Supporting Information

AuLCA: Augmented Life Cycle Assessment for Chemical Data Gaps

Maximilian G. Hoepfner[1,2‡], Dion Jakobs[1‡], Lucas F. Santos[1,2], Gonzalo Guillén-Gosálbez[1,2*]

[1]*Institute for Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, ETH Zurich, Vladimir-Prelog-Weg 1, 8093 Zurich, Switzerland*
[2]*NCCR Catalysis, Zürich CH-8093, Switzerland*
* *E-mail: gonzalo.guillen.gosalbez@chem.ethz.ch*
‡ The authors M.G.H and D.J contributed equally.

Contents

1. Data Curation Strategy for Chemical Reaction Networks

1.1. Open-Source Chemical Reaction Database

1.37 million reactions in the SMILES notation are sourced from the open-source CRD [1] and are converted into structured reaction dictionaries within python. Reaction compounds were classified as reactants, reagents, solvents, or products, followed by the computation of their mass-based stoichiometries based on their mole stoichiometry sourced from the reaction database itself. Solvent classification was carried out using a selection of 188 most commonly used solvents [2], which were translated into their corresponding SMILES representations. For each reaction, the presence of any of these solvents was checked. If a match was found, the identified compound was excluded from the reaction stoichiometry and instead assigned the role of solvent.

Subsequently, the reactions were connected by matching common reagents or products, to build a as highly interconnected chemical reaction network (CRN) as possible. As the OS data is mostly based on very complex reactions, e.g., patents rather than well-established technologies, it is assumed that reactions can be reversed, if necessary, to enable a higher degree of interconnection in the CRN. This is purely a modelling decision, and must not be taken as evidence that the reactions are reversible under experimental conditions. For an exemplary reaction, e.g. A → B + C with B and C in the chemical corpus, it is assumed that we can compute the impact of A by reversing the mass-based allocation to the compute the reactants A.

1.2. Reaxys© Database

Reaction data from the Reaxys© database*[3] is queried via an API that selects target chemicals and reactions from the available >347 million chemicals and >70 million reactions. The queries are structure to select chemicals and reactions based on user-defined criteria, such as CAS number, chemical structure, chemical properties and role in the chemical reaction. By querying only for a relevant subset of all available reactions, a local chemical reaction network (CRN) of interest is generated.

To generate the CRN, an initial set of seed chemicals are selected, including target molecules to be predicted and source molecules that have associated LCIA data to be used for the impact propagation (molecules in the corpus). Based on the chemicals selected, reaction queries are sent via the API to identify reactions in which the known chemicals are reactants (forward search) or in which the known chemicals are products (backwards search). The API returns a list of reaction hits that are ordered by the number of citations from which all available information, including reactants, products, reagents, solvents, etc. can be extracted. To ensure the reliability of the reactions collected, a set of filtering criteria is implemented to filter out any improbable or faulty reactions. The new chemicals retrieved from the reaction queries are collected and used as the initial set of chemicals from which the querying process can be repeated again. This recursive expansion of the reaction network can be repeated as many times as necessary to create a dense CRN. By iteratively expanding outwards from chemicals with known environmental impacts and target chemicals of interest, this approach ensures that all chemicals are sufficiently linked to one another to enable the consistent prediction of LCA impacts.

Reactions collected from Reaxys© do not contain any information on the reaction stoichiometry. As the reaction stoichiometry is necessary to complete the impact augmentation, optimization is used in order to estimate the reaction stoichiometry. For each reaction, the known reactants, reagents, and products from Reaxys© are compiled into an LP optimization model that attempts to identify a feasible stoichiometric solution to the reaction equation with regards to the elemental balance.

For any reaction, we first define the set of chemicals $C := R \cup Rg \cup P$ as the union of the non-overlapping sets of reactants R, reagents Rg, and products P of the reaction ($R \cap Rg \cap P = \emptyset$). In the Reaxys© database, reactants and products are chemicals that are known to either be consumed or produced in the chemical reaction, i.e. chemicals that are known to have a non-zero stoichiometric coefficient. Alternatively, chemicals that are classified by Reaxys© as reagents are considered substances that have been added to a chemical system to either cause a chemical reaction or test if a chemical reaction has occurred. This can include additional reactants, as well as non-consumed substances such as catalysts, solvents, indicators, etc.. Therefore, these reagents could be present in the reaction stoichiometry, but do not have to be utilize (i.e. stoichiometric coefficient less than or equal to zero).

To solve the stoichiometric equation, we defined an objective function of the LP stoichiometric optimization function that selects the smallest chemically feasible values of the reaction stoichiometric coefficients v_i for all reaction chemicals i in the set C:

$$\min_{v_i} \sum_{i' \in P} v_{i'} - \sum_{i'' \in R \bigcup Rg} v_{i''} \#(S1)$$

That is, the model seeks to minimize the sum of the positive values of the reaction stoichiometry while subject to the reaction stoichiometry constraints defined as follows:

$$\sum_{i \in C} v_i a_{i,e} = 0 \; \forall \; e \in E \; \#(S2)$$

$$1 \leq v_i \leq 100 \; \forall \; i \in P \; \#(S3)$$

$$-100 \leq v_i \leq -1 \; \forall \; i \in R \; \#(S4)$$

$$-100 \leq v_i \leq 0 \; \forall \; i \in Rg \; \#(S5)$$

To be defined as chemically feasible, the reaction stoichiometry optimization model must fulfill both the elemental balance (Equation S2) and must have a non-zero stoichiometric coefficient for all known reactants and products. The model may choose to utilize reagents as additional reactants to fulfil the elemental balance. To represent the elemental balance as a constraint, we define the set of elements E present in the reaction, where parameter $a_{i,e}$ denotes the number of atoms of element e present in chemical i for all elements in E and all chemicals in C. The sum of all elemental species must be equal on the reactant and product sides.

Should the LP reaction stoichiometry optimization model be found to be infeasible, it is likely that either a necessary side reactant or side product is missing. In this case, an MILP stoichiometric optimization model is used that enables the usage of addition of auxiliary chemicals that can be added to either the reactant or product side of the reaction equation to properly balance the reaction equation.

