

MPA_Pathway_Tool: User-friendly, automatic assignment of microbial community data on metabolic pathways

Daniel Walke^{1,*}, Kay Schallert², Prasanna Ramesh³, Dirk Benndorf⁴, Emanuel Lange⁵, Udo Reichl⁶, Robert Heyer^{7,*}

*, corresponding author

1. Otto von Guericke University, Bioprocess Engineering, Universitätsplatz 2, 39106 Magdeburg, Germany, daniel.walke@ovgu.de, (*: [corresponding author](#))
2. Otto von Guericke University, Bioprocess Engineering, Universitätsplatz 2, 39106 Magdeburg, Germany, kay.schallert@ovgu.de
3. Otto von Guericke University, Database and Software Engineering Group, Universitätsplatz 2, 39106 Magdeburg, Germany, prasanna.ramesh@ovgu.de
4. Otto von Guericke University, Bioprocess Engineering, Universitätsplatz 2, 39106 Magdeburg, Germany
Anhalt University of Applied Sciences, Microbiology, Bernburger Straße 55, 06354 Köthen, Germany
Max Planck Institute for Dynamics of Complex Technical Systems, Bioprocess Engineering, Sandtorstraße 1, 39106 Magdeburg, benndorf@mpi-magdeburg.mpg.de
5. Otto von Guericke University, Bioprocess Engineering, Universitätsplatz 2, 39106 Magdeburg, Germany, emanuel.lange@ovgu.de
6. Max Planck Institute for Dynamics of Complex Technical Systems, Bioprocess Engineering, Sandtorstraße 1, 39106 Magdeburg, benndorf@mpi-magdeburg.mpg.de, ureichl@mpi-magdeburg.mpg.de
7. Otto von Guericke University, Bioprocess Engineering/ Database and Software Engineering Group, Universitätsplatz 2, 39106 Magdeburg, Germany, heyer@mpi-magdeburg.mpg.de, (*: [corresponding author](#))

Highlights

user-friendly generation of pathways, re-using of existent metabolic pathways, automated mapping of data

index of abbreviations

CSV	comma-separated values	JSON	JavaScript Object Notation
SBML	Systems Biology Markup Language	SVG	Scalable Vector Graphics
MPA	MetaProteomeAnalyzer	CSS	Cascading Style Sheets
HTML	Hypertext Markup Language	JS	JavaScript
EC-number	Enzyme Commission number	BGP	biogas plant
KEGG	Kyoto Encyclopedia of Genes and Genomes		

Abstract

Motivation: Taxonomic and functional characterization of microbial communities from diverse environments such as the human gut or biogas plants by multi-omics methods plays an ever more important role. Researchers assign all identified genes, transcripts, or proteins to biological pathways to better understand the function of single species and microbial communities. However, due to the versatility of microbial metabolism and a still increasing number of new biological pathways, linkage to standard pathway maps such as the KEGG (Kyoto Encyclopedia of Genes and Genomes) central carbon metabolism is often problematic.

Results: We successfully implemented and validated a new user-friendly, stand-alone web application, the MPA_Pathway_Tool. It consists of two parts, called 'Pathway-Creator' and 'Pathway-Calculator'. The 'Pathway-Creator' enables an easy setup of user-defined pathways with specific taxonomic constraints. The 'Pathway-Calculator' automatically maps microbial community data from multiple measurements on selected pathways and visualizes the results.

Availability and Implementation: The MPA_Pathway_Tool is implemented in Java and ReactJS. It is freely available on <http://mpa-pathwaymapper.ovgu.de/>. Further documentation and the complete source code are available on GitHub (https://github.com/danielwalke/MPA_Pathway_Tool).

Contact: daniel.walke@ovgu.de, heyer@mpi-magdeburg.mpg.de

Supplementary Information: Additional files and images are available at *MDPI* online.

Motivation

In the last years, studying the taxonomic and functional composition of bacterial, viral, and archaeal species in the human gut [1–5], in biotechnology [6–8], or the environment [9,10] became increasingly important [8]. There are several different approaches for analyzing microbial communities, focussing on the entirety of the genes (metagenomics), transcripts (metatranscriptomics), or proteins (metaproteomics). Metagenomics only reveals the presence of genes, whereas metatranscriptomics and metaproteomics indicate actual protein expression [8]. Based on the protein expression levels, a microbial communities' phenotype can be linked with specific environmental conditions, process parameters, or diseases [11]. Due to the complexity and amount of the multi-omics data, comprehensive bioinformatic workflows were developed for the data evaluation [12–15]. For example, the MetaProteomeAnalyzer (MPA) enables analyzing and interpreting metaproteomics data sets. It offers a free, open-source, end-user-oriented complete pipeline from peak lists to taxonomic and functional result evaluation. Among others, the MPA links identified proteins to functional categories (e.g., biological keywords) and the KEGG pathways [16]. In addition to the KEGG pathway system [17], several other pathway collection and mapping tools such as Reactome [18], Pathview Web [19], Escher [20], Pathway Tools [21] exist supporting the data analysis of omics-datasets. However, due to the microbial metabolism's versatility and constantly newly discovered biological pathways [22], linkage to standard pathway maps is insufficient for many microbial community studies. Therefore new tools are required, tailored for microbial community studies.

In general, a good pathway tool needs to meet at least the following six criteria. (i.) It should provide easy and intuitive creation of pathways to enable the fast generation of multiple pathways. (ii.) Since new reactions are discovered and pathways might be updated, the tool should support modifying the pathway maps, i.e., appending new and deleting existent reactions. (iii.) Already created pathways from different databases should be reusable. Consequently, the pathway tool should provide an import function for standard exchange formats, like comma-separated values (CSV), JavaScript Object Notations (JSON), and Systems Biology Markup Language (SBML) formats. (iv.) The pathway tool should also map experimental data on created pathways and highlight differences between the considered samples. (v.) Since metabolic reactions are taxonomy-specific, the pathway tool needs a filter to distinguish between reactions carried out by a specific taxonomy. One example

of this specificity is hydrogenotrophic methanogenesis and the Wood-Ljungdahl pathway. Both pathways share similar enzymes (i.e., similar EC-numbers (Enzyme Commission numbers)). However, hydrogenotrophic methanogenesis is carried out only by archaea [23], while the Wood-Ljungdahl pathway is carried out mainly by bacteria [8]. (vi.) The tool should be independent of operating systems so that nearly everyone can use the tool, favoring an implementation as a web application.

The paper aims to present the new web application called MPA_Pathway_Tool. It enables easy creation of user-defined pathways with specific taxonomic constraints, automatically mapping microbial community data from multiple measurements on selected pathways, and visualizing the results ('Pathway-Creator'). Additionally, the 'Pathway-Calculator' enables mapping an entire omics data set on multiple pathways to support the automated data analyses. The functionality of the MPA_Pathway_Tool is demonstrated and validated by reproducing the manual assignment of proteins to metabolic pathways from a previous metaproteomics study about biogas plants, focussing on three pathways (hydrogenotrophic methanogenesis, acetoclastic methanogenesis, and Wood-Ljungdahl pathway).

Results

The 'Pathway-Creator' enables users to define their own metabolic pathways

The first part of the MPA_Pathway_Tool is the 'Pathway-Creator' (figure 1). It enables the creation of user-defined pathways by adding reactions iteratively and linking omics data to this specific pathway. The left side of the 'Pathway-Creator' contains a list of buttons for uploading experimental data and pathways (as CSV, JSON, and SBML), adding new reactions from KEGG, adding user-defined reactions, importing multiple reactions, and downloading created pathways (as CSV, SBML, JSON, and Scalable Vector Graphics (SVG)) and mapped data (as CSV). The right side contains a graph for visualizing the created pathway. Circular-shaped nodes are metabolites (in KEGG referred to as compounds), and diamond-shaped nodes are reactions. Nodes are connected by edges, which display the direction of a metabolic reaction. After uploading experimental data, a further user interface emerges, showing the mapping of the data set to the pathway. After sample selection, by clicking on the respective button at the bottom of the tool, reaction nodes are colored dependent on their abundance in the sample and the color settings. Information about abundances in all samples for a specific reaction is available by clicking on the respective reaction- node.

