Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Atom-Resolved Genome-Scale Metabolic Models of the Human Gut Microbiome

Nora Beier

nora@bioinf.uni-leipzig.de

Leipzig University

Casper A. Eriksen

University of Southern Denmark

Natasha A. N. Jorge

Leipzig University

Thomas Gatter

Leipzig University

data-descriptor

Keywords:

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11293845/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Atom-Resolved Genome-Scale Metabolic Models of the Human Gut Microbiome

Nora Beier[1], Casper A. Eriksen[2], Natasha A. N. Jorge[1], and Thomas Gatter[1]

[2]*Department of Mathematics and Computer Science, University of Southern Denmark, Denmark*
[1]*Bioinformatics Department, Leipzig University, Germany*

October 15, 2026

Abstract

Detailed genome-scale metabolic models (GEMs) are essential for mathematical simulation of metabolism, with applications in industry and research, including bacterial community models, host-disease interaction models, and numerous medical applications.
Here, we provide GEMs of microbial species from 7 phyla of the human gut microbiome, with novel, detailed annotations of metabolic reactions using validated molecular structures and complete atom-to-atom mappings (AAMs).
By incorporating structural information with AAMs, GEM applications go beyond traditional stoichiometric analyses and enable atomic-level examination of biochemical transformations, such as isotope-labeling studies, metabolic flux tracing, and investigation of metabolic interactions in gut microbial communities. The dataset additionally contains 65 metabolic pathways, each with complete AAMs, enabling the independent modeling and analysis of individual aspects of microbial metabolism. This dataset provides a basis for developing, validating, and applying full-resolution metabolic models by combining genome-scale metabolic modeling with detailed information about atomic-level reactions.

1 Background & Summary

The increasing worldwide prevalence of obesity and associated metabolic disorders, such as type 2 diabetes, has encouraged research into human– microbiome interactions [8]. Specific microbiome composition phenotypes related to adipose populations have been identified, and studies using fecal microbiota transplantation (FMT) indicate that microbial community structure may affect host metabolism [1, 18]. To facilitate investigation of such microbial metabolism interactions, we produced an atom-resolved dataset for the SIHUMIx (Simplified Human Intestinal Microbiota) community [16]. This well-defined collection of bacteria includes eight strains that perform key intestinal metabolic functions, such as carbohydrate fermentation and short-chain fatty acid production. SIHUMIx functions as a reliable in vitro model; its simplified composition enables consistent monitoring of metabolic reactions and metabolite formation under controlled conditions while preserving functional diversity relevant to intestinal metabolism.

The dataset provides validated, uniform metabolite structures and atom-to-atom mapping (AAM) information for metabolic reactions encoded in each strain's genome-scale networks. These atom-level annotations enable modeling of metabolic fluxes and interspecies metabolite exchange, preserve mass-balance constraints, and allow integration with isotopic labeling experiments. In addition, the dataset provides comprehensive AAMs for 65 distinct metabolic pathways from the well-established *Escherichia coli* MG1655 model. These data facilitate the simulation and analysis of metabolic subnetworks, allowing users to focus on specific functional modules without requiring the full community-scale network. Overall, the dataset comprises 4,356 structured and validated reactions with associated AAMs. It supports quantitative modeling and comparative analyses of microbial metabolism, providing a standardized resource for studying gut microbial ecology.

2 Methods

2.1 Workflow Overview

All included metabolic models and pathways were collected from reputable sources. Structures and AAMs were carefully annotated using a mixture of manual curation and automated methods. As a particular challenge, we addressed the difficulties arising from inconsistent annotation between or within compound and reaction databases. Figure 1 provides an overview of the general workflow used to generate the datasets. The individual steps are

described in greater detail below.