Using the same sets and parameter as in the LP stoichiometry optimization problem, we define the new set Ax to represent the auxiliary chemicals that can be used as side products or reactants and a set $D = \{React, Prod\}$ to define the side of the reaction equation the auxiliary chemicals are used in. With these additions we can define the new continuous and binary variables $\hat{v}_{i,d}$ and $y_{i,d}$ and the parameter $\hat{a}_{i,e}$, where $\hat{v}_{i,d}$ represents the selected stoichiometry of auxiliary chemical i in the set Ax on side of the reaction d, $y_{i,d}$ represents the binary selection of auxiliary chemical i in the set Ax on side of the reaction d and $ax_{i,e}$ represents the number of atoms of element e present in chemical i for all elements in E and all auxiliary chemicals in Ax. The MILP optimization can thus be formulated as follows:

$$\min_{v_i \, \hat{v}_{i,d}} \sum_{i' \in P} v_{i'} - \sum_{i'' \in R \bigcup Rg} v_{i''} + 100 \sum_{i''' \in Ax} \left(-10\left(\hat{v}_{i''',React} + y_{i''',React}\right) + \left(\hat{v}_{i''',Prod} + y_{i''',Prod}\right)\right) \#(S6)$$

Subject to the following constraints:

$$\sum_{i \in C} v_i a_{i,e} + \sum_{i' \in Ax \, d \in D} \hat{v}_{i',d} \hat{a}_{i',e} = 0 \; \forall \; e \in E$$

$$1 \leq v_i \leq 100 \; \forall \; i \in P \; \#(S7)$$

$$-100 \leq v_i \leq -1 \; \forall \; i \in R \; \#(S8)$$

$$-100 \leq v_i \leq 0 \; \forall \; i \in Rg \; \#(S9)$$

$$-100 y_{i,React} \leq \hat{v}_{i,React} \leq 0 \; \forall \; i \in Ax \; \#(S10)$$

$$0 \leq \hat{v}_{i,Prod} \leq 100 y_{i,Prod} \; \forall \; i \in Ax \; \#(S11)$$

Hence, the MILP seeks to minimize the sum of the positive values of the reaction stoichiometry including a penalty term for selecting auxiliary chemicals to balance the stoichiometric reaction equation. Both the auxiliary chemical stoichiometric coefficient and the binary selection variable are included in the objective function to both penalize the selection of singular auxiliary chemicals with very large stoichiometric coefficients and the selection of more auxiliary chemicals than necessary. The inclusion of auxiliary reactants is more heavily penalized than the inclusion of auxiliary products, as it is more likely that

the Reaxys© database is missing a product than a reactant. The MILP model is subject to a similar elemental balance constraint and constraints of the stoichiometric coefficients as in the LP formulation. The elemental balance constraint is extended to include the stoichiometry of any selected auxiliary chemicals. As before, the model must select non-zero coefficients for the reactant and product chemicals, and may decide to include reagent chemicals on the reactant side with no additional penalty in the objective function. Standard additional upper and lower bound constraints are applied to the stoichiometric coefficient of the auxiliary chemicals that activate or deactivate the variable $\hat{v}_{i,d}$ based on the value of $y_{i,d}$. When the binary $y_{i,d}$ is set to 1, this auxiliary chemical is included in the reaction stoichiometry, objective function, and the elemental balance, while when $y_{i,d}$ is 0, then that auxiliary chemical is not included in the reaction stoichiometry and in fulfilling the elemental balance. An additional constraint that forces the auxiliary chemicals to only appear on one side of the reaction could also be included, that is:

$$y_{i,React} + y_{i,Prod} \leq 1 \; \forall \; i \in Ax\#(S12)$$

However, this constraint is not necessary as a solution with a stoichiometric coefficient for the same auxiliary chemical on both sides of the reaction equation is always more heavily penalized than (i.e. is always non-optimal compared to) an equally feasible solution with the auxiliary chemical only being selected on one side of the reaction stoichiometry. Thus solutions with auxiliary chemicals on both sides of the reaction equation will never be selected.

A key advantage of utilizing Reaxys© as a data source for the CRN is the ability to specifically tailor the CRN to improve the AuLCA prediction. By controlling the size and density of the network, and ensuring that all specific chemicals of interest are present and densely connected in the network, it is possible to improve the amount of data available to the AuLCA methodology and, thus, improve the performance of the algorithm.

Despite the added benefits, generating a CRN with Reaxys© can be a very time intensive process, being an input-output bound problem requiring many thousands of API calls to construct the network. Additionally, accessing data from Reaxys© requires purchasing a License and an API key.

1.3. Comparison of Both Strategies

Both CRN strategies have advantages and disadvantages. While for the OS CRN, the data is readily available in form of open-access chemical reaction databases, Reaxys© data requires a license and an API key.

Besides this, Reaxys© provides fundamental advantages to build chemical reaction networks tailored to our needs. While with Reaxys© we can query to construct a CRN based on chemicals we have from ecoinvent, for the OS CRN we need to work with the reactions available therein. The only parameter to construct a denser reaction network potentially containing more chemicals from ecoinvent is the size of the reaction network itself. Specifically, we showed that larger CRN such as in CS III with about 300k fail to include all chemicals from ecoinvent. Moreover, the data available in OS CRD is limited to around 1.3 million reactions[1], while this is not the case for Reaxys© with around >70 million of reactions.

In Figure S1, for illustrative purposes, the different strategies are compared graphically. For the OS CRN (left) the size of the training set is heavily dependent on the size of the CRN. The training set SK_0 is always a subset of the corpus (SEI), as not all chemicals in the latter appear necessarily in the CRN. For the Reaxys© case, the corpus chemicals SEI can be queried to be included fully in the CRN, thereby ensuring a better coverage. Therefore, the training set SK_0 can match SEI when all chemicals are queried.

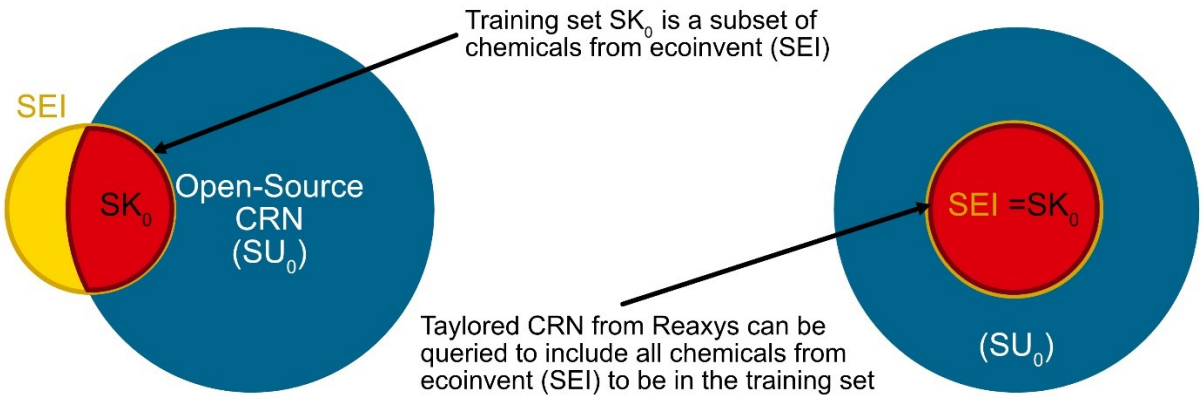

Figure S1: Visualization of the two employed CRN curation strategy, left based on open-source CRNs and on the right based on Reaxys© CRNs.

