantly lower computational cost (at least one order of magnitude faster). This allows us to handle a large set for molecules (both in terms of molecular size and quantity), which is in line with the HTS concept.

4.2 Analysis of computed reduction potentials

The reduction potentials computed for the 6204 compounds of the database are provided in a tabulated form as a supplementary information (separate Excel document; see Appendix 3). Herein, we provide a brief analysis of the results.

The distribution of calculated reduction potentials is illustrated in terms of histograms, such as the one shown in Figure 5. In this particular example, different colors refer to the four molecular sets (**pyr1** – **pyr4**). It is apparent that reduction potential region around -1.5 V is the most populated, and the electrochemically relevant range (above -1 V) has considerably lower populations. In this region, almost

CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

all compounds are from the **pyr1** set. Molecules from the **pyr3** set have reduction potentials around -1.5 V, while the potentials of the **pyr2** members are shifted more towards -2.5 V.

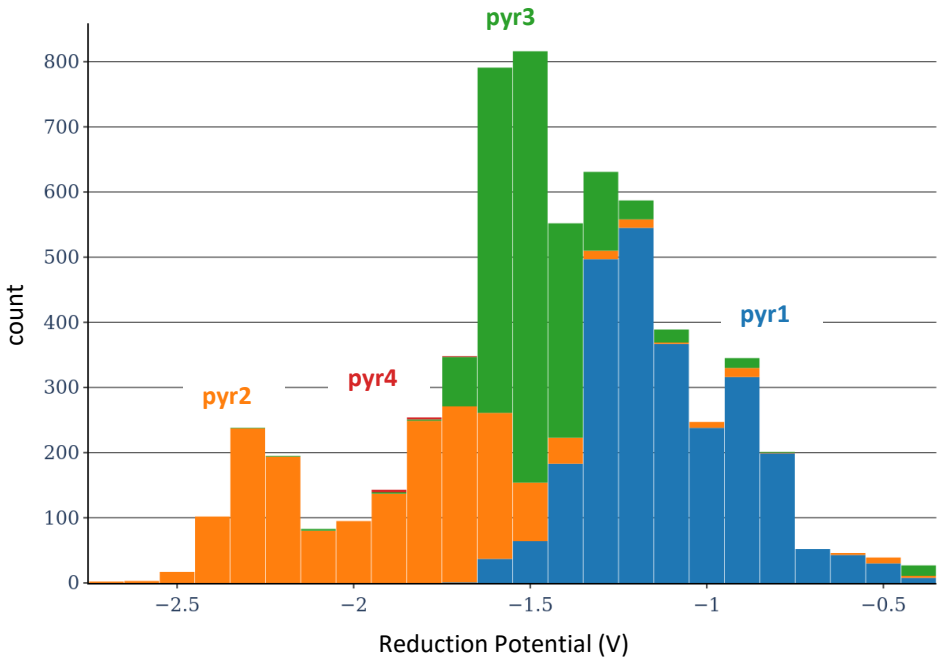


Figure 5: Distribution of calculated redox potentials. The columns are colored according to the molecular sets.

Figure 6 displays the distribution over the total charge of the species (oxidized forms) indicating that the majority of the molecules having reduction potentials above -1 V are neutral.

The distribution over the R_3 substituents in the **pyr1** set is shown in Figure 7. Most of the structures appear around -1.2 V, but the effect of the electronic nature of the R_3 group is apparent. The potential range of electrochemical relevance (above -1.0 V) is dominated by compounds with R_3 = $C(O)CF_3$, whereas molecules involving group COOH are shifted to -1.2 V and with groups C(O)Ph to -1.4 V.

For further analysis, the R_2 substituents in the -1.0 – -0.4 V potential range were inspected (Figure 8). The number of total R_2 substituents is 113, but only a few from this diverse set appear in the potential range above -0.7 V, and actually, there are only 7 out of the total number of R_2 groups that are present in these structures.

CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

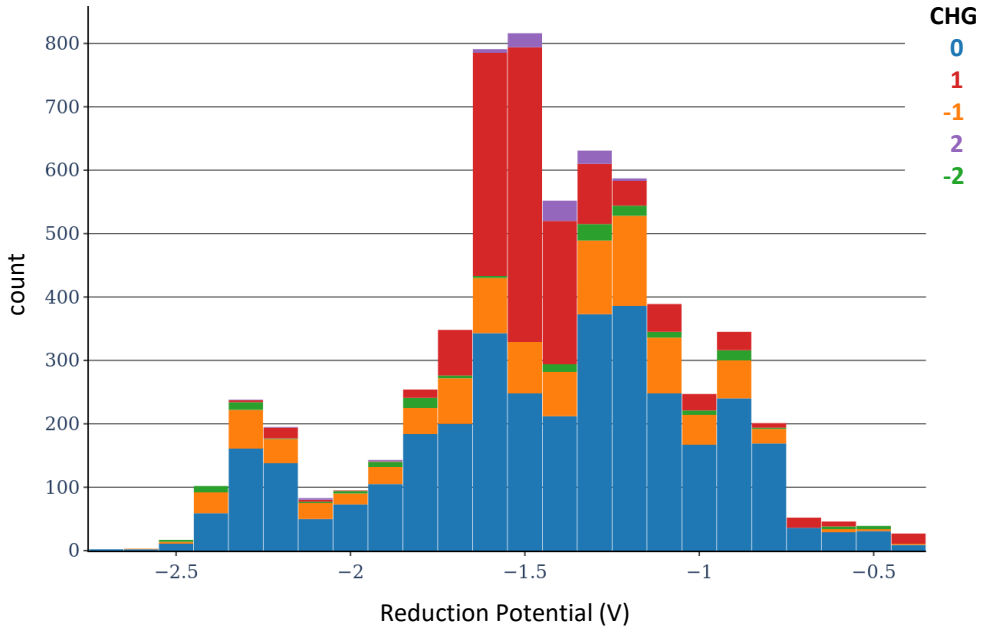


Figure 6: Distribution over charges of calculated redox potentials. The columns are colored according to the charges of the starting compounds.