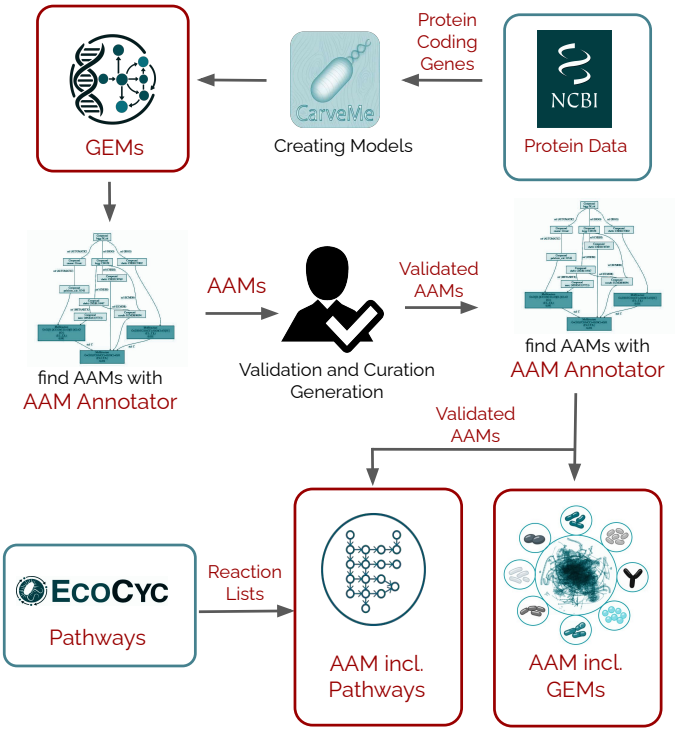


Figure 1: Overview of the workflow for generating AAM-integrated GEMs and pathways. Petrol-colored boxes represent inputs, while red boxes indicate workflow outputs. First, we generated GEMs from protein coding sequences obtained from NCBI. Subsequently, AAMs were assigned to all reactions within the GEMs. We manually validated these assignments and corrected them where necessary. We then incorporated the validated AAM dataset as curation information into the AAM annotation process and regenerated the GEMs with the curated AAM assignments. The validated AAM dataset was also used to assign the corresponding AAMs to reactions within the respective pathways.

2.2 Generation of the Genome-Scale Models of the Gut Bacteria

We evaluated and refined the *universe* metabolic template network files for Gram-positive and Gram-negative networks provided by CarveMe 1.5.1 [11]

using the report snapshot function of MEMOTE 0.14.0 [10].

For reactions marked as mass- or charge unbalanced by MEMOTE, we manually inspected and corrected metabolite charges, molecular formulas, and reaction stoichiometry against the consensus among KEGG [7], BiGG [13], MetaNetX [12], ModelSEED [17], and ChEBI [3]. Edits were directly applied to the model files until no unbalanced reactions remained. Additionally, we added the flavin mononucleotide (FMN) metabolite to the Growth reaction to satisfy MEMOTE quality standards and accurately reflect its essential role as an indispensable cofactor in bacterial metabolism. We then downloaded translated CDS protein FASTA files from NCBI for each organism in SIHUMIX, as follows:

- Anaerostipes caccae: GCA_014131675.1_ASM1413167v1
- Bacteroides thetaiotaomicron: GCA_014131755.1_ASM1413175v1
- Bifidobacterium longum: GCA_000007525.1_ASM752v1
- Blautia producta: GCA_014131715.1_ASM1413171v1
- Clostridium butyricum: GCA_014131795.1_ASM1413179v1
- Erysipelatoclostridium ramosum: GCA_014131695.1_ASM1413169v1
- Escherichia coli: GCA_000005845.2_ASM584v2
- Lactobacillus plantarum: GCA_014131735.1_ASM1413173v1

Using these FASTA files and the manually refined universe files, we created organism-specific metabolic models using CarveMe, retaining only the top 3-5 protein hits identified via DIAMOND v.2.19.163 [2]. Periplasmatic transport and missing periplasmatic metabolites were manually incorporated where appropriate. Finally, we re-evaluated organism-specific networks using MEMOTE report snapshot and refined using the same criteria as applied TO the *universe* templates. Model simulations were constrained to represent growth in CIM-VIII medium, reflecting experimental bioreactor conditions used to cultivate all SIHUMIX species.

2.3 Generation of Initial Atom-to-Atom Mapped Networks

We automatically annotated the GEMs using AAM Annotator [4], which uses the meta-database ChemRecon [5] to retrieve consensus molecular structures

and curated AAMs from several biochemical databases. The ChemRecon database contains 342 AAMs derived from MCS-A and 1,003 AAMs derived from MetaMDB, providing a comprehensive set of curated AAMs for reaction annotation.

We matched candidate AAMs to model reactions based on their similarity to known chemical structures. For reactions with no curated AAMs, we generated AAMs using the Reaction Decoder Tool [14]. AAM Annotator also reports confidence scores, the provenance of chemical structures and AAMs, and potential inconsistencies and errors, which were used to guide manual validation and refinement.