2. Computational Sequence of the AuLCA Algorithm

The following examples illustrate how the algorithm works. The approximation of reactants by using proxies as visualized in Figure S3 mostly applies to open-source CRN, as for Reaxys© curated CRNs the availability factor (AF) is 1 in most of the cases.

2.1. Calculation Sequence of the AuLCA Algorithm based on Reaction Ranking using the Availability Factor (AF)

This section provides, with the help of some examples, details on how the algorithm computes the unknown LCIs sequentially. In Figure S2, a straightforward calculation sequence of the AuLCA algorithm is visualized based on three reactions using three precomputed chemicals SK_0 from ecoinvent. In the first iteration, the compound with the highest according AF is selected, being 3. Subsequently, the next iteration of AF calculation reveals AF values of 1 for all chemicals SU, which will therefore be calculated in the 2-4 calculations. After all iterations, all substances belong to the set SK, since all previously unknown chemicals have been calculated.

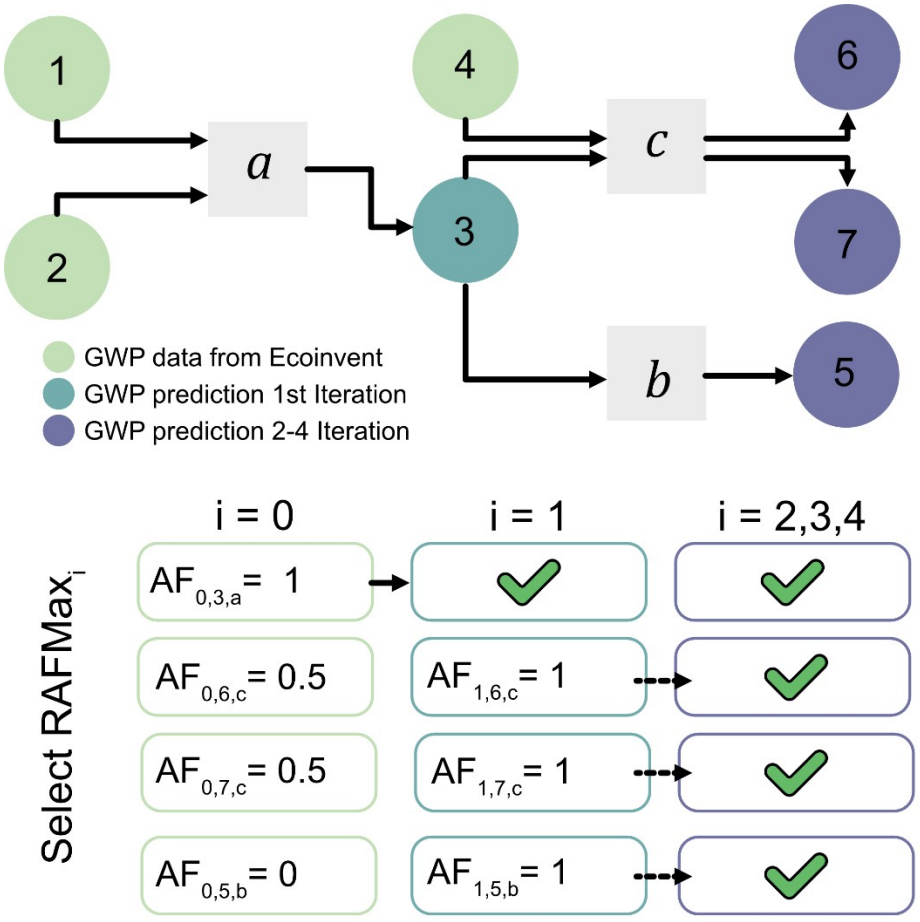

Figure S2: General calculation order of AuLCA based on a simple reaction system.

In Figure S3, a different system of equations (e.g., part of a CRN) is visualized which shows the impact of the AF on the calculation sequence. As before in Figure S2, chemical 3 is calculated first, due to the AF of 1. The subsequent step differs in Figure S3, since a new compound N is introduced, which does not have precomputed GWP, or is computed elsewhere in the CRN. Therefore, after calculation of 6 and 7 (due to the higher AF of 1 in both cases, against AF of 0.5 for 5), 5 is calculated by using a proxy GWP (compare eq. 12).