The 'Pathway-Calculator' enables automated mapping of experimental data on multiple metabolic pathways

The 'Pathway-Calculator' (figure 2) consists of two upload zones, one for experimental data and another for multiple pathway files (as CSV, JSON, or SBML). The 'Pathway-Calculator' performs mapping of experimental data on multiple uploaded pathways. The mapping algorithm is explicitly explained in methods. After uploading all previously created pathways (CSV) and the experimental data in the 'Pathway-Calculator', the calculation starts. Subsequently, the result table can be exported as CSV. Furthermore, a list with all unmatched features (e.g., proteins) can be downloaded.

Finally, we tested the performance of the 'Pathway-Calculator' by uploading experimental data with different file sizes (10,000 proteins, 100,000 proteins, and 1,000,000 proteins with 44 samples per test) and different a number of pathways on one of our local desktop computers (AMD Ryzen 5 3600, 16 GB DDR4 RAM 3000MHz, Chrome Browser version 89.0.4389.90). We utilized 1, 10, and 100 copies of the user-defined Wood-Ljungdahl pathway used in Heyer et al. 2019b (Biogas plant (BGP)-Wood-Ljungdahl pathway) for the performance test. Finally, we measured the elapsed time between pressing the 'start'-button and the availability of the download link (table 1). Each test with 10,000 and 100,000 proteins performed under 1 minute indicating a good performance for most files. Tests with 1,000,000 proteins took much longer (up to 12 minutes) caused by higher upload times and high requirements on memory and on CPU performance (table 1).

Tool validation with experimental data

Validation of the MPA_Pathway_Tool was carried out by comparing our pathway assignment against a manual pathway assignment of a previous metaproteomics study about biogas plants [8]. Additionally, we assigned the proteins to the standard KEGG modules to illuminate the demand for user-specific pathways.

For the evaluation, we created six pathway maps (BGP-hydrogenotrophic-methanogenesis, BGP-acetoclastic-methanogenesis, BGP-Wood-Ljungdahl-pathway, KEGG-hydrogenotrophic-methanogenesis, KEGG-acetoclastic-methanogenesis, and KEGG-Wood-Ljungdahl-pathway). The three user-defined pathways (BGP-, Supplementary information figure 3-5) were created by importing EC- numbers under 'import multiple reactions' and selecting the appropriate reactions (for further details, see the tutorial in supplementary information file 17) based on the pathway assignment in [8]. The creation of each of the other three KEGG pathways (Supplementary information figure 6-8) was performed by importing the respective KEGG-MODULE using 'import KEGG MODULE' under 'import multiple reactions'. For each pathway, a taxonomic classification was added (table 2) by using 'Add Taxonomy' under 'node configurations' and adding the respective taxonomy. According to the publication of the dataset, we decided to exclude Archaea from the Wood-Ljungdahl pathway

and include only Archaea for the hydrogenotrophic and acetoclastic methanogenesis [8]. Subsequently, the experimental metaproteomics data (Supplementary information file 9) were mapped on each pathway using single pathway (Supplementary information file 10-15) and multiple pathway mapping (figure 4)(Supplementary information file 1). Unmatched proteins were downloaded as CSV (Supplementary information file 2).

The summed spectral counts mapped on each user-defined pathway were slightly greater than in the published experimental data for each sample. The observed differences had three causes. First, some reactions have multiple EC-numbers. In the published experimental data, only one EC-number was considered for each reaction. The second reason for the differences is the missing matching of proteins by K-numbers, which also catalyze the same reaction. They were not considered in the published data. The last reason is the usage of deprecated EC-numbers in the published data. EC- numbers are transferred to new EC- numbers occasionally. All three reasons lead to a minor loss of information in the published dataset due to less matched proteins. The stated issues were corrected in the MPA_Pathway_Tool, resulting in a higher accuracy during the pathway assignment.

Furthermore, we compared the results of each BGP-pathway with the corresponding KEGG-pathway. We observed higher spectral counts for the created KEGG MODULEs for most analyzed pathways compared to the respective BGP-pathways (figure 3). Especially acetoclastic methanogenesis exhibits higher spectral counts in most samples for the KEGG MODULE. One reason for higher spectral counts in KEGG MODULEs is the integration of more reactions in the KEGG MODULEs than in the BGP- pathways (figure 4). Whereas KEGG modules contain all reactions for a certain function, BGP pathways focus on the pathway specific for the function. For example, KEGG-acetoclastic-methanogenesis, and KEGG-hydrogenotrophic methanogenesis possess three additional reactions (2.1.1.86, 2.8.4.1., and 1.8.98.1), which catalyze the last steps of the methanogenesis. These were purposely integrated into another separate BGP-pathway, because of the occurrence in hydrogenotrophic and acetoclastic methanogenesis. This separation prevented repeated summation of spectral counts of equal reactions. The analyzed KEGG MODULEs additionally possess similar reactions with minor differences (e.g., NADH+H⁺ or NADPH+H⁺ as reductant).

Conclusion: The MPA_Pathway_Tool provides an easy and fast option to set up multiple pathways

We successfully implemented the pathway tool to meet all of our six defined criteria. (i.) The MPA_Pathway_Tool provides an easy and fast setup of multiple pathways. Multiple reactions can be imported using various options, e.g., import by EC numbers, import of a KEGG MODULE, or entire SBML files. (ii.) A further modification of the generated pathways is possible by deleting reactions and adding new reactions (from the KEGG database and user-defined reactions). (iii.) As interchange formats, JSON, CSV, and SBML were implemented. (iv.) Experimental data from metaproteomics, metatranscriptomics, and metagenomics studies can be automatically mapped on single pathways ('Pathway-Creator') and multiple pathways ('Pathway-Calculator'). The results of pathway mapping can also be exported as CSV. (v.) The mapping algorithm includes a taxonomic filter, which was successfully applied by comparing our results with published experimental data for the stated pathways. (vi.) The MPA_Pathway_Tool was implemented as a stand-alone web application to guarantee the independence from users' operating systems.

Future work: integration to other features and addition of new functions will increase the flexibility of the MPA_Pathway_Tool

The MPA_Pathway_Tool is developed as a stand-alone application for the characterization and analysis of microbial community data. To broaden its scope of application, it will be integrated into the next version of the MPA software. However, in the future other tools such as UniPept [24] or Propane [25] might be linked, too.

Since the MPA_Pathway_Tool already provides stoichiometric data, further options like identification-driven flux balance analysis and flux vector analysis might be added to the workflow using tools such as the CellNetAnalyzer [26], COBRApy [27] or Escher [20]. As it allows the calculation of metabolic flows in metabolic networks, it can be used to predict the specific growth rate of organisms and the specific production rate of

182 certain metabolites. Accordingly, identification of strategies for optimizing biotechnological processes might be
183 feasible [28].

Availability and Implementation

General workflow

The MPA_Pathway_Tool represents an intuitive web application for mapping (meta)-proteome and other omics data on metabolic pathways and the creation and modification of pathways. It consists of two different parts. First, the 'Pathway-Creator' allows the creation and modification of pathways. The 'Pathway-Creator' also supports single pathway mapping. The second part is the 'Pathway-Calculator', which provides multiple pathway mapping.

Implementation

The MPA_Pathway_Tool is a stand-alone web application consisting of JavaScript (JS), Hypertext Markup Language 5 (HTML), and Cascading Style Sheets (CSS) on the client-side and Java on the server-side. The client-side is built with the help of ReactJS. We used Create React App for the setup of the project. Create React App is used to create single-page React applications, and it provides a setup without configurations [29]. Third-party dependencies have been pulled by using Node Package Manager (npm) (version 6.14.9). The most important dependencies are ReactJS (version 16.8.6) for using JSX syntax in our project, Redux (version 4.0.5) [30] and mobx (version 7.0.5) [31]- used for storing and handling states, React-d3-graph (version 2.5.0)- used for visualization of pathways [32], Axios (version 0.21.0) [33]- used for making HTTP requests from the browser, Material-UI (version 4.11.2) [34] used for the implementation of some user-interface components, Lodash (version 4.17.21) [35] for deep cloning objects, and File-saver (version 2.0.5) [36] used for saving files. A complete list of used dependencies on the client-side is listed in the package.json file (https://github.com/danielwalke/MPA_Pathway_Tool/tree/main/keggcalculator-frontend).