2.4 Validation and Extension of Atom-to-Atom Maps

All AAMs, independent of their source, underwent systematic manual curation. Each mapping was reviewed for consistency. Where necessary, corrections were applied. We manually annotated reactions lacking AAM information in all source databases to ensure comprehensive coverage of the metabolic networks.

Where available, reaction mechanisms from the MCS-A database [15] served as a priority reference for constructing AAMs. These mechanistic descriptions guided atom correspondence assignments and ensured consistency with known biochemical transformations. When no mechanistic information was available, we constructed AAMs based on biochemical expertise. We then reviewed the resulting mappings for internal consistency. We retained them in the dataset but explicitly annotated them to indicate the absence of a validated mechanistic reference.

We obtained reaction sets for pathway-level organization from the EcoCyc database [9]. Together, these pathways represent the complete metabolic network of *Escherichia coli*. We extracted the corresponding AAMs from the curated genome-scale metabolic models and mapped them onto the pathway definitions. This establishes a consistent link between reaction-level AAMs and higher-level pathway structures.

3 Data Record

3.1 Genome-Scale Models (GEMs)

This dataset provides GEMs in two distinct variants. The first consists of the GEMs we generated in pure SBML format. The second comprises the

identical GEMs with comprehensive structure and AAM annotations in a custom format.

Accordingly, both variants share an identical reaction set, albeit neither is a subset of the other. The following schema provides a detailed overview of the structure and specific characteristics of both datasets.



Table 1 shows the GEMs of SIHUMIx for the following gut-associated microbial species.

Bacterial Strains	Score	Filename
Anaerostipes caccae DSMZ 14662	88%	Acaccae.json
Bacteroides thetaiotaomicron DSMZ 2079		Btheta.json
Blautia producta DSMZ 2950	89%	Bproducta.json
Clostridium butyricum DSMZ 1702	89%	Cbuty.json
Clostridium ramosum DSMZ 1402	89%	Eramosum.json
Escherichia coli K-12 MG1655	89%	Ecoli.json
Lactobacillus plantarum DSMZ 20174	89%	Lplantarum.json

Table 1: The score measures network quality and was calculated using MEMOTE. The GEM files are available at https://github.com/NoraBeier/cAAMsGEMs_of_SIHUMIx/tree/main.

Compared with the ANGORA2 models [6] of the same species published in the Metabolic Atlas database (Anton et al. 2020), our networks demonstrate superior quality, as the MEMOTE scores of the ANGORA2 models average only 73 %.

3.2 Pathways

The pathway data is provided in CSV format and contains detailed information on individual metabolic reactions and their atom-level transformations. The file includes the following columns:

- **Reaction_index**: The unique reaction identifier from EcoCyc, serving as the reference for each metabolic transformation.
- **Substrate and Product**: Metabolites involved in the reaction, annotated with their corresponding EcoCyc identifiers to enable precise mapping of reaction participants.
- **Bigg_IDs**: The corresponding reaction identifiers in the BiGG database, facilitating cross-references between metabolic reconstruction frameworks.
- **Atom-to-Atom Maps**: Atom-to-atom mappings in SMILES format that capture the precise correspondence of atoms from substrate to product for each reaction.

This structure enables detailed pathway analysis, integration with genome-scale metabolic models, and atom-level tracing of metabolic transformations, including for specific metabolic sub-networks. The dataset contains 65 pathways, listed in the appendix.