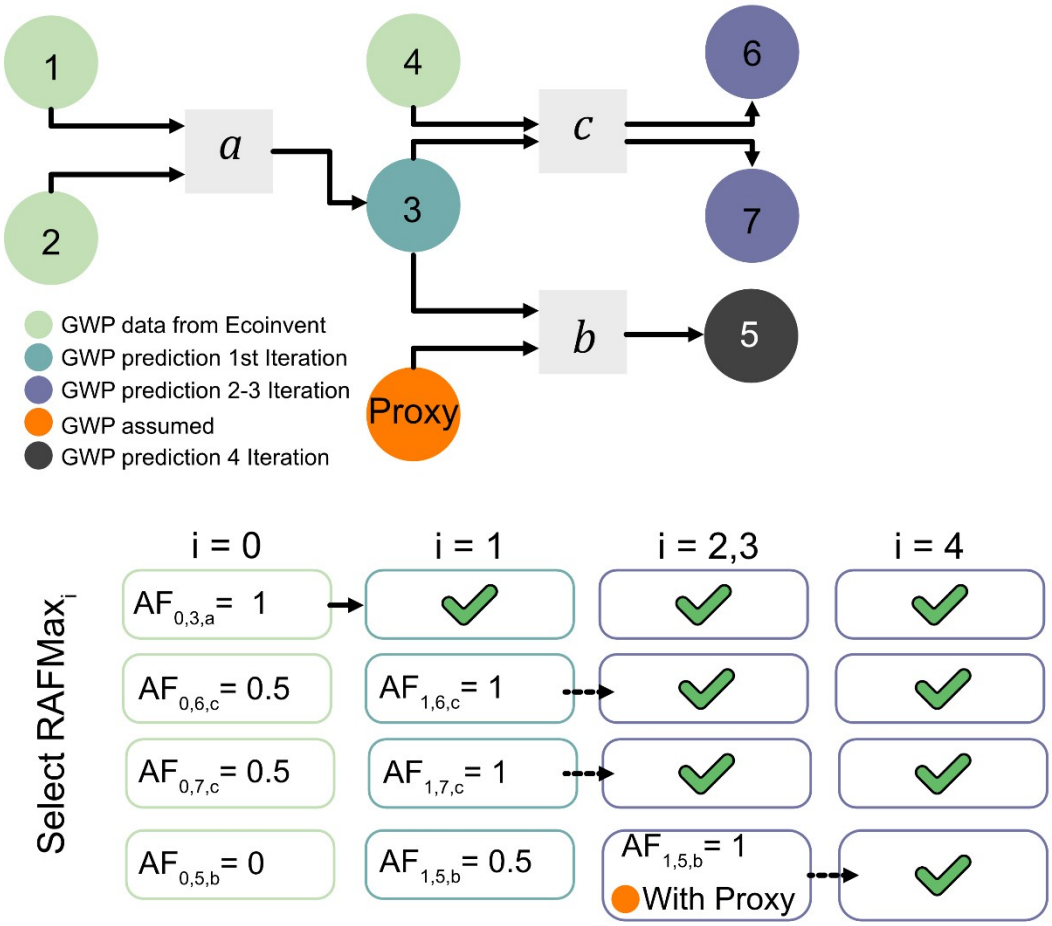

Figure S3: Calculation order of AuLCA based on a reaction system, where proxy data is essential due to a lack of data. For simplicity, the iteration index i is not included.

Figure S4 visualizes the case, where multiple reactions might be used to calculate chemical 7. The AF guided calculation indicates that 7 can be estimated through reaction c and d. However, due to the higher AF in reaction d, the LCI of 7 is computed from *d*. In later iterations, reaction c is solely used to calculate 6. If both reactions, c and d, showed the same AF (e.g., if 3 is in the set SEI), then both would be used for the impact augmentation and the mean would be calculated.

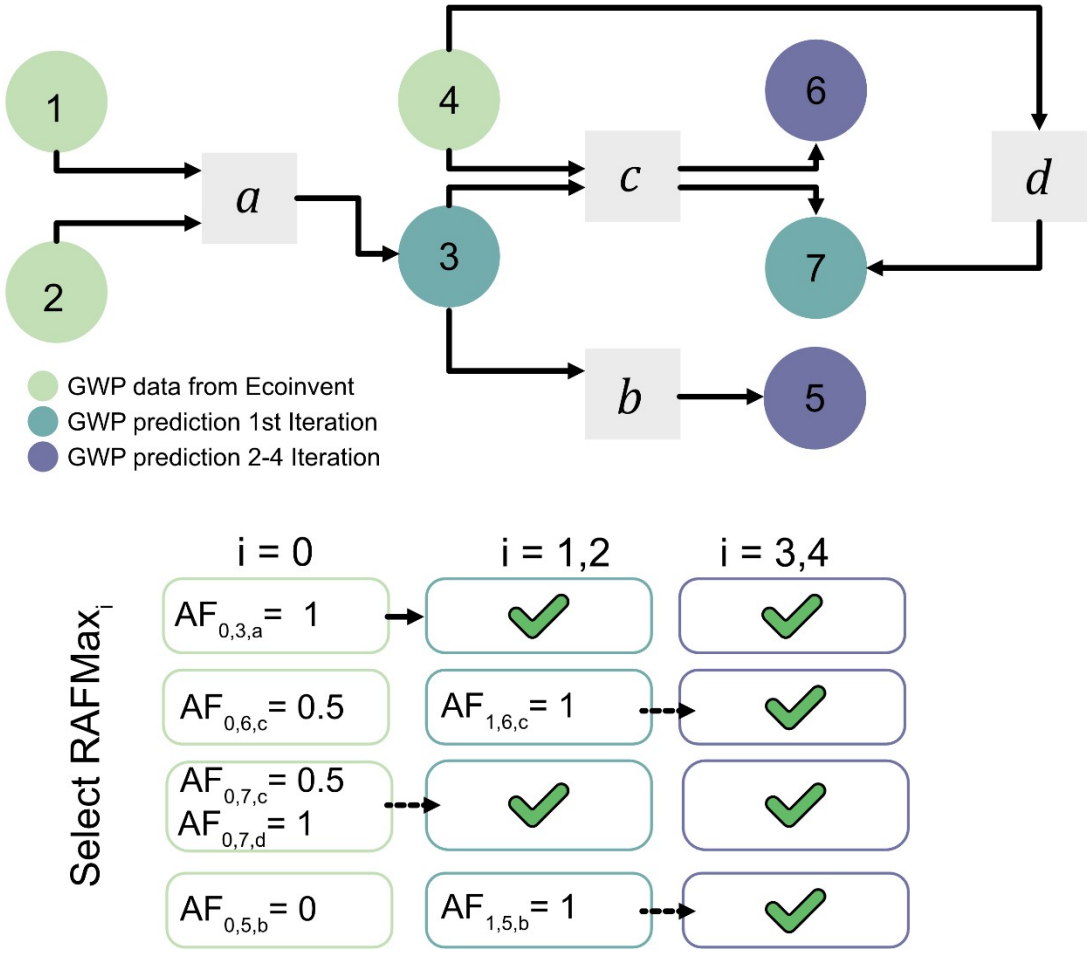

Figure S4: Calculation order of AuLCA, based on a reaction system with multiple reactions for one chemical. For simplicity, the iteration index i is not included.

2.2. Assumptions for chemicals involved in the impact propagation

If proxies are required during impact propagation (e.g., Figure S3), the life-cycle inventory (LCI) for a missing reactant is estimated using the equation below. This formula takes the weighted average of the available LCIs of the co-reactants. We then assign this average as the proxy LCI for the missing reactant. It should be noted that the AF ranking already aims to avoid such cases, but depending on the CRN and the associated data availability (e.g., low interconnected) it might be necessary to provide ad-hoc approximations.

$$ProxyLCI_{j,n} = \frac{\sum_{n' \neq n, n' \in SIN_r} \nu_{n',r} \, LCI_{j,n'}}{\sum_{n' \neq n, n' \in SIN_r} \nu_{n',r}} \quad \#(S13)$$

$\forall\, i,j,n \in SIN_r \cap SU_i$

3. Leave-one-Out-Validation Overview

3.1. Reference values from ecoinvent

Figure S5 shows that most chemicals in ecoinvent have a molecular weight in the range 118.8 - 133.7 g/mol and a GWP from 4.31 - 5.18 $kgCO_2$-eq (95% confidence interval, n = 429, outliers removed). This data indicates that most chemicals in ecoinvent v3.9.1 are bulk chemicals, while only few complex fine chemicals are covered.