We have implemented a REST-API (figure 7) and an algorithm for mapping data on multiple pathways in the programming language Java (Java SE-1.8). Dependencies have been imported using Maven. We used Gson (version 2.8.5) [37] for converting Java objects to JSON and vice versa, Spark (version 2.6.0) for setting up the REST-API, and JSBML (version 1.5) [38] for reading SBML- files.

MPA_Pathway_Tool: 'Pathway-Creator'

In the 'Pathway-Creator' compounds, i.e., substrates and products (in KEGG referred to as compounds), are visualized as circular nodes and reactions as diamond-shaped nodes. Links connect compounds and reactions, and arrows show the direction of a reaction. Users have the opportunity to create their pathway from scratch by defining an initial substrate and then adding a reaction with the desired product iteratively. For each added reaction, the additional substrates and products are added automatically. The reaction library contains all reactions from the KEGG database and may extend by user-defined reactions. Multiple reactions can also be imported by using the KEGG-MODULE import, EC- number reactions import, K- number reactions import, and SBML- files. Created pathways can be exported as CSV, SBML, and JSON. In addition, to the metabolic reactions, we store in these files the opacity of nodes, their position, and the abbreviations for later usage or modifications of the pathways. This information is stored in a specific tag ('layout:ListOfLayouts') of SBML- files (figure 5). Users can also download the created pathway as an image (SVG). Figure 6A provides an overview of the 'Pathway-Creator'. For a detailed description of pathway creation, please refer to the tutorial in supplementary file 17.

Identified features (e.g., proteins) can be mapped on the created pathway (figure 6A). Colors visualize differences in abundances for a chosen sample for each reaction (single pathway mapping). The results can be downloaded as CSV.

MPA_Pathway_Tool: 'Pathway-Calculator'

Multiple created pathways (CSV, JSON, or SBML) and a file with experimental data (e.g., protein identifications) as CSV can be uploaded on the 'Pathway-Calculator' to perform multiple pathways mapping (figure 6B). After calculation, the result file can be downloaded as CSV, which contains in each row a pathway name, the total number of reactions in this pathway, the number of reactions identified in the data, and the summed quantified values for each sample.

Mapping Algorithm

For a better understanding of the mapping algorithm, we explain the application with experimental data containing proteins. Of course, the mapping algorithm can be applied for other omics- data too. Each pathway contains a pathway name (filename) and multiple reactions. On the other hand, each reaction contains an ID, a name, substrates, and products with stoichiometric coefficients, EC- numbers, K- numbers, taxonomic constraints, and a list of matching proteins, which is empty at the beginning of the mapping calculation. For a better overview, only the last four properties are visualized as a scheme (figure 8). The imported file with protein identifications contains multiple proteins, which comprise an identifier, EC- numbers, K- numbers, a taxonomic tree (superkingdom, kingdom, phylum, class, order, family, genus, species), a description, and quantified values for each measured sample (Supplementary File 16, example input). For each reaction in a pathway, all proteins are analyzed. To add a protein to the list of matched proteins, two conditions must be fulfilled. The first condition is fulfilled if the EC-numbers or K- numbers of the reaction contain at least one EC-number or K- number of the protein. The second condition is fulfilled if the taxonomic tree of the protein contains the taxonomic requirement specified for the reaction. The fulfilment of both conditions adds the protein to the list of matched proteins. The quantitative values of all proteins are summed up for each sample. For single pathway mapping, the results can be visualized on the created pathway and exported as CSV. For multiple pathway mapping, all quantified values of each reaction (sums of quantified values of matched proteins) are summed up. Additionally, each pathway's completeness is checked by returning the number of reactions identified in the data and the total number of reactions in the pathway. Finally, results can be exported as CSV. Multiple pathway mapping also provides the calculation of multiple imported pathways.

Experimental data

Experimental data used for validating the MPA_Pathway_Tool originate from a previous metaproteomics study of 11 different biogas plants sampled at two different time points [8]. As input, we used the final result matrix of all identified metaproteins (see Supplementary information file 5, in [8])

Availability

The MPA_Pathway_Tool is freely available on the web at <http://mpa-pathwaymapper.ovgu.de/>. Further documentation and the complete source code are deposited on GitHub (https://github.com/danielwalke/MPA_Pathway_Tool).

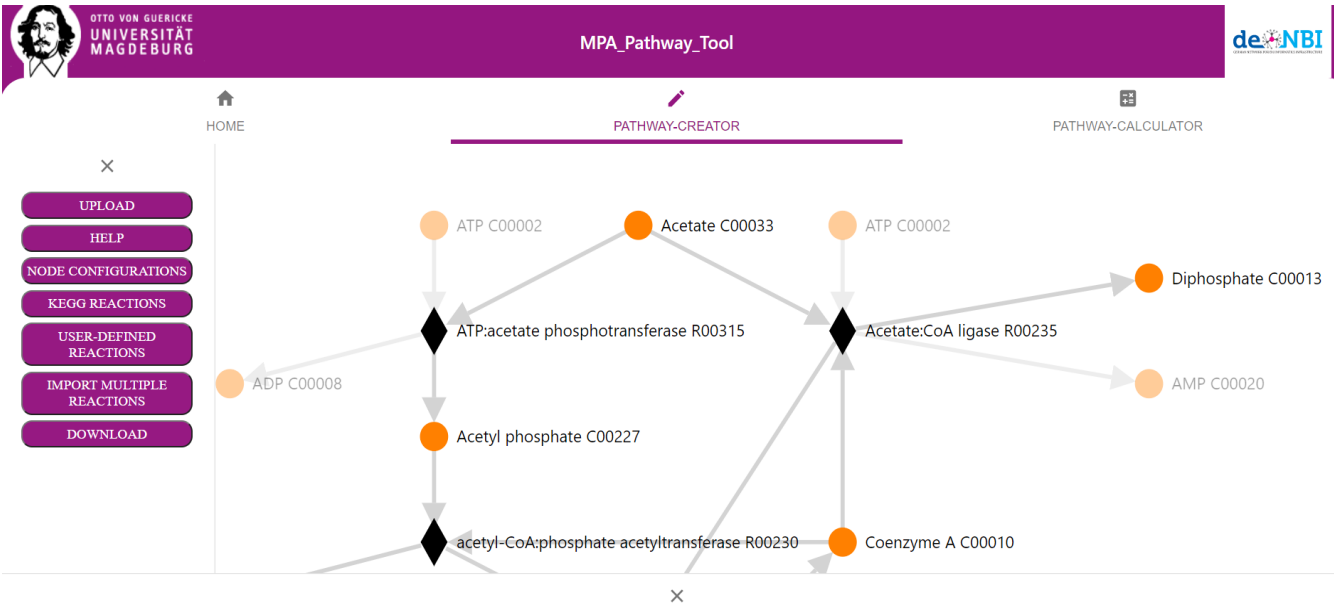
Funding and acknowledgment

This work was supported by the German Federal Ministry of Education and Research (de.NBI network. project MetaProtServ. grant no. 031L0103). We highly appreciate their funding.

Competing interests

The authors declare that they have no competing interests.