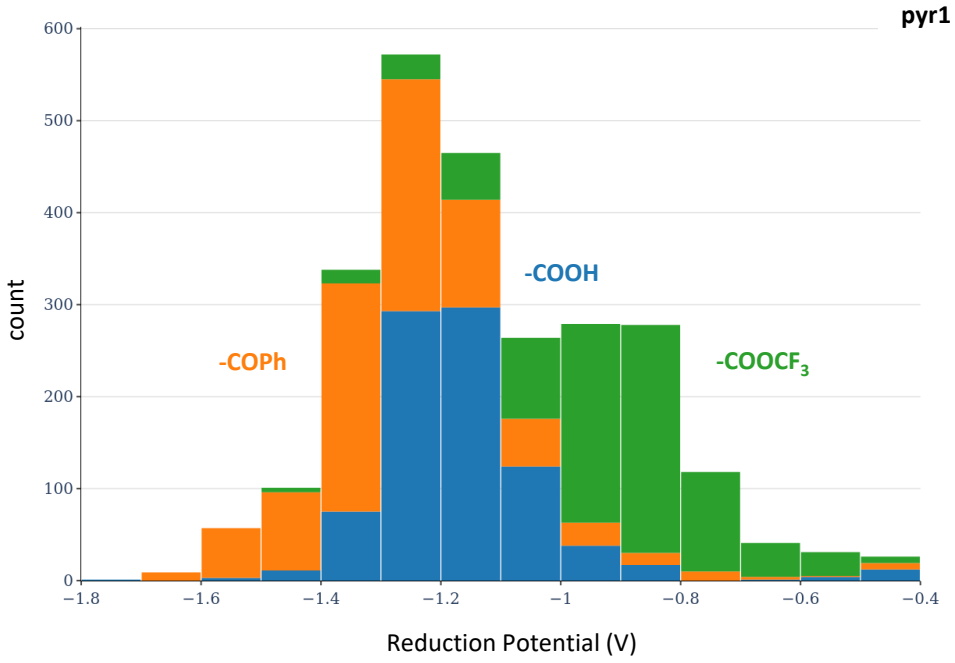


Figure 7: Distribution of calculated redox potential for the **pyr1** set as a function of R_3.

CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

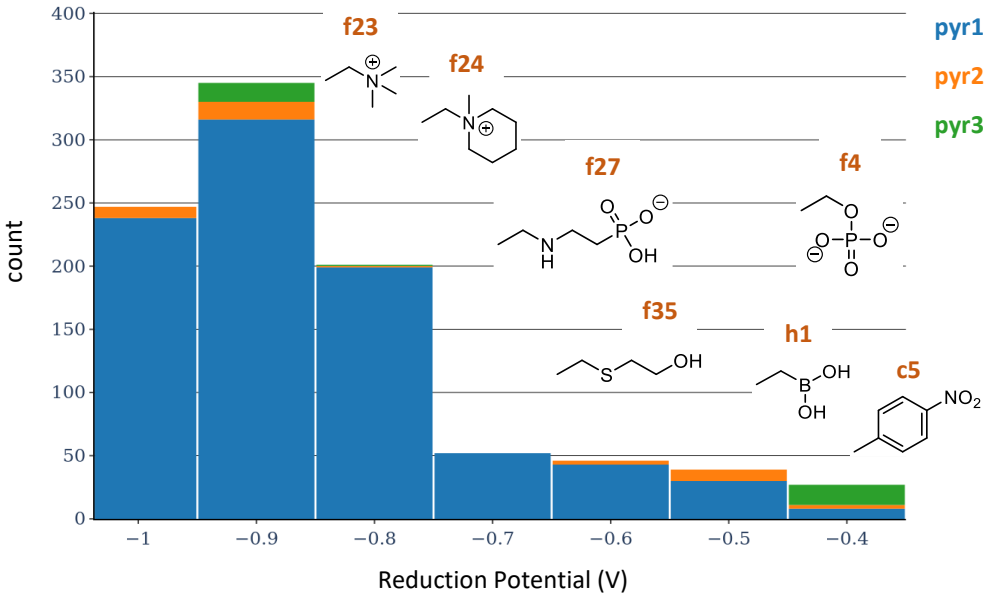


Figure 8: Distribution of calculated redox potentials in the most relevant region. A few representative R_2 groups are depicted on the diagram.

4.3 Analysis of computed solubilities

Solvation free energies were computed by using two different quantum chemical methods (xTB/GBSA and DFT/SMD) as described in Section 3. The scatter plot of the results obtained by the two approaches (Figure 9) shows reasonably good correlation for each group of molecules having different charges. This level of correlation indicates that the inexpensive xTB computational approach can be efficiently used to estimate solvation free energies of near DFT accuracy.

In Figure 10, the structures with reduction potentials in the range of interest are highlighted in green. The solvation free energies of these molecules are either in the 0 – -100 kcal/mol range, or below -200 kcal/mol, which are all double charged anions.

The majority of neutral species have solvation free energies below -50 kcal/mol (Figure 11), but a group from the **pyr3** set is displayed in the -80 – -60 kcal/mol range pointing to higher solubilities. This group, however, is characterized by reduction potentials out of electrochemical relevance (between -1.7 and -1.2 V).

Figure 12 depicts the overall distribution of computed solubility data in a histogram representation. Additional scatter plots focusing on the reduction potential region above -1.0 V are shown in Figure 13. These results can contribute to the design of additional molecular sets in the present project.

CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

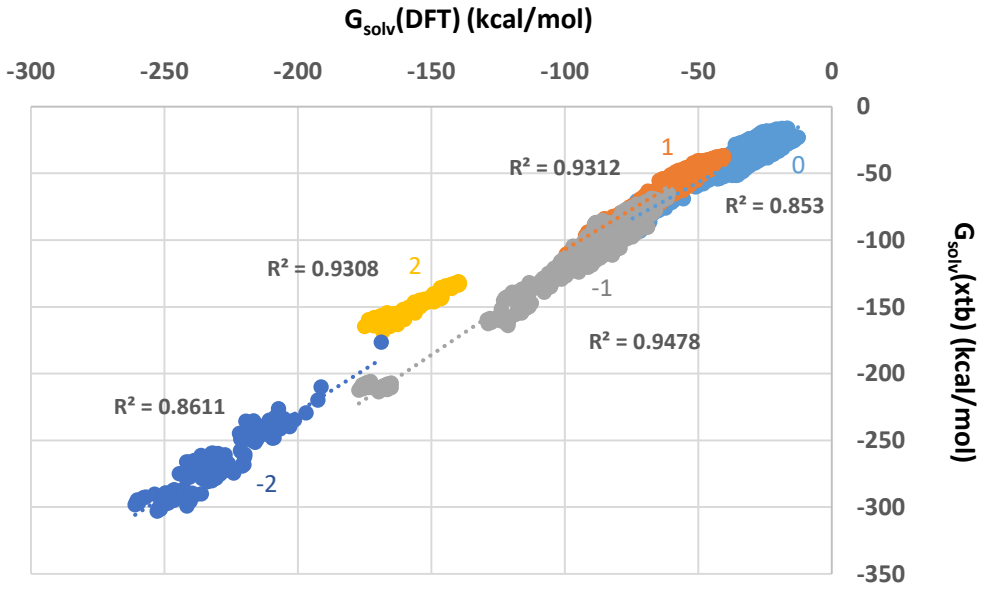


Figure 9: Scatter plot for correlation between the solubility calculated by DFT versus xTB methods. Color codes represent systems with different charges. For each group of a given charge the correlation coefficient is depicted.