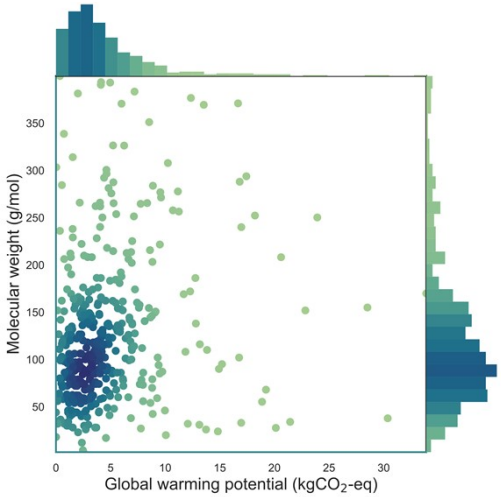

Figure S5: Molecular weight and GWP distribution of reference chemicals in ecoinvent v3.9.1. GWP calculation based on IPCC 2021 GWP 100a using brightway. This dataset shows the general trends in the ecoinvent data. Not all data curation strategies such as in chapter 3.2 are applied.

3.2. Data Curation Strategy for Corpus

In order to retrieve a suitable corpus of chemicals, several filtering criteria need to be applied, in addition to manually checking the data. This is of particular importance, as the impacts will be propagated throughout the whole CRN. We follow the filtering criteria by Lucas et. al [4] to the chemical activities in the ecoinvent database (Figure S6, Step 1). All chemicals are translated into their SMILES, prior to subsequently filtering out duplicates (e.g., one chemical with different activity location; activities with location "GLO" were kept if possible). In the following step, the retrieved data was manually analyzed (compounds with a GWP > 150 $kgCO_2$-eq/kg were removed), to identify any outliers (e.g., heavy water with a GWP > 1000 $kgCO_2$-eq/kg). Outliers, such as the heavy water, will heavily influence the impact augmentation and, therefore, lead to less accurate predictions (Figure S6, Step 2). The dataset is denoted as the corpus of chemicals SEI. Each CRN can utilize the chemicals in the corpus as training set SK_0. However, in most cases (e.g., small open-source CRNs) the training set SK_0 is much smaller than the chemical corpus SEI. This is because not all chemicals in the corpus SEI appear in the CRN, so they cannot be used for data augmentation. Last but not least, filter criteria discussed in the next section apply to build the LOOV validation data set SK_{LOOV} (Figure S6, Step 3). The procedure is visualized in Figure S6.

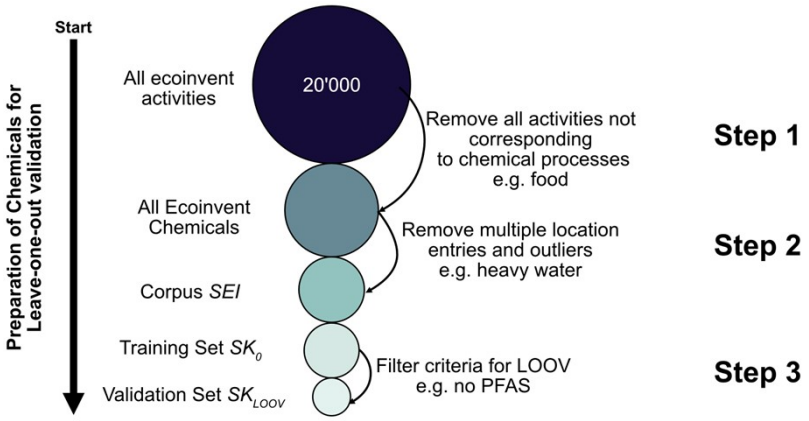

Figure S6: Filtering criteria [4] and additional approaches to retrieve chemicals for the leave-one-out validation from the LCA database ecoinvent.

3.3. Data Curation for the LOOV Data Set

Many chemicals in the training set SK_0 are not suited for the leave-one-out validation (LOOV). For example, chemicals such as $SiCl_4$, which are not in the application domain of AuLCA (e.g., we focus on organic chemicals), were excluded (Figure S6, Step 3) for the LOOV and are, therefore, removed from the validation set SK_{LOOV}.

Figure S7 shows the results generated when using raw data, e.g., the case when none of the chemicals from the training set SK_0 are excluded for the LOOV $(SK_0 = SK_{LOOV})$. Results indicate that chemicals outside the application domain cannot be predicted with sufficient accuracy (e.g., chemicals outside the grey area).