268 **Figures and tables**



BGP_02_T1_Ext1

269 **Figure 1: Screenshots of the 'Pathway-Creator' of the MPA_Pathway_Tool.** The 'Pathway-Creator' enables
270 the creation of user-defined pathways by adding reactions iteratively. A user- interface (left side) allows users
271 to upload experimental data and pathways, add new reactions, modify the pathway's visualization, and
272 download the created pathways and the mapped data. A network graph (right side) visualizes the created
273 pathway. Metabolites are visualized as circular nodes, while reactions are visualized as diamond-shaped nodes.
274 The connecting edges show the direction of each reaction. On the bottom are buttons for selecting samples of
275 the experimental data and a heatmap for adjusting the colors. Users can choose a sample, and as a
276 consequence, the reaction nodes are colored according to their abundance.
277

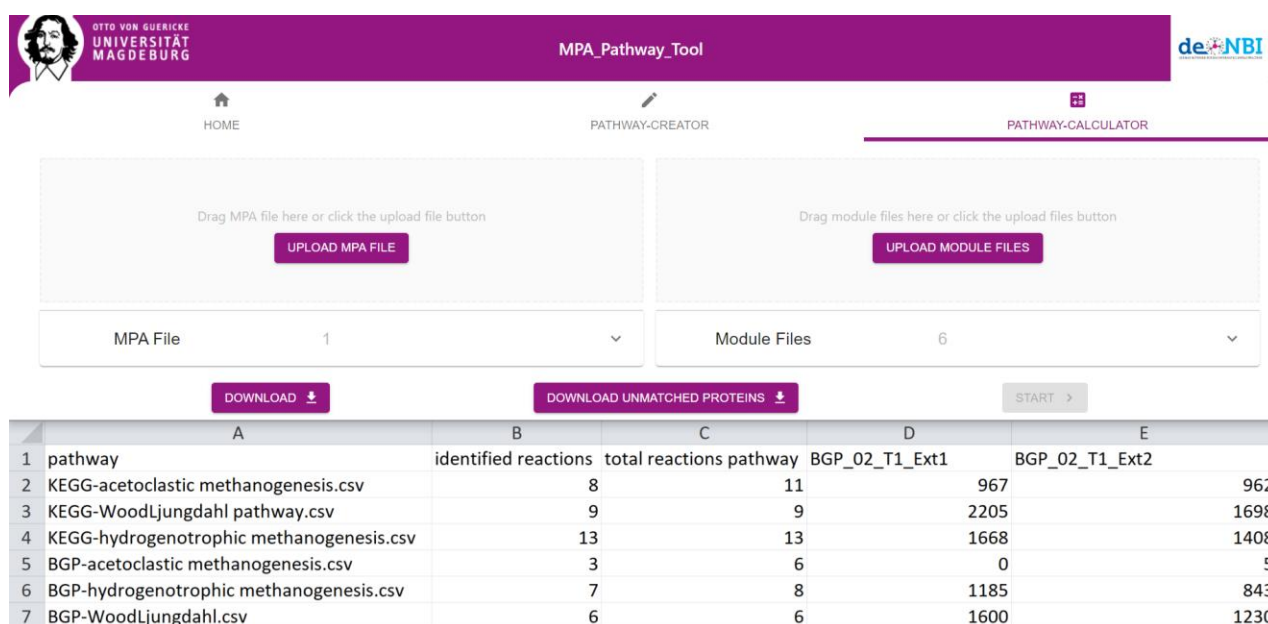


Figure 2: Screenshot of the 'Pathway-Calculator' of the MPA_Pathway_Tool. Users can upload experimental data as CSV on the left and multiple created pathways (CSV, SBML, and JSON) on the right side. After starting by clicking 'start', the 'Pathway-Calculator' matches the features (e.g., proteins) on all pathways and returns a file with the results (for further details, refer to the mapping algorithm in the material and methods part). Additionally, users can download a file of all unmatched features (e.g., proteins).

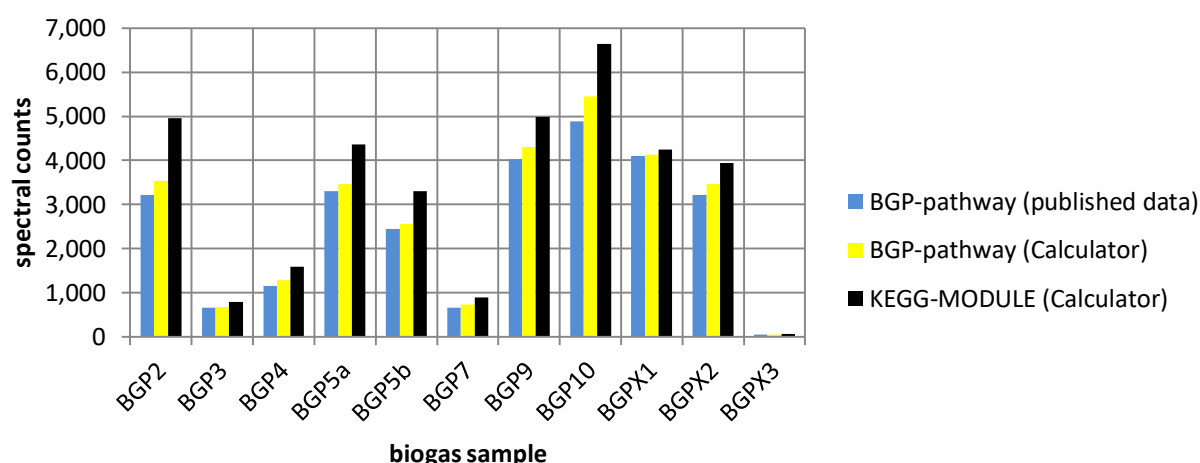
Table 1: Performance test for the 'Pathway-Calculator'. It lists the elapsed time in seconds between pressing 'start' and receiving the download link (AMD Ryzen 5 3600, 16 GB DDR4 RAM 3000MHz, Chrome version 89.0.4389.90) Each test was performed with 44 samples and different numbers of proteins.

Experimental data	1 pathway	10 pathways	100 pathways
10,000 proteins	5 s	5 s	6 s
100,000 proteins	5 s	10 s	36 s
1,000,000 proteins	192 s	244 s	712 s

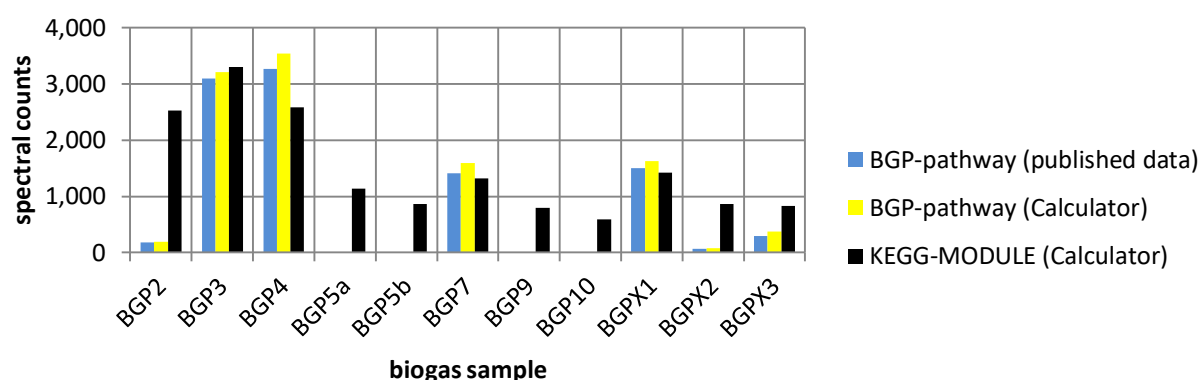
Table 2: Taxonomic classification of each pathway. Taxonomic assignments for the BGP-pathways were carried out according to the manual assignments in the used data set for the validation.