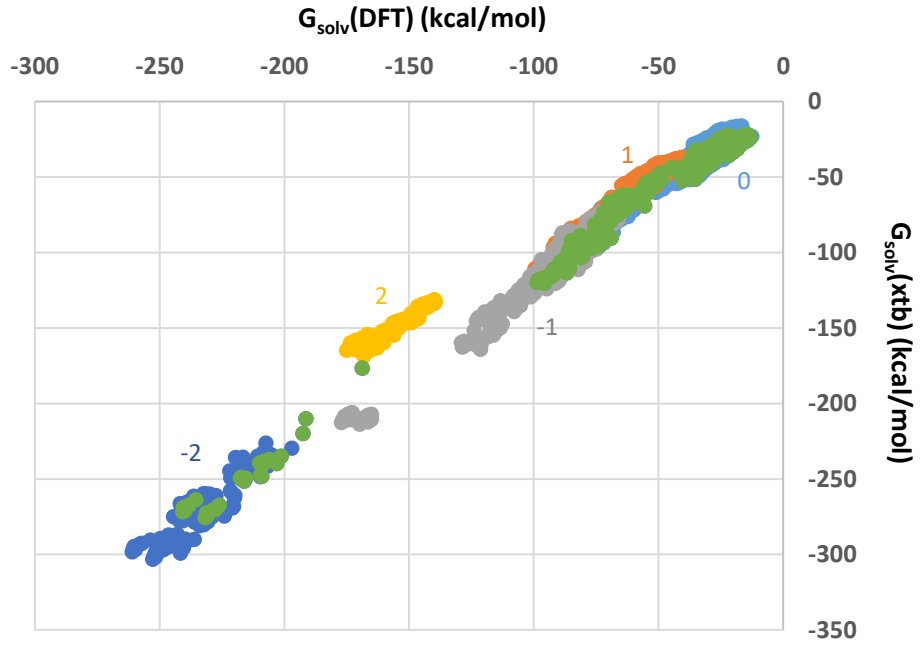


Figure 10: Scatter plot for correlation between the solubility calculated by DFT versus xTB methods. Color codes are according to the charges of the systems. Green dots show the structures with redox potentials above -1V.

CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

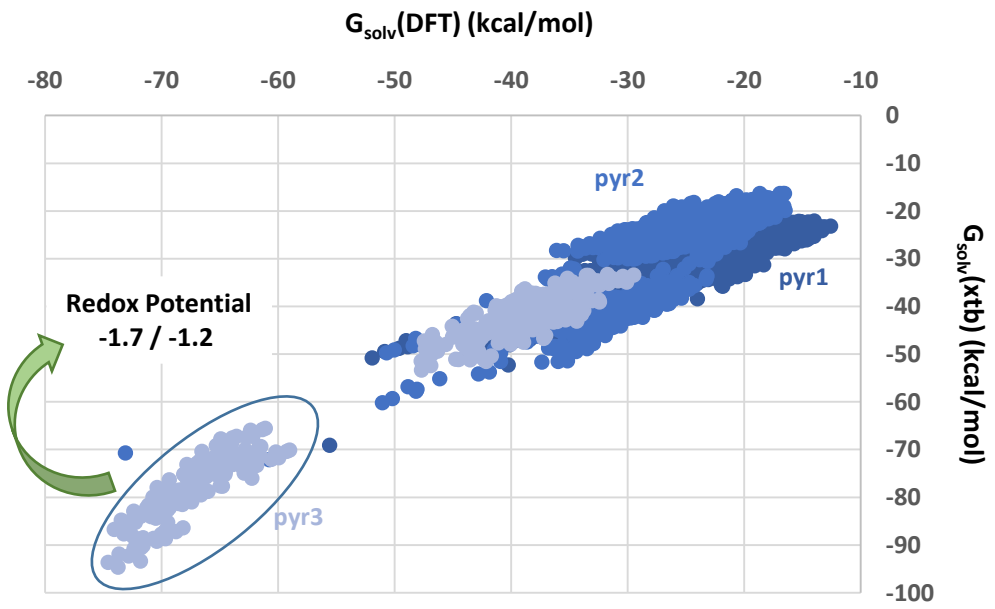


Figure 11: Scatter plot for correlation between the solubility calculated by DFT methods vs. xtb for the neutral molecules.

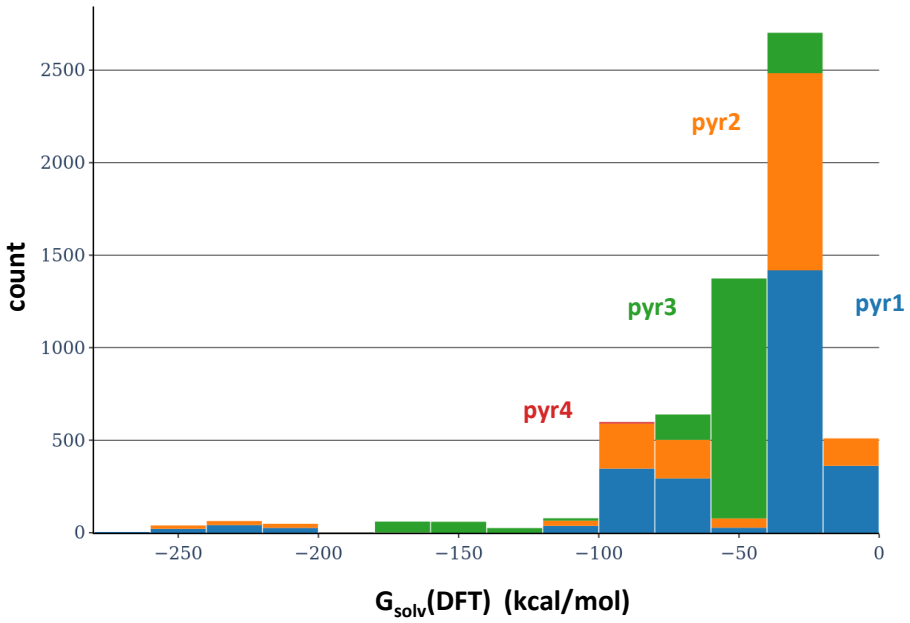


Figure 12: Distribution of the calculated solubilities. The color codes represent the molecular sets.

CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

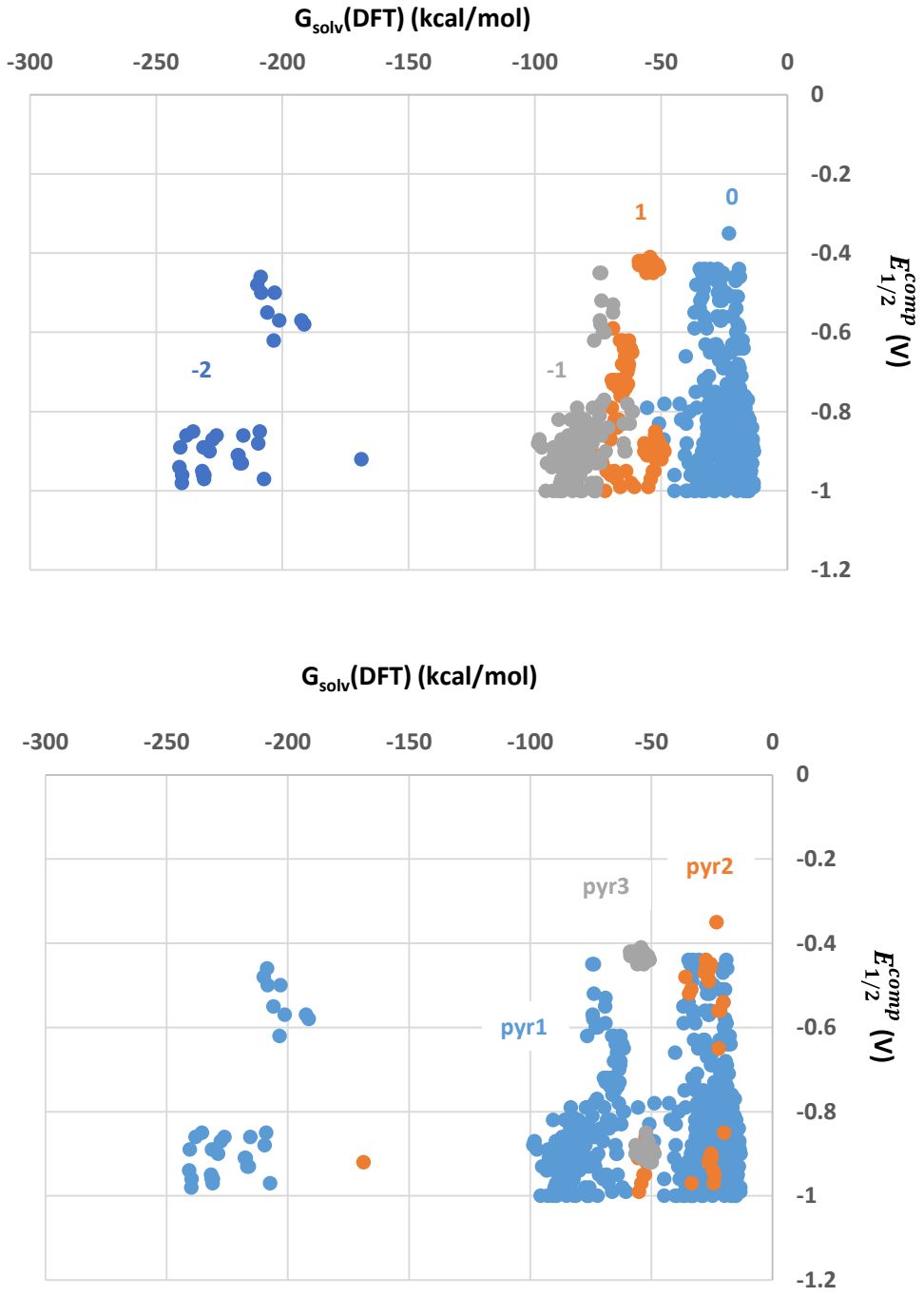


Figure 13: Scatter plots for calculated redox potentials as a function of the solubilities calculated by the DFT method in the most relevant region for the potential. The color codes are according to the charges of the systems (upper part), or according to the molecule sets (lower part).

CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

4.4 Descriptors for subsequent machine learning studies

Molecular properties derived from quantum chemical calculations can be considered as descriptors for subsequent machine learning analysis. The computed properties include electronic parameters, such as atomic charges (Mulliken or NBO), HOMO-LUMO energies, vibrational frequencies, which are all available from the output files of xTB and DFT calculations. In addition, the optimized molecular structures can be used to derive steric parameters as well, such as Sterimol parameters, for instance.

A variety of descriptors can be generated from the structure of a molecule using the *rdkit* tool, which is an open-source cheminformatics software [8]. These descriptors correspond to 2D and 3D representations of the molecule. The topological (2D) representation includes partial charges, number of hydrogen bond donors/acceptors, number of rotatable bonds, etc.; whereas the Coulomb matrix is common 3D representation. The descriptors can be generated instantly with the *rdkit* tool; thus, these data were not collected into tables. The selection of the relevant descriptors will be the subject of the subsequent machine learning analysis.

We also tested the generation of fingerprints and graphs with the *rdkit* tool, which are common molecular representations used in combined quantum chemistry – machine learning approaches [9], which will be exploited in our project as well.

4.5 Histogram analyzer

To assist the analysis of the present database, we developed a Python utility that enables the visualization of computed data via histograms and also viewing the Lewis structures of the molecules included in the database. The *HistPlotly* software is currently available for project members only as an online application. The built-in database for the 6204 compounds contains the calculated redox potentials, notations for substituents R_1, R_2, R_3, total charges of the original (oxidized) forms of the molecules, and the name of the molecular set: pyr1, pyr2, pyr3, pyr4. The data are tabulated, and specific columns can be selected for a colored histogram plot. Histograms presented in sections 4.3 and 4.4 were generated with this utility.

In addition to the histogram plots, we implemented several other functionalities. By selecting a cell of the potential value in the table, the Lewis structure of the corresponding molecule is shown under the View tab. By selecting a cell of columns R_1, R_2, R_3 or SYS, the structure of the corresponding substituent or molecular set is displayed. Each displayed structure can be downloaded in PNG or xyz format. Furthermore, the entire table, or a range of selected columns/rows can be saved in excel (.csv) format.

A few illustrative screenshots from the *HistPlotly* utility are shown in Appendix 4.

5 Conclusions

Virtual library of over 6200 pyridoxal-based molecules has been built, which includes the structures, reduction potentials and aqueous solvation free energies of these neutral and charged species. The properties were obtained using a composite quantum chemical approach. The suggested protocol

CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

comprises a recently parameterized semiempirical extended tight-binding method and high-level DFT calculations, and it provides reliable predictions for the target properties at a reasonable cost. Fortunately, restrictions due to Covid-19 pandemic had only minor effects on the activities of Tasks 1.1 and 1.2, because the work could be advanced in home office.

To reach milestone MS2, WP1 will continue with assessing various machine learning algorithms using the present molecular database. Related literature and our initial experience within Task 1.2 indicate. that the application of graph-based methods is a promising approach in high-throughput screening. However, the diversity of the molecular frameworks need to be clearly increased to arrive at an efficient and widely applicable screening method.

6 References

[1] Schrödinger modules used in the present project are licenced to TTK. For Schrödinger software package, see: https://www.schrodinger.com/platform#product-list-collapse.

[2] For key references on the semiempirical quantum mechanical methods GFNn-xTB and the CREST method, see: (a) Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parameterized for All spd-Block Elements (Z = 1-86), *J. Chem. Theory Comput.* 2017, *13*, 1989-2009; (b) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions, *J. Chem. Theory Comput.* 2019, *15*, 1652–1671; (c) Grimme, S. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations, *J. Chem. Theory Comput.*, 2019, *155*, 2847-2862.

[3] User guide of the *xtb* software, see: https://xtb-docs.readthedocs.io/en/latest/contents.html#

[4] Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals, *Theor. Chem. Acc.* 2008, *120*, 215-241.

[5] Gaussian 16, Revision A.03, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A. Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.;

CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2016.

[6] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions *J. Phys. Chem. B* 2009, *113*, 6378-6396.

[7] Roth, H. G.; Romero, N. A.; Nicewicz, D. A. Experimental and Calculated Electrochemical Potentials of Common Organic Molecules for Applications to Single-Electron Redox Chemistry, *Synlett.* 2016, *27*, 714-723.