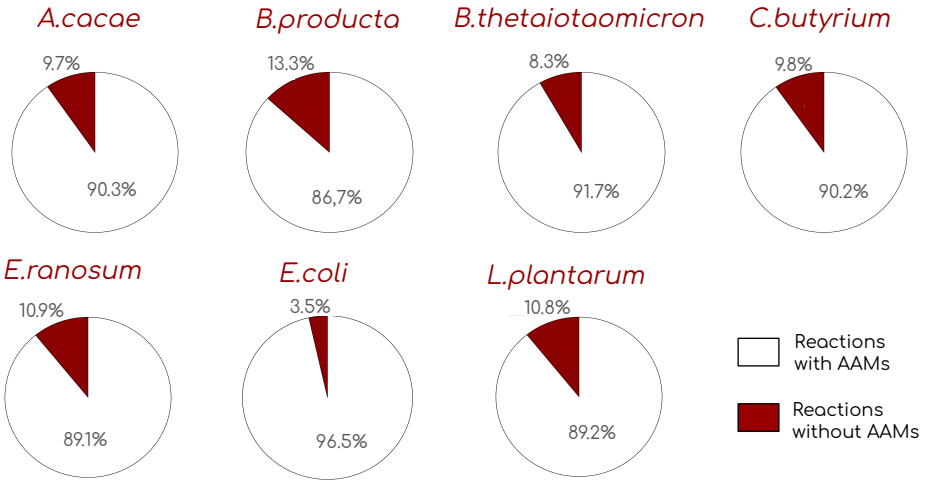
4 Technical Validation

4.1 Uniform Molecule Structure Test

After manually correcting the atom-to-atom mappings (AAMs), we re-evaluated all metabolites to ensure structural consistency across the dataset. We standardized variations arising from common biochemical conventions for representing molecular residues and substructures to the most detailed molecular representation available. This harmonization step ensured uniform structural detail across the dataset.

4.2 Statistics

Across the seven gut microbiome species models considered, we achieved an average reaction coverage of up to 90.5% with AAM annotations (Fig. 2). Overall, the models showed consistently high atom-mapping coverage, indicating that most reactions in the reconstructed metabolic networks could be associated with atom-mapping information. The proportion of reactions with AAM annotations ranged from approximately 86% in *B.producta* to 93% in *B.thetaiotaomicron*.



Model	Metabolites	Reactions	Reactions with AAMs
A.cacae	781	1486	1324
B.producta	759	1518	1320
B.thetaiotaomicron	757	1538	1375
C.buty	771	1611	1428
E.coli	894	2417	2211
E.ramosum	611	1159	1026
L.plantarum	643	1236	1107

Figure 2: AAM coverage of all SIHUMIx GEMs. The figure shows the AAM coverage of all models, while the table summarises the corresponding total number of reactions and reactions with annotated atom mappings (AAMs).

The total number of reactions varied considerably across models, reflecting differences in metabolic network size. *E. coli* contained the largest network with 2,417 reactions, of which 2,211 were associated with AAM annotations. In contrast, *E.ramosum* had the smallest model, with 1,159 reactions, including 1,026 reactions with AAM annotations. Despite these differences in network size, all models exhibited a substantial number of reactions with available atom mappings.

The number of metabolites also differed between the models, ranging from 611 metabolites in *E.ramosum* to 894 metabolites in *E.coli*. Thus, network size variation was reflected in both the number of metabolites and reactions. Taken together, these results demonstrate that the generated GEMs provide broad AAM coverage across the investigated species, facilitating subsequent analyses that depend on atom-level information.

4.3 Limitations

Symmetric molecules may, under certain conditions, yield multiple plausible AAMs. However, this work did not explicitly consider such symmetries, and future modeling efforts should incorporate them. A small number of reactions within the dataset contain rational stoichiometric coefficients. These reactions are exceptions because they are not atom-balanced. Therefore, atoms that are not conserved and consequently absent on the opposite side of the reaction lack corresponding atom mapping IDs.

A reaction can be catalyzed by different enzymes, just as a single enzyme may catalyze multiple reactions. Consequently, these processes may also result in different AAMs. Unfortunately, enzyme-catalyzed multi-step reaction mechanisms are often unavailable. Therefore, in this dataset, the biochemically most plausible variant was used in most cases.

For 169 reactions, no AAMs are available. These reactions involve molecules with sufficient structural complexity or with missing molecular structure information, making AAM generation infeasible. The appendix lists the affected reactions.

5 Data Availability

The data is publicly available at the following GitHub repository:
https://github.com/NoraBeier/cAAMsGEMs_of_SIHUMIx/tree/main/data
The dataset is distributed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

6 Code Availability

The code used to generate the genome-scale metabolic models (GEMs) for the SIHUMIx consortium is publicly available on GitHub:
https://github.com/NoraBeier/cAAMsGEMs_of_SIHUMIx/tree/main/code

7 Author Contributions

Nora Beier validated and supplemented the AAMs in the networks. Casper A. Eriksen queried the database for already validated AAM data. Natasha A. N. Jorge manually curated and generated the GEMs for the SIHUMIx species. Thomas Gatter supervised the project.