Pathway	Added taxonomic requirement
BGP-Hydrogenotrophic methanogenesis (user-defined)	only Archaea
KEGG-Hydrogenotrophic methanogenesis (KEGG-MODULE)	only Archaea
BGP-Acetoclastic Methanogenesis (user-defined)	only Archaea
KEGG-Acetoclastic Methanogenesis (KEGG-MODULE)	only Archaea
BGP-Wood- Ljungdahl pathway (user-defined)	all except Archaea
KEGG-Wood- Ljungdahl pathway (KEGG-MODULE)	all except Archaea

Wood-Ljungdahl pathway



Acetoclastic methanogenesis



Hydrogenotrophic methanogenesis

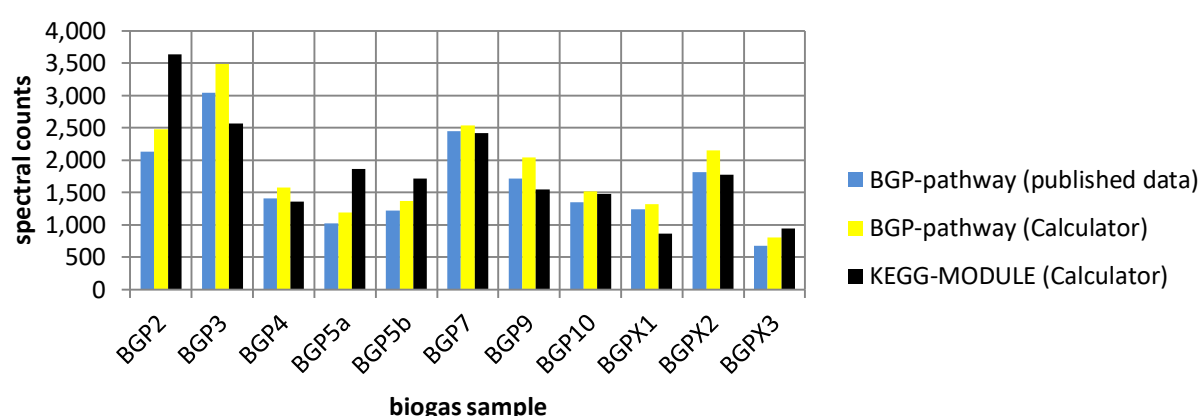


Figure 3: Comparison of experimental data mapped on each pathway (Wood-Ljungdahl pathway, acetoclastic methanogenesis, and hydrogenotrophic methanogenesis) with published data [8]. Summed spectral counts plotted against all biogas samples are blue and yellow for user-defined pathways and black for KEGG-MODULES. The results were obtained from the 'Pathway-Calculator' (yellow and black) and published data (blue). Especially acetoclastic methanogenesis shows higher spectral counts in most samples, indicating the occurrence of unspecific reactions.

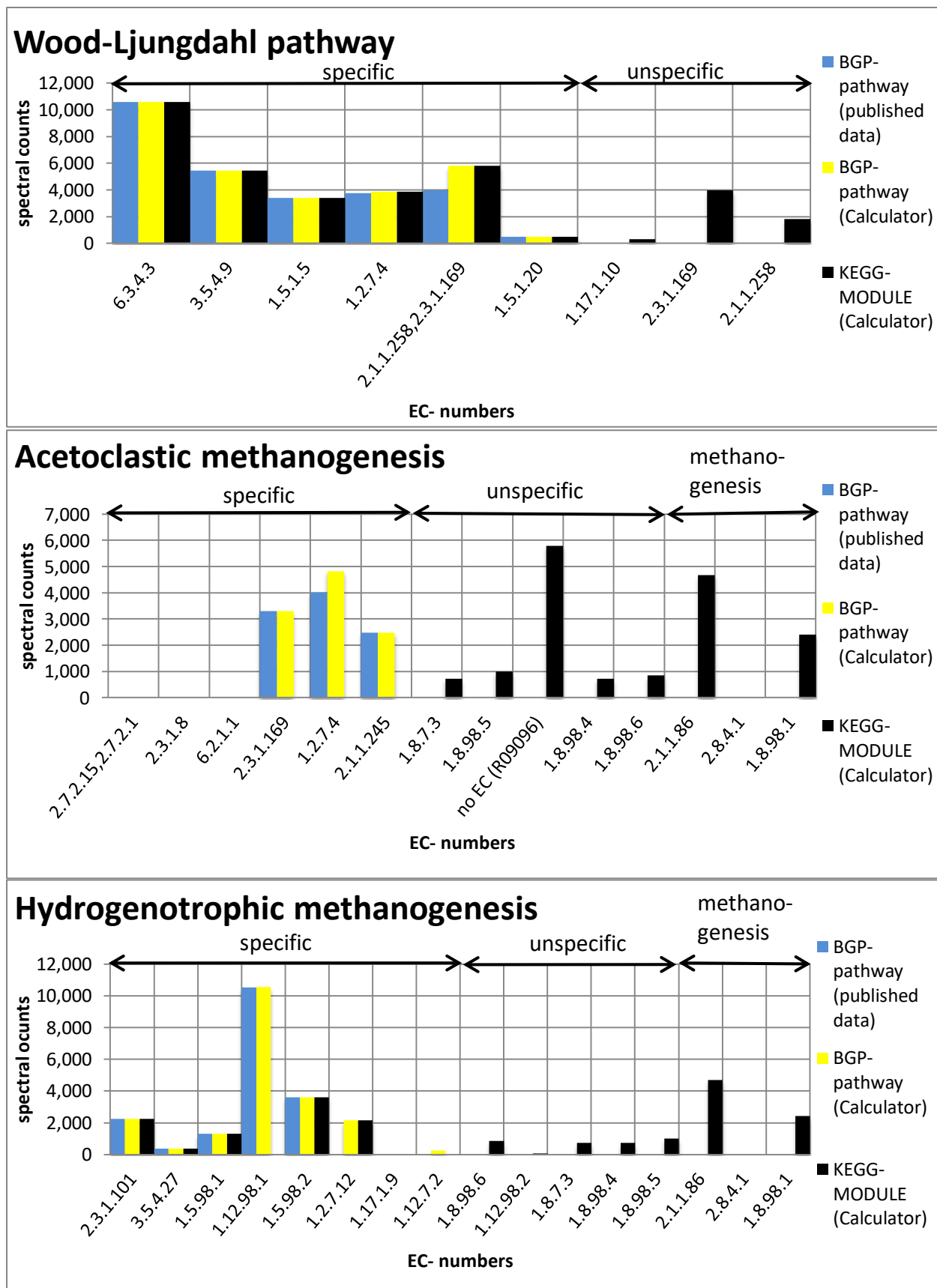


Figure 4: Summed spectral counts of each reaction in the Wood-Ljungdahl pathway, acetoclastic methanogenesis, and hydrogenotrophic methanogenesis. Results from published data [8] were visualized in blue, results from the 'Pathway-Calculator' for the user-defined in yellow, and results from the 'Pathway-Calculator' for KEGG-MODULEs in black. All KEGG-MODULEs possess additional unspecific reactions. KEGG-