[8] For the *rdkit* tool, see: https://www.rdkit.org/

[9] For selected contributions, see: (a) Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswingd, K.; Pande, V. MoleculeNet: a benchmark for molecular machine learning, *Chem. Sci.*, 2018, *9*, 513-530; (b) Jinich, A.; Sanchez-Lengeling, B.; Ren, H.; Harman, R.; Aspuru-Guzik, A. A Mixed Quantum Chemistry/Machine Learning Approach for the Fast and Accurate Prediction of Biochemical Redox Potentials and Its Large-Scale Application to 315 000 Redox Reactions, *ACS Central Science* 2019, *5*, 1199-1210; and references therein.

CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

compbat

Appendix 1: Consortium

COMPUTER AIDED DESIGN FOR NEXT GENERATION FLOW BATTERIES
COMPBAT

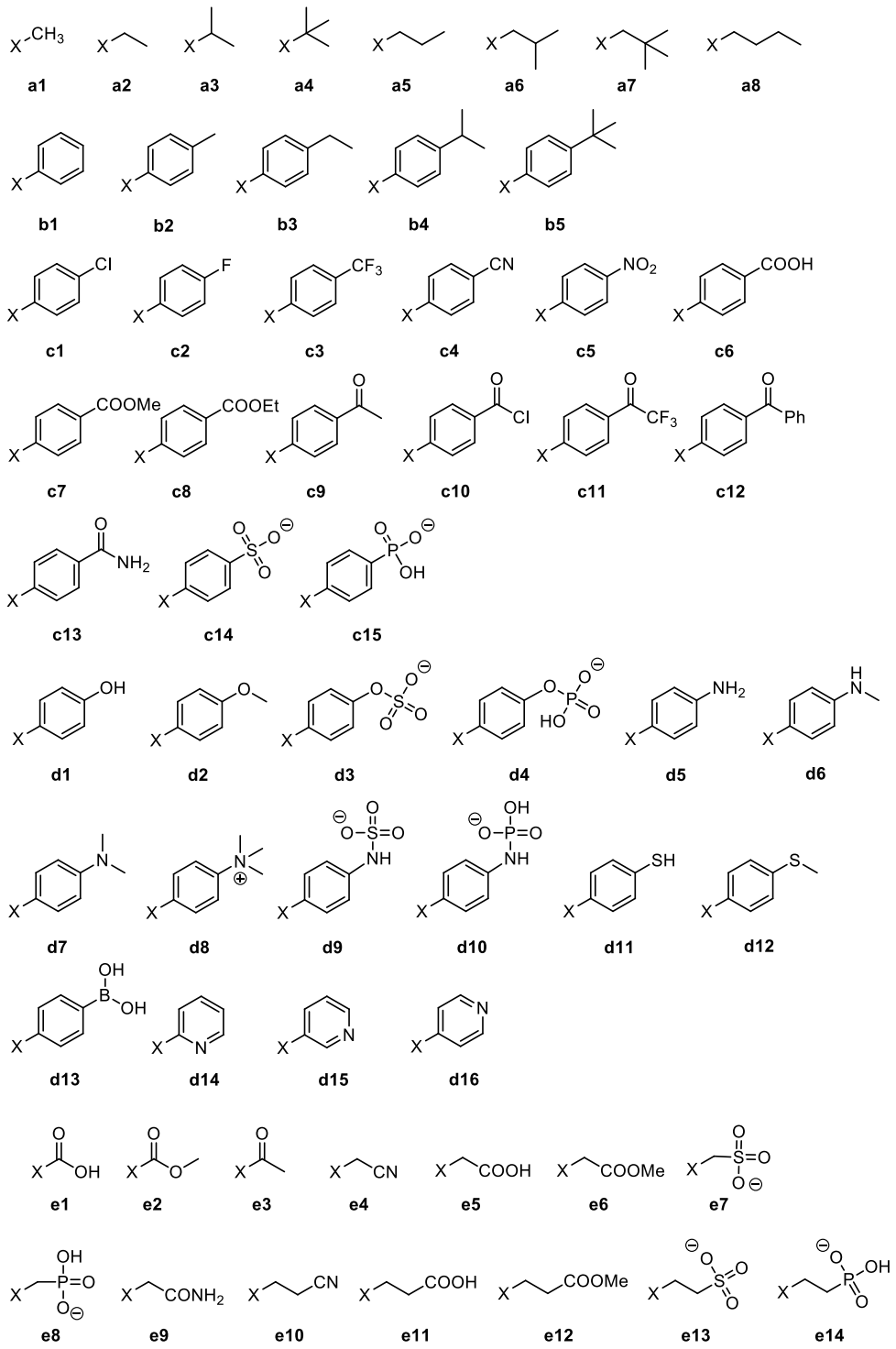
List of participants

Participant No.	Participant organisation name	Country
1 (Coordinator)	**Aalto Korkeakoulusaatio sr** Aalto University (Aalto)	Finland
2	**Természettudományi Kutatóközpont** Research Centre for Natural Sciences (TTK)	Hungary
3	**Uppsala Universitet** Uppsala University (UU)	Sweden
4	**Universita Di Pisa** Pisa University (UNIPI)	Italy
5	**Skolkovo Institute of Science and Technology** (SKOLTECH)	Russia
6	**Jyvaskylan Yliopisto** University of Jyväskylä (JYU)	Finland
7	**Turun yliopisto** University of Turku (UTU)	Finland