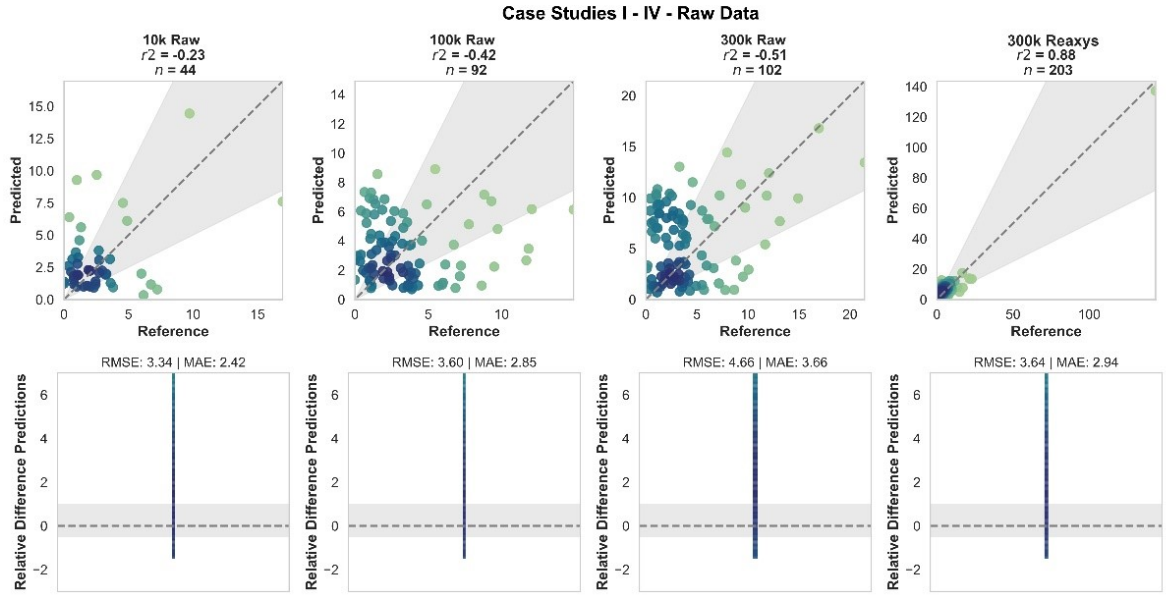

Figure S7: Raw data LOOV

In order to understand which chemicals can be predicted more accurately, a systematic approach was followed to exclude all chemicals in the validation set SK_{LOOV} outside the desired accuracy range. Figure S8 shows the results of systematically removing the chemicals, which were not predicted within the desired accuracy range. We conclude that foremost (small) inorganic molecules, heavy halogenated and elements are not suitable for the LOOV and are removed (Figure S6, Step 3).

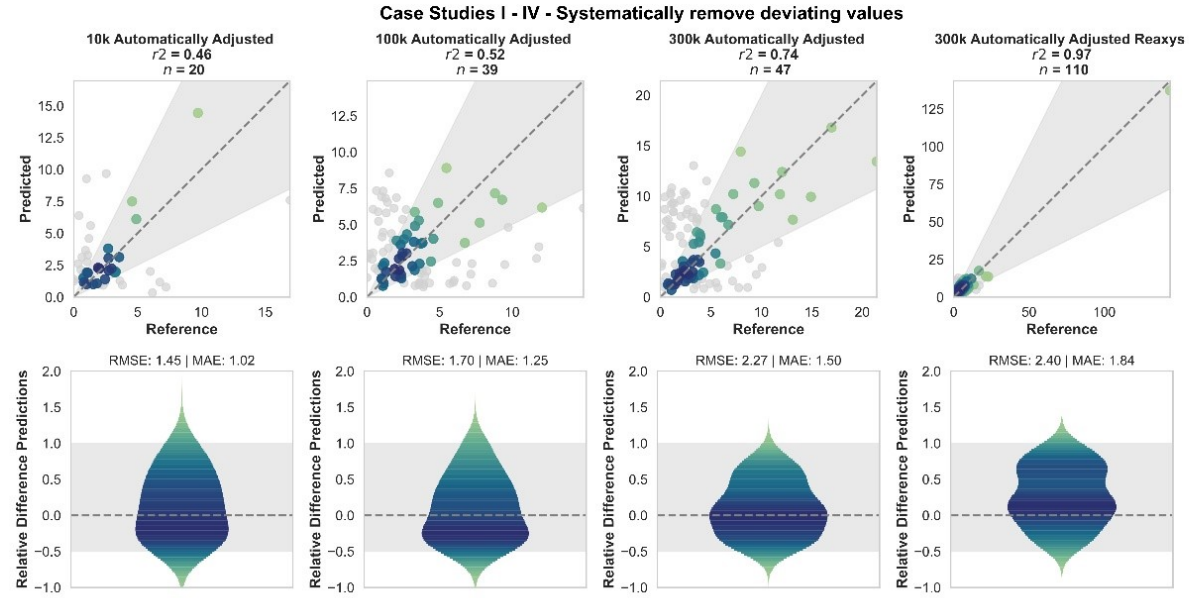

Figure S8: Results of the LOOV for all four case studies, CS I to CS IV. Top: raw data without adjusting, bottom systematically adjusting the data.

3.4. Filter Criteria for the LOOV

After following the filter criteria of Lucas et. al [4] (Figure S6, Step 1) and the subsequent removal of duplicates and outliers (Figure S6, Step 2), the following filter criteria (Figure S6, Step 3), based on the knowledge gained in Figure S7 and S8, were applied to obtain the validation set SK_{LOOV}:

- Only chemicals with the following characters/symbols in their SMILES notation:
    - Cl, Br, F, B, C, N, O, P, S, =, #, \, (, \,), 1, 2).
- No single atoms.
- No pure halogens.
- No PFAS, number of "F" atoms < 4 for every molecule.

Ultimately used chemicals for each case studies are given in a separate excel file.

3.5. Results of all Case Studies

Table S1 provides the results of the LOOV for all the four-case studies CS I – CS IV.

Table S1: Results of all case studies.

	Data	R^2	RMSE [kgCO2eq]	MAE [kgCO2eq]	n	Mean Relative Difference [%]	Mean Absolute Relative Error [%]
CS I	OS	0.28	3.78	2.53	12	43.7	72.3
CS II	OS	0.11	3.65	2.52	34	39.4	81.1
CS III	OS	0.41	2.79	1.98	41	29.7	59.3
CS IV	Reaxys©	-0.08	2.97	2.27	110	60.2	72.1

3.6. Prediction Error Analysis Case Study I – IV

Figure S9 provides the prediction error of chemicals in the LOOV of all case studies. As discussed, no strong bias towards over- nor underestimation is noticed for CS I – CS III. Case study IV shows a slight trend towards overestimation.