308 acetoclastic-methanogenesis and KEGG-hydrogenotrophic methanogenesis possess three additional reactions
 309 (2.1.1.86, 2.8.4.1., and 1.8.98.1), which catalyze the last steps of the methanogenesis. (6.3.4.3:
 310 Formate:tetrahydrofolate ligase (ADP-forming); 3.5.4.9: 5,10-Methenyltetrahydrofolate 5-hydrolase; 1.5.1.5:
 311 5,10-methylenetetrahydrofolate:NADP+ oxidoreductase; 1.2.7.4: carbon-monoxide,water:ferredoxin
 312 oxidoreductase; 2.1.1.258,2.3.1.169: Tetrahydrofolate + Acetyl-CoA <=> 5-Methyltetrahydrofolate + CoA + CO;
 313 1.5.1.20: 5-methyltetrahydrofolate:NADP+ oxidoreductase; 1.17.1.10: formate:NADP+ oxidoreductase;
 314 2.3.1.169: acetyl-CoA:corrinoid protein O-acetyltransferase; 2.1.1.258: 5-
 315 methyltetrahydrofolate:corrinoid/iron-sulfur protein methyltransferase; 2.3.1.101:
 316 Formylmethanofuran:5,6,7,8-tetrahydromethanopterin 5-formyltransferase; 3.5.4.27: 5,10-
 317 Methenyltetrahydromethanopterin 10-hydrolase; 1.5.98.1: 5,10-
 318 methylenetetrahydromethanopterin:coenzyme-F420 oxidoreductase; 1.12.98.1: Hydrogen:Coenzyme F420
 319 oxidoreductase; 1.5.98.2: 5,10-Methylenetetrahydromethanopterin:coenzyme-F420 oxidoreductase; 1.2.7.12:
 320 formylmethanofuran:ferredoxin oxidoreductase; 1.17.1.9: formate:NAD+oxidoreductase; 1.12.7.2:
 321 hydrogen:ferredoxin oxidoreductase; 1.8.98.6: coenzyme B,coenzyme M,ferredoxin:formate oxidoreductase;
 322 1.12.98.2: hydrogen:N5,N10-methenyltetrahydromethanopterin oxidoreductase; 2.1.1.86: 5-methyl-5,6,7,8-
 323 tetrahydromethanopterin:2-mercaptoethanesulfonate 2-methyltransferase; 1.8.7.3: CoB,CoM:ferredoxin
 324 oxidoreductase; 1.8.98.4: CoB,CoM,ferredoxin:coenzyme F420 oxidoreductase; 2.8.4.1: 2-
 325 (methylthio)ethanesulfonate:N-(7-thioheptanoyl)-3-O-phosphothreonine S-(2-sulfoethyl)thiotransferase;
 326 1.8.98.5: CoB,CoM,ferredoxin:H2 oxidoreductase; 1.8.98.1: Coenzyme B:coenzyme M:methanophenazine
 327 oxidoreductase; 2.7.2.15,2.7.2.1: ATP:acetate phosphotransferase; 2.3.1.8: acetyl-CoA:phosphate
 328 acetyltransferase; 6.2.1.1: Acetate:CoA ligase; 2.1.1.245: 5-methyltetrahydrosarcinapterin:corrinoid/iron-sulfur
 329 protein methyltransferase, R09096: acetyl-CoA decarboxylase/synthase)

```

<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level3/version2/core" xmlns:layout="http://www.sbml.org/sbml/level3/version1/layout/version1" layout:required="false" level="3" version="2">
  <model id="pathway">
    <listOfSpecies>
    </listOfSpecies>
    <listOfReactions>
    </listOfReactions>
    <listOfCompartments>
    </listOfCompartments>
    <layout:listOfLayouts xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:layout="http://www.sbml.org/sbml/level3/version1/layout/version1">
      <layout:layout layout:id="Layout1">
        <layout:dimensions layout:width="400" layout:height="230" />
        <layout:listOfReactionGlyphs>
          <reactionGlyph layout:id="glyph_R03115" layout:reaction="R03115">
            <layout:curve>
              <listOfCurveSegments>
                <layout:curveSegment xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:type="LineSegment">
                  <layout:start layout:x="611.3006056839773" layout:y="161.4198572567345" />
                  <layout:end layout:x="611.3006056839773" layout:y="161.4198572567345" />
                </layout:curveSegment>
              </listOfCurveSegments>
            </layout:curve>
            <layout:speciesReferenceGlyphs>
              <layout:speciesReferenceGlyph layout:id="SpeciesReferenceGlyph_alpha-D-Glucose" layout:speciesReference="SpeciesReference_alpha-D-Glucose" layout:speciesGlyph="SpeciesGlyph_alpha-D-Glucose" layout:role="substrate">
                <layout:curve>
                  <listOfCurveSegments>
                    <layout:curveSegment xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:type="LineSegment">
                      <layout:start layout:x="610.906460775097" layout:y="85.33197252959872" />
                      <layout:end layout:x="611.3006056839773" layout:y="161.4198572567345" />
                    </layout:curveSegment>
                  </listOfCurveSegments>
                </layout:curve>
              </layout:speciesReferenceGlyph>
            </layout:speciesReferenceGlyphs>
          </reactionGlyph>
        </layout:listOfReactionGlyphs>
      </layout:layout>
    </layout:listOfLayouts>
  </model>

```

Figure 5: An example of updates in an SBML-file. In the tag 'layout:listOfLayouts' all positions for the nodes of the created pathway are stored.

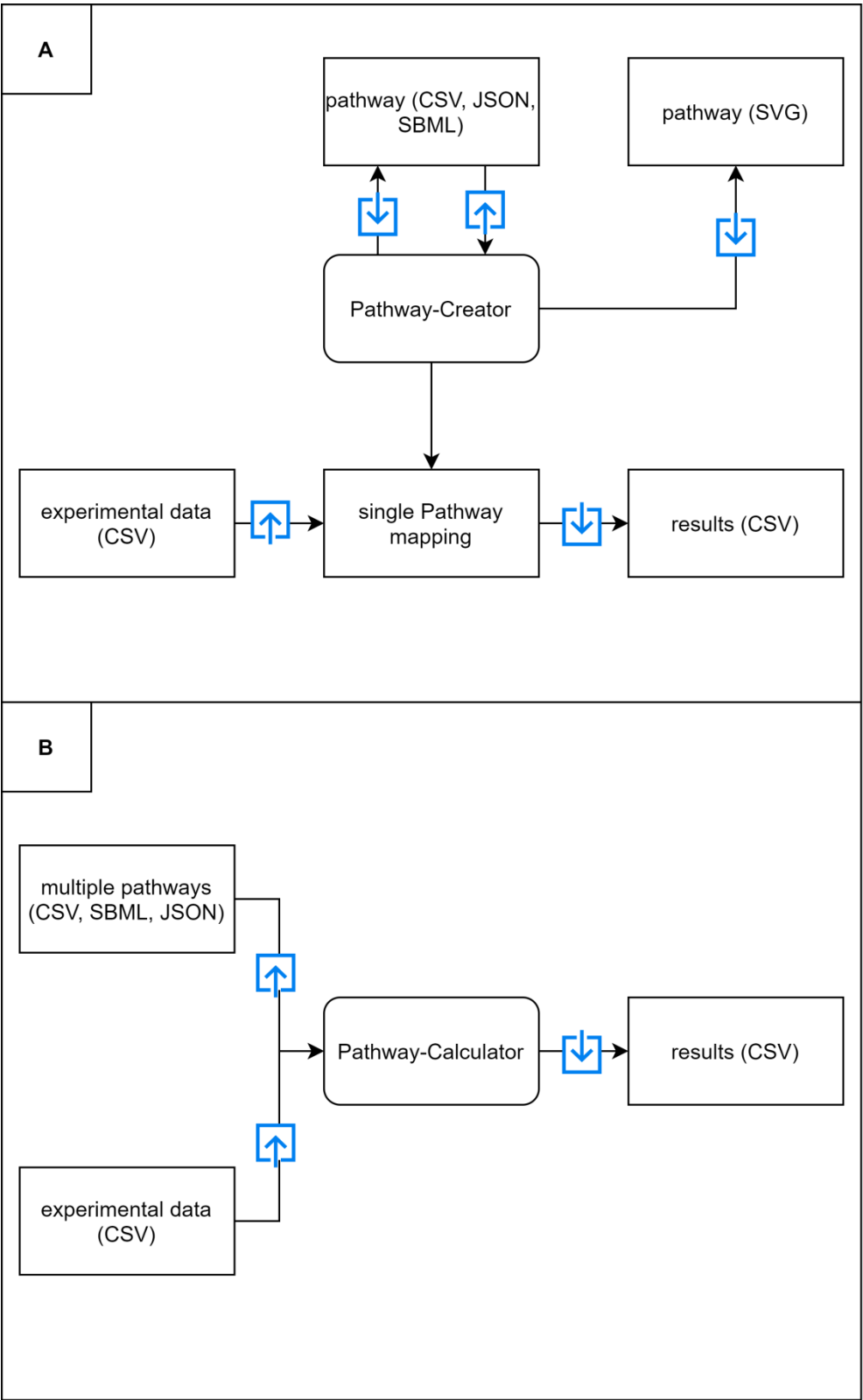


Figure 6: An overview of the structure of the MPA_Pathway_Tool with the 'Pathway-Creator' (A) and the 'Pathway-Calculator' (B). A) The 'Pathway-Creator' provides the creation of user-defined pathways. These pathways can be exported as SBML, SVG, CSV, or JSON. For later modifications, the user can also upload these pathways (CSV, JSON, SBML). Furthermore, the 'Pathway-Creator' enables the mapping of experimental data (CSV) on the created pathway. Results can be exported as CSV. B) The 'Pathway-Calculator' provides the mapping of experimental data (CSV) on multiple pathways (CSV, JSON, and SBML). The resulting mapped data can also be exported as CSV. Blue arrows represent an upload or download of specific files.