8 Competing interests

The authors declare no competing interests.

9 Funding

This research was funded by the Novo Nordisk Foundation (NNF21OC0066551"MATOMIC"). Natasha A. N. Jorge was also funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project number 517078227.

References

[1] Fredrik Bäckhed et al. "The gut microbiota as an environmental factor that regulates fat storage". In: *Proceedings of the National Academy of Sciences* 101.44 (2004). DOI: 10.1073/pnas.0407076101.

[2] Benjamin Buchfink, Klaus Reuter, and Hajk-Georg Drost. "Sensitive protein alignments at the tree-of-life scale using DIAMOND". In: *Nature Methods* 18 (2021), pp. 366–368. DOI: 10.1038/s41592-021-01101-x.

[3] Kirill Degtyarenko et al. "ChEBI: a database and ontology for chemical entities of biological interest". In: *Nucleic Acids Research* 36.Database issue (2008), pp. D344–D350. DOI: 10.1093/nar/gkm791.

[4] Casper Asbjørn Eriksen. *AAM Annotator*. Python package. 2024. URL: https://libraries.io/pypi/aam-annotator.

[5] Casper Asbjørn Eriksen et al. "ChemRecon: a Consolidated Meta-Database Platform for Biochemical Data Integration". In: *arXiv* (2026). DOI: 10.48550/arXiv.2602.12310. arXiv: 2602.12310 [q-bio.QM].

[6] Almut Heinken et al. "AGORA2: Large scale reconstruction of the microbiome highlights wide-spread drug-metabolising capacities". In: *bioRxiv* (2020). DOI: 10.1101/2020.11.09.375451. eprint: https://www.biorxiv.org/content/early/2020/11/10/2020.11.09.375451.full.pdf. URL: https://www.biorxiv.org/content/early/2020/11/10/2020.11.09.375451.

[7] Minoru Kanehisa and Susumu Goto. "KEGG: Kyoto Encyclopedia of Genes and Genomes". In: *Nucleic Acids Research* 28.1 (2000), pp. 27–30. DOI: 10.1093/nar/28.1.27.

[8] Fredrik Karlsson et al. "Assessing the Human Gut Microbiota in Metabolic Diseases". In: *Diabetes* 62.10 (Sept. 2013), pp. 3341–3349. ISSN: 0012-1797. DOI: 10.2337/db13-0844. eprint: https://diabetesjournals.org/diabetes/article-pdf/62/10/3341/569702/3341.pdf. URL: https://doi.org/10.2337/db13-0844.

[9] Peter D. Karp et al. "The EcoCyc database (2025)". In: *EcoSal Plus* 13.1 (2025), eesp-0019–2024. DOI: 10.1128/ecosalplus.esp-0019-2024. eprint: https://journals.asm.org/doi/pdf/10.1128/ecosalplus.esp-0019-2024. URL: https://journals.asm.org/doi/abs/10.1128/ecosalplus.esp-0019-2024.

[10] Christian Lieven, Moritz E. Beber, Brett G. Olivier, et al. "MEMOTE for standardized genome-scale metabolic model testing". In: *Nature Biotechnology* 38 (2020), pp. 272–276. DOI: 10.1038/s41587-020-0446-y.

[11] Daniel Machado et al. "Fast automated reconstruction of genome-scale metabolic models for microbial species and communities". In: *Nucleic Acids Research* 46.15 (2018), pp. 7542–7553. DOI: 10.1093/nar/gky537.

[12] Simone Moretti et al. "MetaNetX: a bridge between metabolic resources for enhanced curation and multi-omics data harmonization". In: *Nucleic Acids Research* 54.D1 (2026), pp. D617–D622. DOI: 10.1093/nar/gkaf1286.

[13] Corey J. Norsigian et al. "BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree". In: *Nucleic Acids Research* 48.D1 (2020), pp. D402–D406. DOI: 10.1093/nar/gkz1054.

[14] Syed Asad Rahman et al. "Decoder Tool (RDT): extracting features from chemical reactions". In: *Bioinformatics* 32.23 (2016), pp. 2065–2066. DOI: 10.1093/bioinformatics/btw096.

[15] António J M Ribeiro et al. "Mechanism and Catalytic Site Atlas (M-CSA): a database of enzyme reaction mechanisms and active sites". In: *Nucleic Acids Research* 46.D1 (Nov. 2017), pp. D618–D623. ISSN: 0305-1048. DOI: 10.1093/nar/gkx1012.