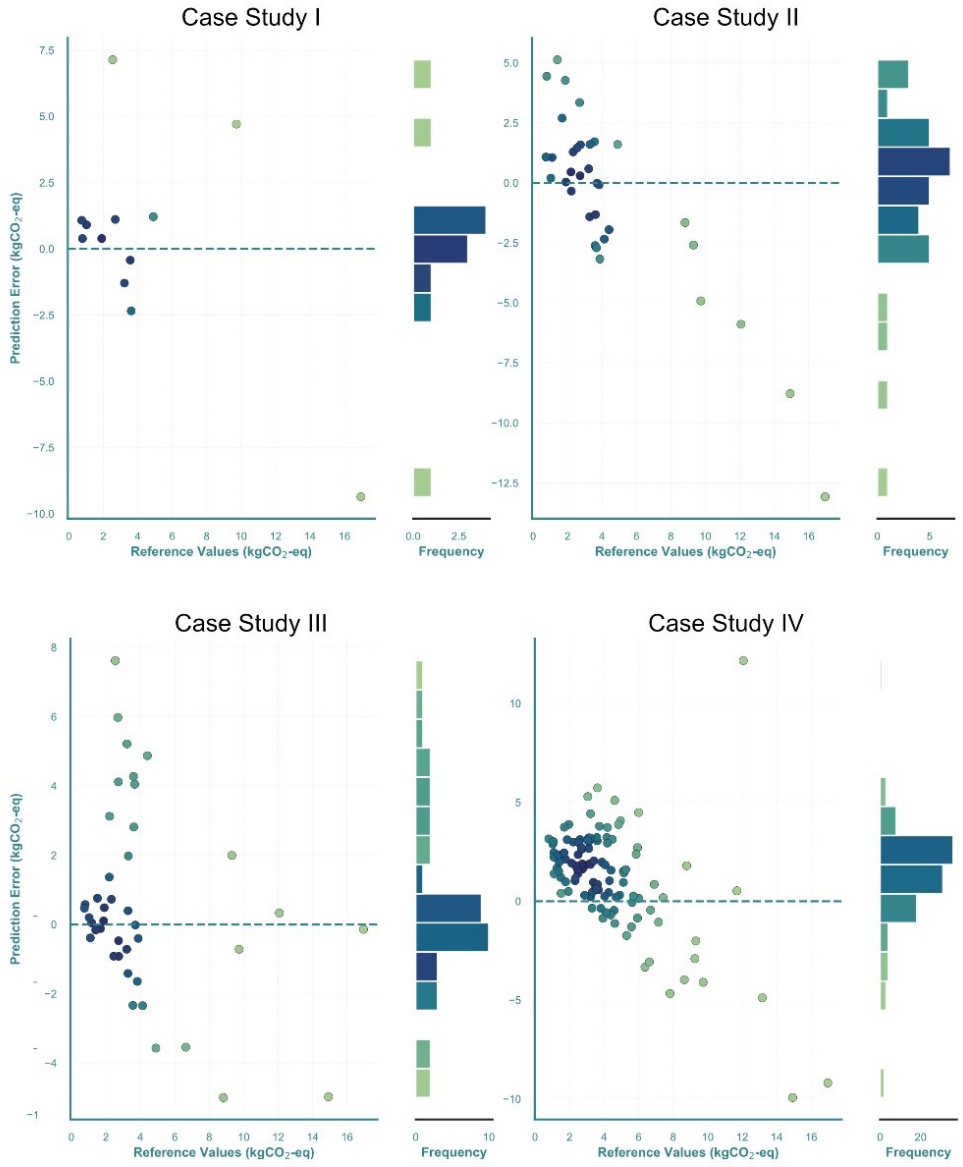

Figure S9: Prediction error analysis for all case studies I - IV.

3.7. Gate to Gate Analysis Case Study I – IV

Figure S10 provides the distribution of the impacts related to reaction energy, separation energy and the mass based allocation for all four case studies. Mass based allocation related impacts dominate in all four cases, while being more emphasized in CS III and CS IV.

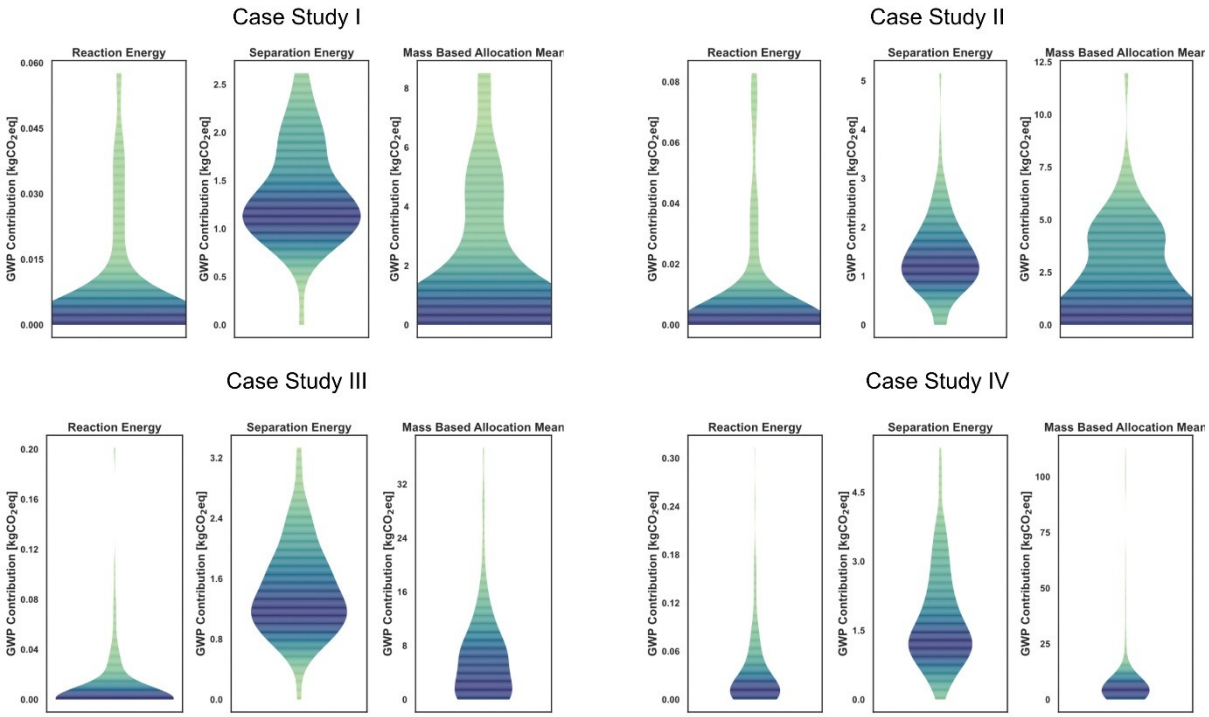

Figure S10: Gate-to-gate analysis for all case studies I - IV.

3.8. Cradle-to-Gate Prediction Accuracy Analysis

To validate our prediction performance, we replicated the pathways reported in ecoinvent for selected chemicals. The ecoinvent inventories were analyzed and translated into chemical reactions. Owing to the inventory structure, each chemical is associated with a single network with a single reaction. For the mass-based allocation, we adopted the same raw materials specified in the ecoinvent inventory and designated them as the "corpus" of the reaction network. The global warming potential values predicted by AuLCA were assessed relative to the corresponding ecoinvent values. Furthermore, energy-related impacts (e.g., steam for heat generation) were compared between our gate-to-gate predictions and the values reported in ecoinvent.

Overall, the total values (including reaction-, separation energy and mass-based allocation impacts) demonstrate good agreement between the predictions and the reference data.

The share of energy-related impacts (e.g., impacts related to reaction energy and separation energy) is comparatively low (on average, 18% for the reference data and 16% for the predictions) in respect to the total values.