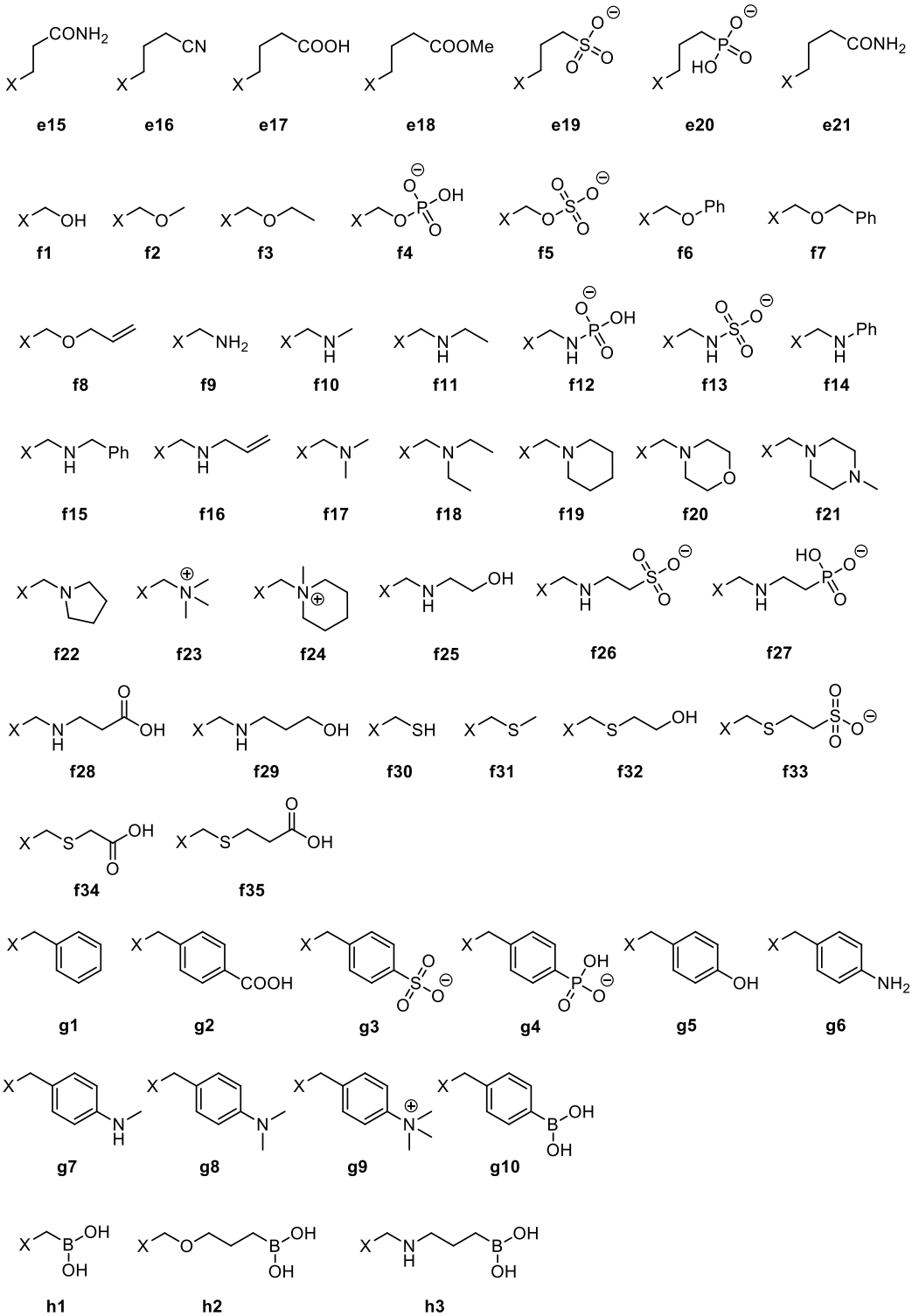
CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

compbat

Appendix 2. Full list of R_2 szubsztituents



CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.



CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

compbat

Appendix 3. Database in tabulated form

The computed reduction potentials and solubility data are provided in separate Excel documents (select-pot.csv and select-solv.csv). The first few lines of these tables are shown below for illustration. The reduction potentials are given in V, the Gibbs free energies of solvation are in kcal/mol.

Table select-pot.csv

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

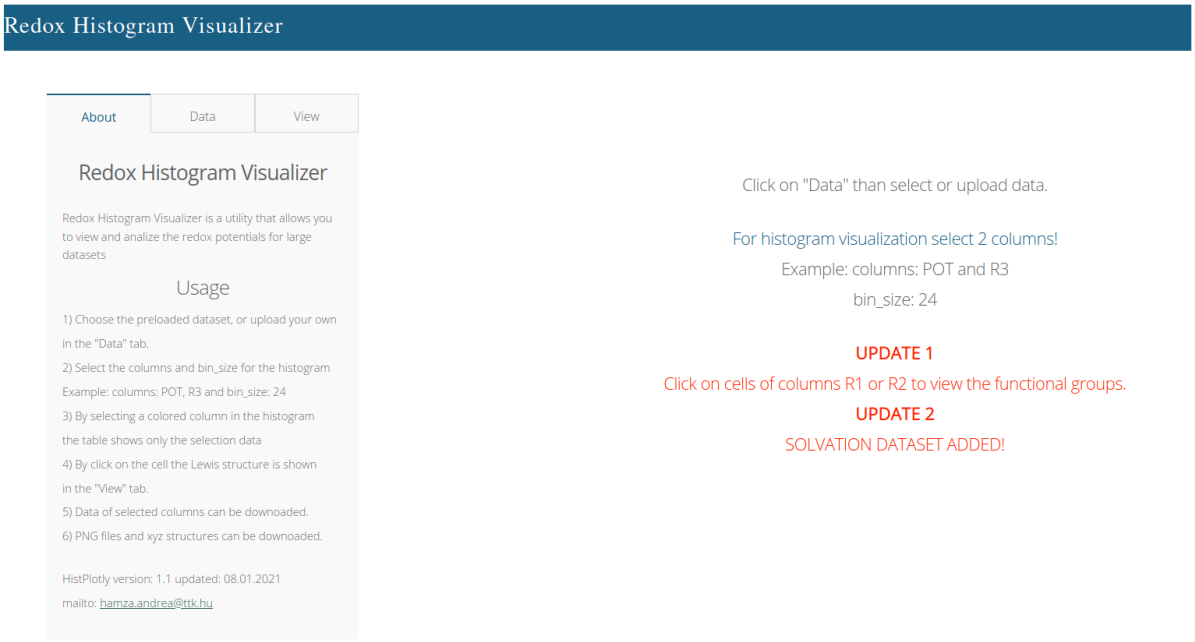
Table select-sol.csv

NR	Solv-DFT	Solv-XTB	R3	R1	R2	CHG	SYS
1	-29.4131	-37.7264	COOH	P1	a1	0	pyr1
2	-29.2332	-38.7235	COOH	P1	a2	0	pyr1
3	-28.6223	-39.1564	COOH	P1	a3	0	pyr1
4	-28.3208	-39.6555	COOH	P1	a4	0	pyr1
5	-29.0904	-38.9419	COOH	P1	a5	0	pyr1
6	-28.5269	-39.1215	COOH	P1	a6	0	pyr1
7	-28.2802	-39.0049	COOH	P1	a7	0	pyr1
8	-28.8423	-39.0854	COOH	P1	a8	0	pyr1
9	-33.4616	-40.4635	COOH	P1	b1	0	pyr1
10	-33.2069	-40.8989	COOH	P1	b2	0	pyr1
11	-32.9433	-41.0249	COOH	P1	b3	0	pyr1
12	-32.746	-41.3413	COOH	P1	b4	0	pyr1