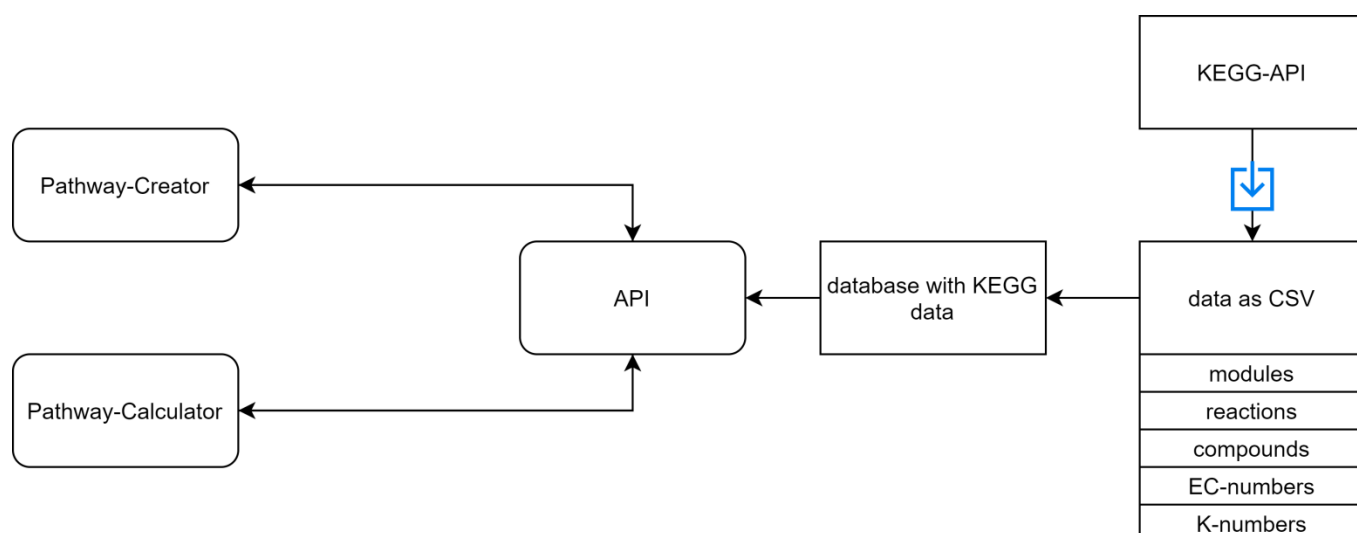


Figure 7: An overview of the REST-API. A REST- API is used for sending the required data to the 'Pathway-Creator' and the 'Pathway-Calculator'. The KEGG database has been downloaded as CSV files and parsed into our database. The available KEGG data includes pathway modules, compounds, reactions, EC numbers, and K-numbers.

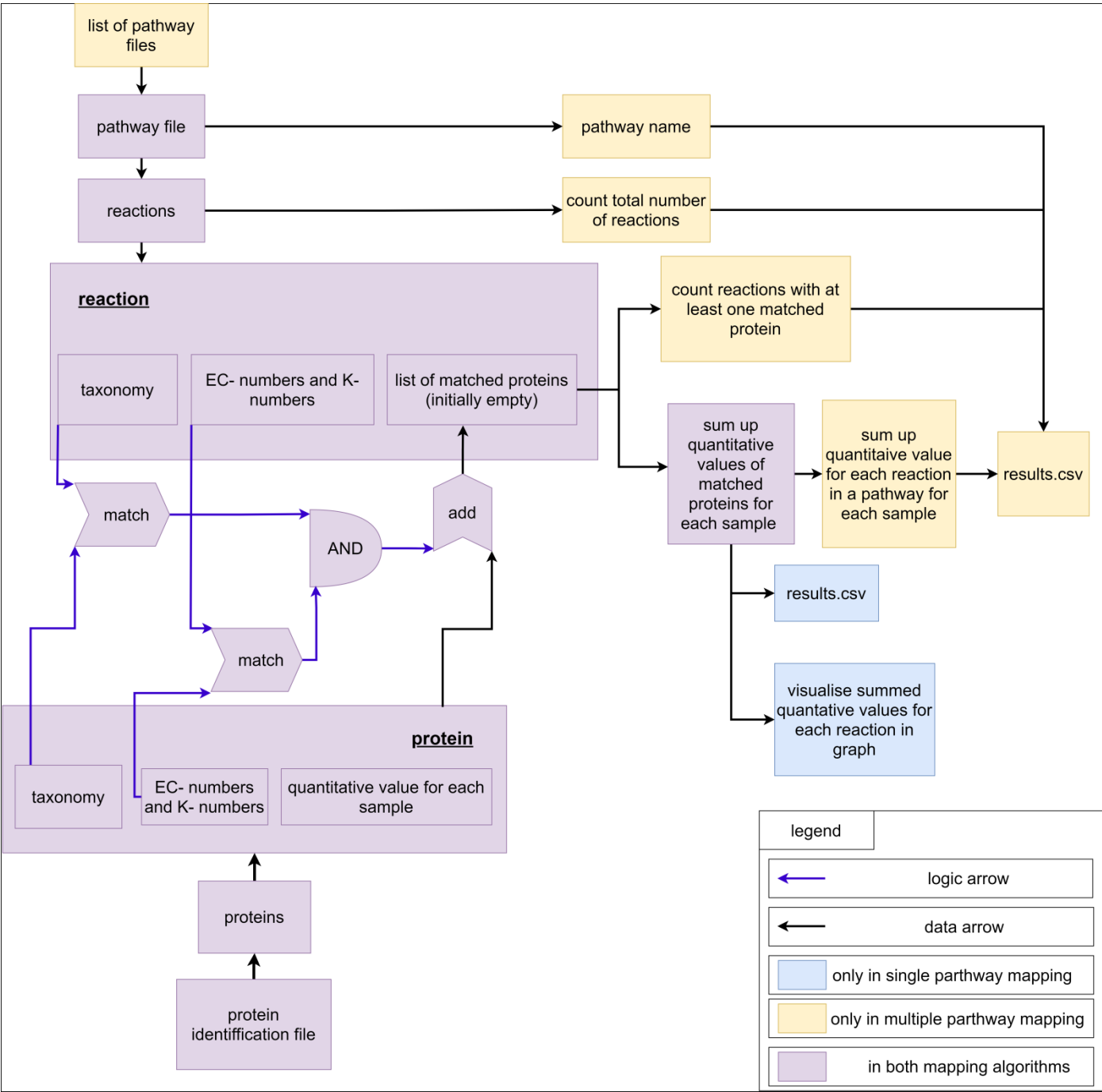


Figure 8: Scheme for visualization of the Mapping Algorithm.