The predicted energy impact values follow the same general trend as the ecoinvent reference values.

Table S2: Results of the cradle-to-gate GWP prediction accuracy analysis. Values have been normalized in respect to the reference values. Total number of case studies n = 13.

Chemical	Reference Values Normalized	AuLCA Prediction Normalized	Energy Share Reference	Energy Share AuLCA
Aniline	1	1.01	21%	11%
Azodicarbonamide	1	0.99	2%	14%
Hydrazine sulfate	1	0.99	7%	9%
Phenol	1	1.04	16%	14%
Propanal	1	0.69	22%	11%
2-pyridinol	1	0.94	1%	3%
Salicylic acid	1	0.92	7%	13%
p-nitrotoluene	1	0.81	33%	30%
Dimethylaminopropylamine	1	0.99	5%	11%
Triphenyl phosphate	1	1.04	5%	21%
Acetic acid	1	1.24	21%	15%
Trimethyl borate	1	1.04	25%	11%
Trichloroethylene	1	0.58	62%	48%

3.9. External Validation using the IDEA database

The English version of the IDEA database v. 3.3.3 [5] was utilized to perform an external validation of AuLCAs predictions. For this matter, the results of the prior performed LOOV were compared against the corresponding data in the IDEA database. Due to the different content of the IDEA and ecoinvent databases, not all prior used LOOV chemicals were found in the IDEA database. All chemicals used for the validation are provided in a separate excel file. In addition, it should be noted that the IDEA database does not provide Europe as a standalone region. Hence, only global regionalized data was utilized in this comparison.

Figures S11a-d illustrates the results of the external validation, while Table S3 shows the corresponding prediction metrics. Given the limited sample size for external validation, these metrics are highly sensitive to individual outliers and should be interpreted with this observation in mind.

Figure S11e compares the reference values from the ecoinvent and IDEA databases. While several chemicals have comparable GWP values across both sources, significant deviations are observed for others. The prediction performance for Case Studies I, III, and IV slightly worsens when validated against the IDEA database compared to ecoinvent. Conversely, the performance in CS II improves in the IDEA database.

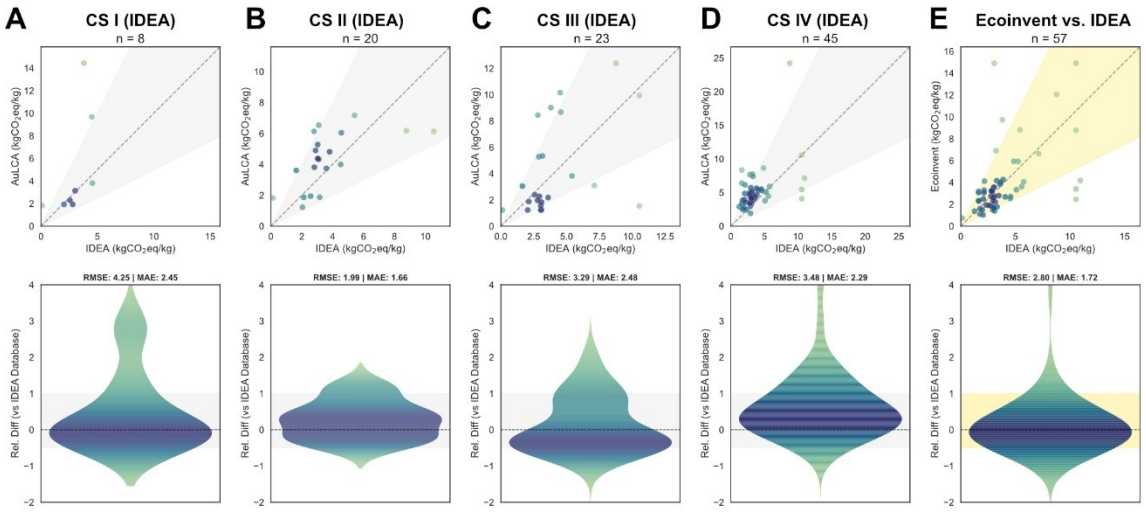

Figure S11: External validation of the LOOV results using the IDEA database. Values for RMSE and MAE in kgCO2-eq/kg.

Note that in all cases we rely on a reduced data set for validation compared to the amount of chemicals that can be predicted from the corpus. Furthermore, Figure S11e shows that the IDEA and ecoinvent data do not fully align. Because AuLCA is trained with the ecoinvent-based corpus, one would expect that the same type of discrepancies would also appear when comparing predictions made with AuLCA with the values in IDEA, for which our approach was not trained. Hence, when validating against an alternative source like the IDEA database, the reported error is to some extent expected to include both the algorithm's predictive deviation and the systematic differences between the ecoinvent and IDEA datasets.

Nevertheless, the majority of the GWP predictions for the LOOV chemicals are still within the grey highlighted desired accuracy range of +100%/-50%.

Table S3: Prediction performance metrics for the external validation against the IDEA database.

Networks	Open-Source (OS)			Reaxys©
Case Study	CS I	CS II	CS III	CS IV
RMSE [kgCO2eq/kg]	4.25	1.99	3.29	3.48
MAE [kgCO2eq/kg]	2.46	1.61	2.48	2.29
MRE [%]	247	121	101	72
R2	-8.7	0.25	-0.55	-0.94

4. References

(1) van der Lingen, R. Reaction SMILES CRD 1.37M Dataset, 2025. https://doi.org/10.6084/m9.figshare.28230053.v1.
(2) Diorazio, L. J.; Hose, D. R. J.; Adlington, N. K. Toward a More Holistic Framework for Solvent Selection. *Org. Process Res. Dev.* **2016**, *20* (4), 760–773. https://doi.org/10.1021/acs.oprd.6b00015.
(3) Elsevier. Reaxys, 2025. https://www.reaxys.com/.
(4) Lucas, E.; Martín, A. J.; Mitchell, S.; Nabera, A.; Santos, L. F.; Pérez-Ramírez, J.; Guillén-Gosálbez, G. The Need to Integrate Mass- and Energy-Based Metrics with Life Cycle Impacts for Sustainable Chemicals Manufacture. *Green Chem.* **2024**, *26* (17), 9300–9309. https://doi.org/10.1039/D4GC00394B.
(5) National Institute of Advanced Industrial Science and Technology (AIST); Research Institute of Science for Safety and Sustainability; Research Laboratory for IDEA. IDEA Ver.3.3.3 Regionalized Type JPN+GLO, 2024.