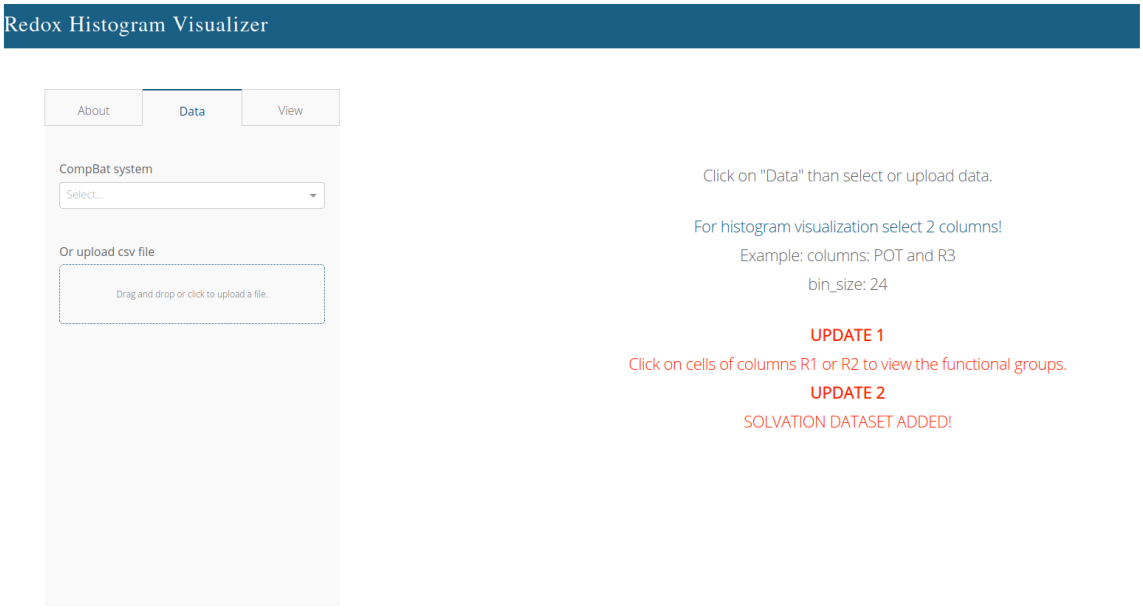
CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

Appendix 4: Screenshots from the *HistPlotly* utility

The introductory page of ***HistPlotly.***

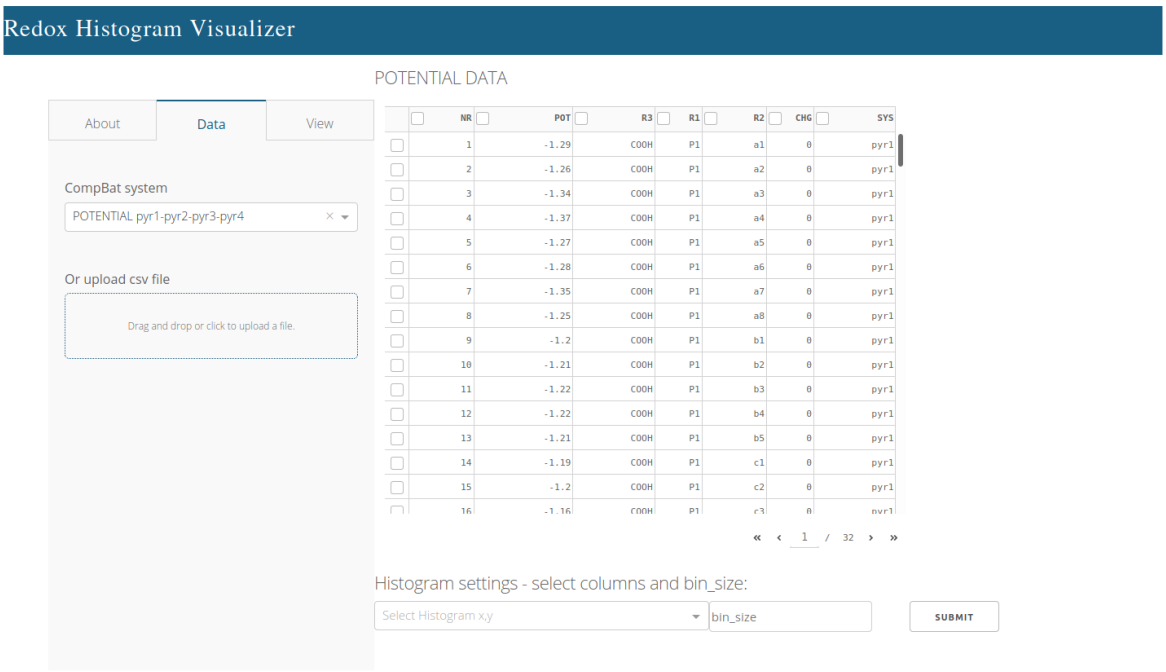


Choice of the dataset.

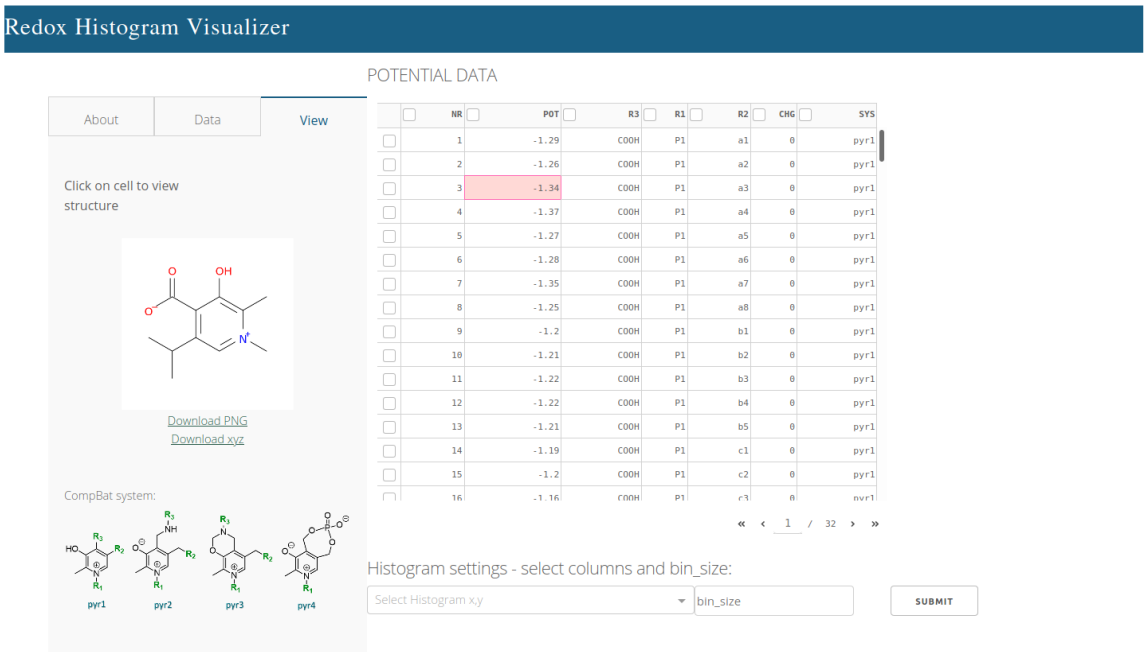


CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

Reduction potential data appearing in a tabulated form and histogram settings.