[16] Stephanie Serena Schäpe et al. "The Simplified Human Intestinal Microbiota (SIHUMIx) Shows High Structural and Functional Resistance against Changing Transit Times in In Vitro Bioreactors". In: *Microorganisms* 7.12 (2019), p. 641. DOI: 10.3390/microorganisms7120641. URL: https://www.mdpi.com/2076-2607/7/12/641.

[17] Samuel M. D. Seaver et al. "The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes". In: *Nucleic Acids Research* 49.D1 (2021), pp. D575–D588. DOI: 10.1093/nar/gkaa746.

[18] Zhengxiao Zhang et al. "Impact of Fecal Microbiota Transplantation on Obesity and Metabolic Syndrome-A Systematic Review". In: *Nutrients* 11.10 (2019), p. 2291. DOI: 10.3390/nu11102291.

Appendix

Table 2: List of pathways with filenames and number of reactions

Pathway	**Total Reactions**
2-O-α-Mannosyl-D-glycerate Degradation	2
3-Dehydroquinate Biosynthesis I	2
4-Aminobenzoate Biosynthesis I	2
4-Aminobutanoate Degradation I	2
4-Aminobutanoate Degradation II	2
4-Hydroxybenzoate Biosynthesis II	1
6-Hydroxymethyl Dihydropterin Diphosphate Biosynthesis I	5
Adenine and Adenosine Salvage II	4
Adenine and Adenosine Salvage III	4
Adenosine Ribonucleotides De Novo Biosynthesis	3
Aminopropylcadaverine Biosynthesis	3
Cadaverine Biosynthesis	1
Chorismate Biosynthesis from 3-dehydroquinate	5
CMP Phosphorylation	2
Demethylmenaquinol-8 Biosynthesis I	1
Entner-Doudoroff Shunt	2
Glycolysis II	20
Glyoxylate Cycle	6
Guanosine Ribonucleotides De Novo Biosynthesis	4
Inosine-5-phosphate Biosynthesis I	6
L-Alanine Biosynthesis I	1
L-Arginine Biosynthesis I via L-Ornithine	9
L-Arginine Degradation II	6
L-Asparagine Degradation I	1
L-Aspartate Degradation II aerobic	3
L-Aspartate Degradation III anaerobic	3
L-Cysteine Biosynthesis I	2
L-Glutamate Biosynthesis	3
L-Glutamate Degradation II	4
L-Glutamate Degradation IX via 4-Aminobutanoate	1
L-Glutamine Degradation I	1
L-Histidine Biosynthesis	10
L-Homoserine Biosynthesis	3
L-Leucine Biosynthesis	6
L-Lysine Biosynthesis I	9
L-Ornithine Biosynthesis I	5
L-Proline Biosynthesis I from L-Glutamate	3
L-Threonine Biosynthesis	6
L-Threonine Degradation II	2
L-Threonine Degradation III	3
L-Threonine Degradation IV	4
L-Valine Biosynthesis	4
Menaquinol-8 Biosynthesis	1

Pathway	**Total Reactions**
Mixed Acid Fermentation	28
Pentose Phosphate Pathway	8
Phosphopantothenate Biosynthesis I	5
Proline Degradation	3
Putrescine Biosynthesis I	2
Putrescine Biosynthesis III	1
Pyridoxal-5-Phosphate Salvage I	6
Pyrimidine Deoxyribonucleotide Phosphorylation	4
Pyrimidine Deoxyribonucleotides De Novo Biosynthesis I	9
Pyrimidine Nucleobases Salvage I	1
Pyrimidine Nucleobases Salvage II	2
Pyrimidine Ribonucleosides Salvage I	3
Pyrimidine Ribonucleosides Salvage II	2
Pyrimidine Ribonucleosides Salvage III	2
S-Adenosyl-L-Methionine Biosynthesis	1
Salvage Pathways of Pyrimidine Deoxyribonucleotides	5
Salvage Pathways of Pyrimidine Ribonucleotides	8
Spermidine Biosynthesis I	2
Superpathway of 4-Aminobutanoate Degradation to Succinate	3
Tetrahydrofolate Biosynthesis I	3
UMP Biosynthesis I	6
UTP and CTP De Novo Biosynthesis	4

Table 3: Complex Reactions without Atom-to-Atom Mapping.