Supplementary information

References

1. Johnson, C.L.; Versalovic, J. The human microbiome and its potential importance to pediatrics. *Pediatrics* **2012**, *129*, 950–960, doi:10.1542/peds.2011-2736.
2. Ursell, L.K.; Metcalf, J.L.; Parfrey, L.W.; Knight, R. Defining the human microbiome. *Nutr. Rev.* **2012**, *70 Suppl 1*, S38-44, doi:10.1111/j.1753-4887.2012.00493.x.
3. Xiong, W.; Abraham, P.E.; Li, Z.; Pan, C.; Hettich, R.L. Microbial metaproteomics for characterizing the range of metabolic functions and activities of human gut microbiota. *Proteomics* **2015**, *15*, 3424–3438, doi:10.1002/pmic.201400571.
4. Zhang, X.; Deeke, S.A.; Ning, Z.; Starr, A.E.; Butcher, J.; Li, J.; Mayne, J.; Cheng, K.; Liao, B.; Li, L.; et al. Metaproteomics reveals associations between microbiome and intestinal extracellular vesicle proteins in pediatric inflammatory bowel disease. *Nat. Commun.* **2018**, *9*, 2873, doi:10.1038/s41467-018-05357-4.
5. Biemann, R.; Buß, E.; Benndorf, D.; Lehmann, T.; Schallert, K.; Püttker, S.; Reichl, U.; Isermann, B.; Schneider, J.G.; Saake, G.; et al. Fecal Metaproteomics Reveals Reduced Gut Inflammation and Changed Microbial Metabolism Following Lifestyle-Induced Weight Loss. *Biomolecules* **2021**, *11*, 726, doi:10.3390/biom11050726.
6. Sun, L.; Liu, T.; Müller, B.; Schnürer, A. The microbial community structure in industrial biogas plants influences the degradation rate of straw and cellulose in batch tests. *Biotechnology for Biofuels* **2016**, *9*, 128, doi:10.1186/s13068-016-0543-9.
7. Kohrs, F.; Wolter, S.; Benndorf, D.; Heyer, R.; Hoffmann, M.; Rapp, E.; Bremges, A.; Sczyrba, A.; Schlüter, A.; Reichl, U. Fractionation of biogas plant sludge material improves metaproteomic characterization to investigate metabolic activity of microbial communities. *Proteomics* **2015**, *15*, 3585–3589, doi:10.1002/pmic.201400557.
8. Heyer, R.; Schallert, K.; Siewert, C.; Kohrs, F.; Greve, J.; Maus, I.; Klang, J.; Klocke, M.; Heiermann, M.; Hoffmann, H.; et al. *Metaproteome analysis reveals that syntrophy, competition, and phage-host interaction shape microbial communities in biogas plants*, 2019. Available online: <https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-019-0673-y>.
9. Morris, S.J.; Blackwood, C.B. The Ecology of the Soil Biota and their Function. In *Soil microbiology, ecology and biochemistry*, Fourth edition; Paul, E.A., Ed.; Academic Press is an imprint of Elsevier: Waltham MA, 2015; pp 273–309, ISBN 9780124159556.
10. McGuire, K.L.; Treseder, K.K. Microbial communities and their relevance for ecosystem models: Decomposition as a case study. *Soil Biology and Biochemistry* **2010**, *42*, 529–535, doi:10.1016/j.soilbio.2009.11.016.
11. Heyer, R.; Schallert, K.; Büdel, A.; Zoun, R.; Dorl, S.; Behne, A.; Kohrs, F.; Püttker, S.; Siewert, C.; Muth, T.; et al. *A Robust and Universal Metaproteomics Workflow for Research Studies and Routine Diagnostics Within 24 h Using Phenol Extraction, FASP Digest, and the MetaProteomeAnalyzer*, 2019. Available online: <https://www.frontiersin.org/articles/10.3389/fmicb.2019.01883/full>.
12. Giardine, B.; Riemer, C.; Hardison, R.C.; Burhans, R.; Elnitski, L.; Shah, P.; Zhang, Y.; Blankenberg, D.; Albert, I.; Taylor, J.; et al. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* **2005**, *15*, 1451–1455, doi:10.1101/gr.4086505.
13. Muth, T.; Behne, A.; Heyer, R.; Kohrs, F.; Benndorf, D.; Hoffmann, M.; Lehtevä, M.; Reichl, U.; Martens, L.; Rapp, E. *The MetaProteomeAnalyzer: a powerful open-source software suite for metaproteomics data analysis and interpretation*, 2015. Available online: <https://pubmed.ncbi.nlm.nih.gov/25660940/>.
14. Lange, E.; Reinert, K.; Groepl, C.; Kohlbacher, O.; Sturm, M.; Hildebrandt, A. *OPENMS; a generic open source framework for chromatography/MS-based proteomics*, 2005.
15. Leipzig, J. *A review of bioinformatic pipeline frameworks*, 2017. Available online: <https://academic.oup.com/bib/article/18/3/530/2562749>.
16. Muth, T.; Behne, A.; Heyer, R.; Kohrs, F.; Benndorf, D.; Hoffmann, M.; Lehtevä, M.; Reichl, U.; Martens, L.; Rapp, E. The MetaProteomeAnalyzer: a powerful open-source software suite for metaproteomics data analysis and interpretation. *J. Proteome Res.* **2015**, *14*, 1557–1565, doi:10.1021/pr501246w.

17. Kanehisa, M.; Goto, S.; Sato, Y.; Furumichi, M.; Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research* **2012**, *40*, D109-14, doi:10.1093/nar/gkr988.
18. Joshi-Tope, G.; Gillespie, M.; Vastrik, I.; D'Eustachio, P.; Schmidt, E.; Bono, B. de; Jassal, B.; Gopinath, G.R.; Wu, G.R.; Matthews, L.; et al. Reactome: a knowledgebase of biological pathways. *Nucleic Acids Research* **2005**, *33*, D428-32, doi:10.1093/nar/gki072.
19. Luo, W.; Pant, G.; Bhavnasi, Y.K.; Blanchard Jr, S.G.; Brouwer, C. *Pathview Web: user friendly pathway visualization and data integration*, 2017. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5570256/>.
20. King, Z.A.; Dräger, A.; Ebrahim, A.; Sonnenschein, N.; Lewis, N.E.; Palsson, B.O. Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. *PLoS Comput. Biol.* **2015**, *11*, e1004321, doi:10.1371/journal.pcbi.1004321.
21. Karp, P.D.; Paley, S.M.; Midford, P.E.; Krummenacker, M.; Billington, R.; Kothari, A.; Ong, W.K.; Subhraveti, P.; Keseler, I.M.; Caspi, R. *Pathway Tools version 24.0: Integrated Software for Pathway/Genome Informatics and Systems Biology*, 2015. Available online: <https://arxiv.org/pdf/1510.03964>.
22. Zhu, X.; Campanaro, S.; Treu, L.; Seshadri, R.; Ivanova, N.; Kougias, P.G.; Kyrpides, N.; Angelidaki, I. Metabolic dependencies govern microbial syntrophies during methanogenesis in an anaerobic digestion ecosystem. *Microbiome* **2020**, *8*, 22, doi:10.1186/s40168-019-0780-9.
23. Wagner, T.; Watanabe, T.; Shima, S. Hydrogenotrophic Methanogenesis. In *Biogenesis of Hydrocarbons*; Stams, A.J., Sousa, D., Eds.; Springer International Publishing; Imprint: Springer: Cham, 2020; pp 1–29, ISBN 978-3-319-53114-4.
24. Mesuere, B.; Devreese, B.; Debyser, G.; Aerts, M.; Vandamme, P.; Dawyndt, P. Unipept: tryptic peptide-based biodiversity analysis of metaproteome samples. *J. Proteome Res.* **2012**, *11*, 5773–5780, doi:10.1021/pr300576s.
25. Schiebenhoefer, H.; Schallert, K.; Renard, B.Y.; Trappe, K.; Schmid, E.; Benndorf, D.; Riedel, K.; Muth, T.; Fuchs, S. A complete and flexible workflow for metaproteomics data analysis based on MetaProteomeAnalyzer and Prophane. *Nat. Protoc.* **2020**, *15*, 3212–3239, doi:10.1038/s41596-020-0368-7.
26. Klamt, S.; Saez-Rodriguez, J.; Gilles, E.D. Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Syst. Biol.* **2007**, *1*, 2, doi:10.1186/1752-0509-1-2.
27. Ebrahim, A.; Lerman, J.A.; Palsson, B.O.; Hyduke, D.R. COBRApy: CONstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **2013**, *7*, 74, doi:10.1186/1752-0509-7-74.
28. Orth, J.D.; Thiele, I.; Palsson, B.Ø. What is flux balance analysis? *Nat. Biotechnol.* **2010**, *28*, 245–248, doi:10.1038/nbt.1614.
29. McCurdy, N. Getting Started | Create React App. Available online: <https://create-react-app.dev/docs/getting-started/> (accessed on 30 March 2021).
30. GitHub. reduxjs/redux. Available online: <https://github.com/reduxjs/redux> (accessed on 30 March 2021).
31. GitHub. mobxjs/mobx. Available online: <https://github.com/mobxjs/mobx> (accessed on 30 March 2021).
32. GitHub. danielcaldas/react-d3-graph. Available online: <https://github.com/danielcaldas/react-d3-graph> (accessed on 30 March 2021).
33. GitHub. axios/axios. Available online: <https://github.com/axios/axios> (accessed on 30 March 2021).
34. Material-UI: A popular React UI framework. Available online: <https://material-ui.com/> (accessed on 30 March 2021).
35. GitHub. lodash/lodash. Available online: <https://github.com/lodash/lodash> (accessed on 30 March 2021).
36. GitHub. infinidat/file-saver. Available online: <https://github.com/infinidat/file-saver> (accessed on 30 March 2021).
37. GitHub. google/gson. Available online: <https://github.com/google/gson> (accessed on 30 March 2021).
38. GitHub. sbmlteam/jsbml. Available online: <https://github.com/sbmlteam/jsbml> (accessed on 6 May 2021).