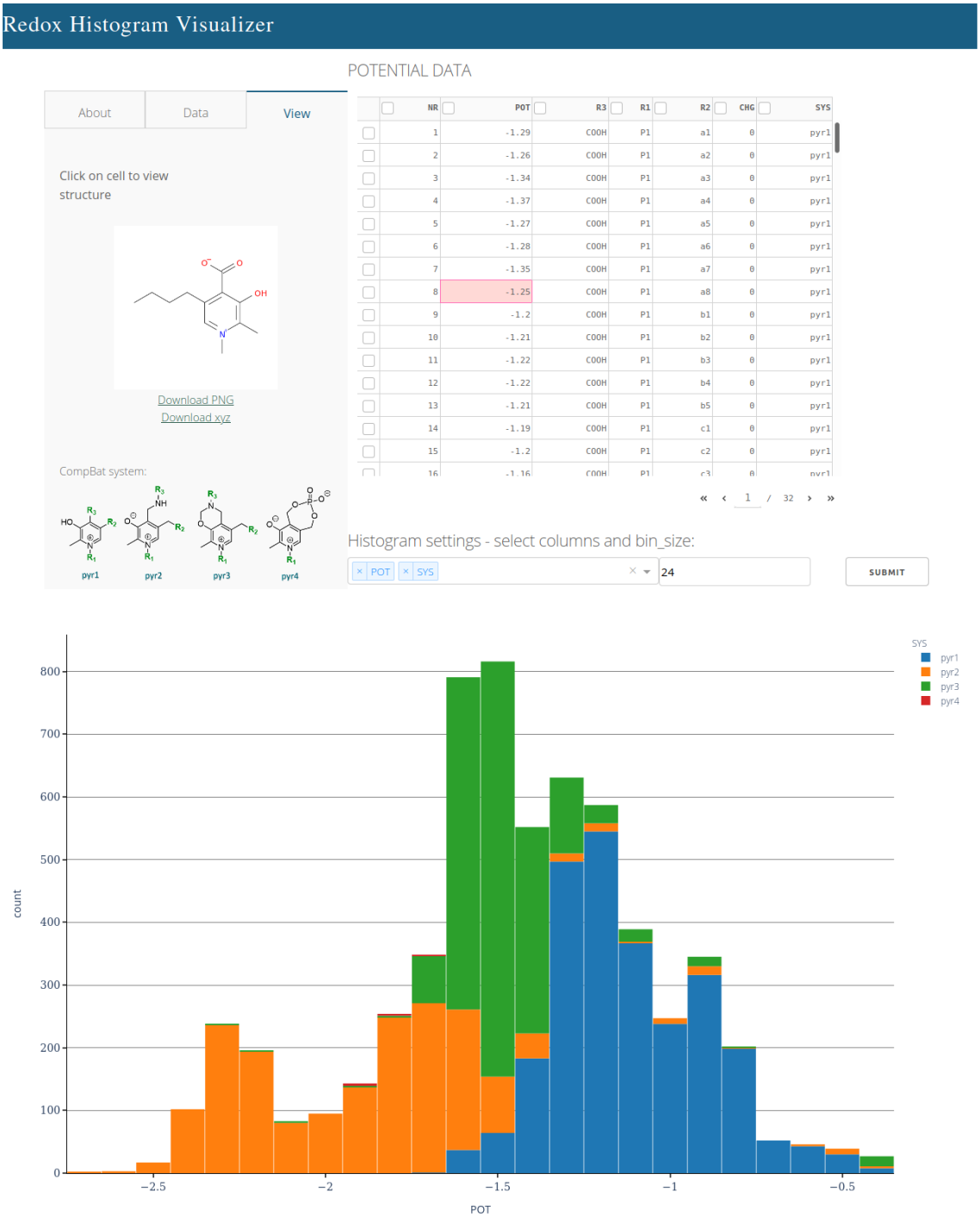


Molecular viewer.



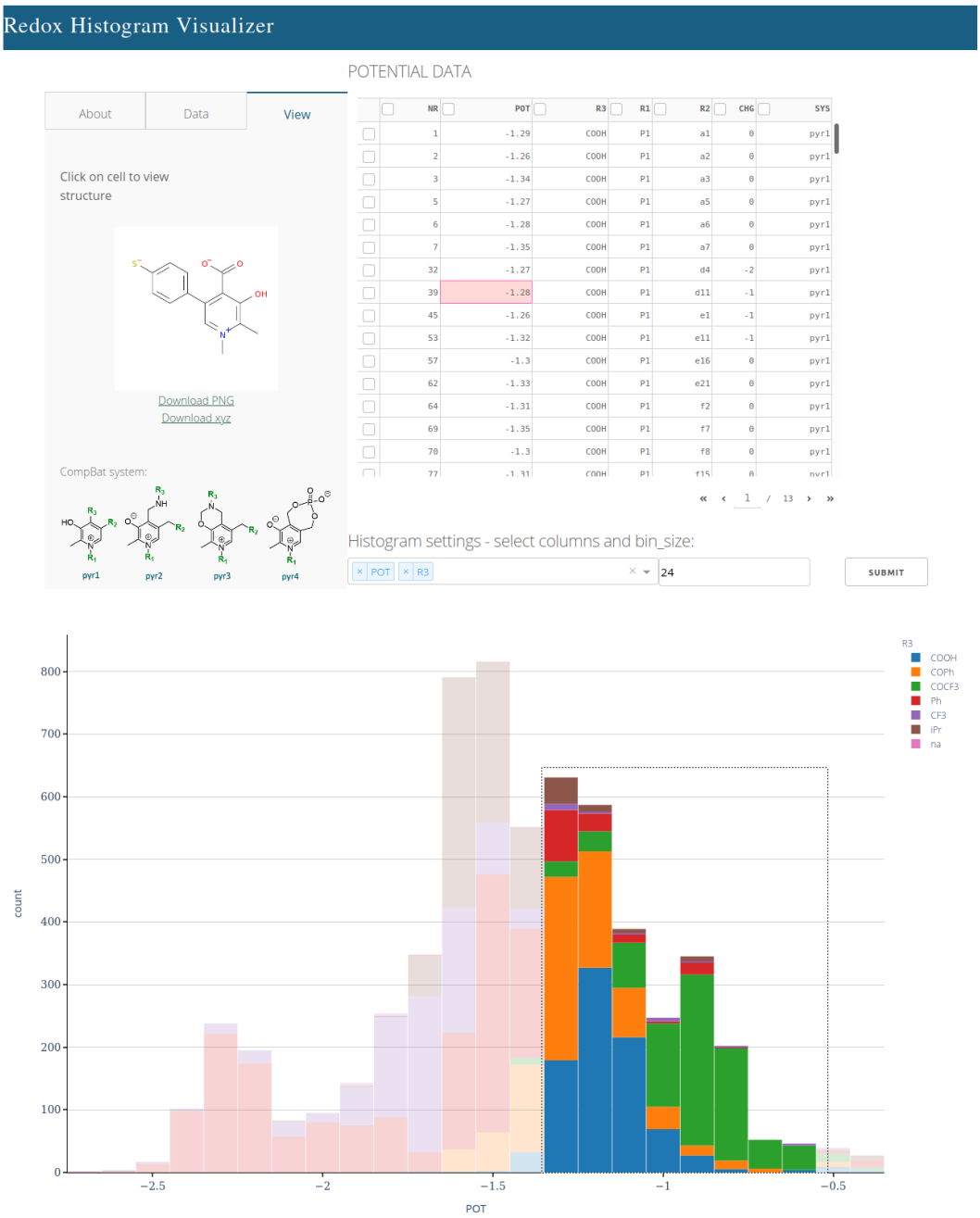
CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

Histogram with respect to molecular sets.



CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

Selection of a particular potential range.



CompBat project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 875565. This document has been produced by the CompBat project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.