R_I2FE2SR	R_I2FE2ST	R_I4FE4ST	R_S2FE2SR
R_S2FE2SS	R_S2FE2SS2	R_S2FE2ST	R_S4FE4SR
R_EX_glcman6_e	R_CBLAT	R_ADCPS1	R_AGPATr_BS
R_ALMPC	R_AMALT1	R_AMALT2	R_AMALT3
R_AMALT4	R_COB1NDAAT1	R_COB1NDAAT2	R_COLIPAabcpp
R_COLIPAabctex	R_COLIPAKpp	R_COLIPAPabctex	R_CYTMQOR3pp
R_ECA4COLIPAabctex	R_ECA4OALpp	R_EDTXS1	R_EDTXS3
R_EDTXS4	R_EX_glucan4_e	R_EX_glucan6_e	R_EX_pdima_e
R_EX_xylan4_e	R_FAO2	R_FASm220	R_FASm240
R_FASm260	R_FASm280	R_FE3HOXR1	R_FE3HOXR2
R_FE3HOXR3	R_G3POA_BS	R_GCCa_1	R_GCCb_1
R_GCCc_1	R_GCVHADPr	R_GCVHRADPr	R_GDOCBIK
R_GLBRAN3	R_GLCMAN4	R_GLCMAN4abc	R_GLCMAN6
R_GLCMAN6abc	R_GLCP4	R_GLCS3_1	R_GLCTR2
R_GLCTR3	R_GLCURS1	R_GLDBRAN3	R_GLS
R_GLS2	R_GLUCAN2abc	R_GLUCANabc	R_HEPT4
R_KAS1	R_KAS11	R_KAS12	R_KAS13
R_KAS17	R_KAS2	R_KAS3	R_KAS4
R_KAS6	R_KAS7	R_KAS8	R_LA4NTpp
R_LIPAHT2ex	R_LIPO1S24_BS	R_LIPO2S24_BS	R_LIPO3S24_BS
R_LIPO4S24_BS	R_MCBTS3	R_MCTP1Bpp	R_MCTP2App
R_MDDCP1pp	R_MDDCP2pp	R_MDDCP3pp	R_MDDCP4pp
R_MDDEP1pp	R_MDDEP2pp	R_MDDEP3pp	R_MDDEP4pp
R_MLDCP1App	R_MLDCP1Bpp	R_MLDCP2App	R_MLDCP2Bpp
R_MLDCP3App	R_MLDEP1pp	R_MLDEP2pp	R_MLTGY4pp
R_MPTG2	R_O16A4COLIPAabctex	R_O16A4Lpp	R_O16AP1pp
R_O16AP2pp	R_O16AP3pp	R_PAP	R_PAP_BS
R_PC2DHG	R_PDBL_3	R_PDBL_4	R_PLPS
R_SMIA1	R_SMIA2abc	R_SMIB1	R_TECA1S45
R_TECA2S45	R_TECA3S45	R_THZPSN2	R_UGT_BS
R_XYLANabc	R_ACBIPGT	R_EX_fe3dhbzs3_e	R_EX_fe3pyovd_kt_e
R_FE3PYOVD2tonex	R_FE3PYOVDDR2	R_FE3PYOVDL2	R_FEENTERES
R_MOADSUx	R_MPTSS	R_TDSR1	R_TDSR2

Table 4: Reactions with missing Molecule Structure Information

R_ACKILE	R_GALMAN4	R_ACKLEU	R_GALMAN4abc
R_ACKVAL	R_GALMAN6	R_CMCBTFL	R_GALMAN6abc
R_CMCBTFtex	R_PDIMAS	R_CMCBTFtpp	R_PDIMAT
R_CMCBTFU	R_PHTHDLS	R_DAGK	R_PHTHS
R_DNTAep	R_PREPHACPH	R_EX_35dnta_e	R_PREPTHS
R_EX_cmcbtt_e	R_RECOAH9	R_EX_galman4_e	R_TNTHAC1
R_EX_galman6_e	R_TNTHAC2	R_FACOAL140_ISO	R_TNTR3
R_FACOAL150_anteiso	R_TNTR4	R_FACOAL150_ISO	R_FACOAL160_ISO
R_FACOAL170_anteiso	R_FACOAL170_ISO		