Generation of a chemoinformatic data base

Author: German Preciat, Analytical BioSciences and Metabolomics, Leiden University.

INTRODUCTION

In a COBRA model, the molecular structure of a metabolite can be represented in different formats such as SMILES InChIs and InChIkeys (See Figure 1), as well as identifiers from different databases such as the Virtual Metabolic Human database [1] (**VMH**), the Human Metabolome Database [2] (**HMDB**), **PubChem** database [3], the Kyoto Encyclopedia of Genes and Genomes [4] (**KEEG**), and the Chemical Entities of Biological Interest [5] (**ChEBI**).

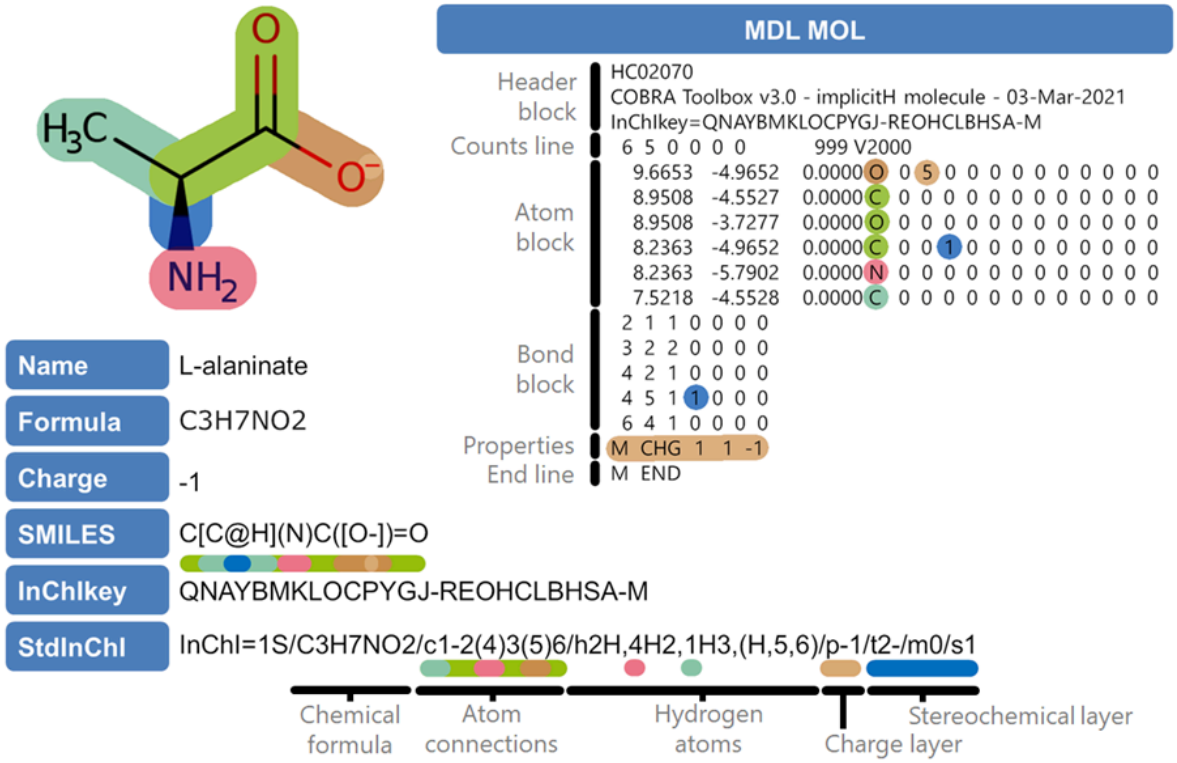


Figure 1. Different chemoinformatic formats

The function `generateChemicalDatabase` generates a chemoinformatic database with the data in a COBRA model (Figure 2). It uses different external softwares among which are openBabel [6], MarvinSuite [7] and JAVA. `generateChemicalDatabase` works if some of these software are not installed but for best results, it is recommended to have all of them in the system.

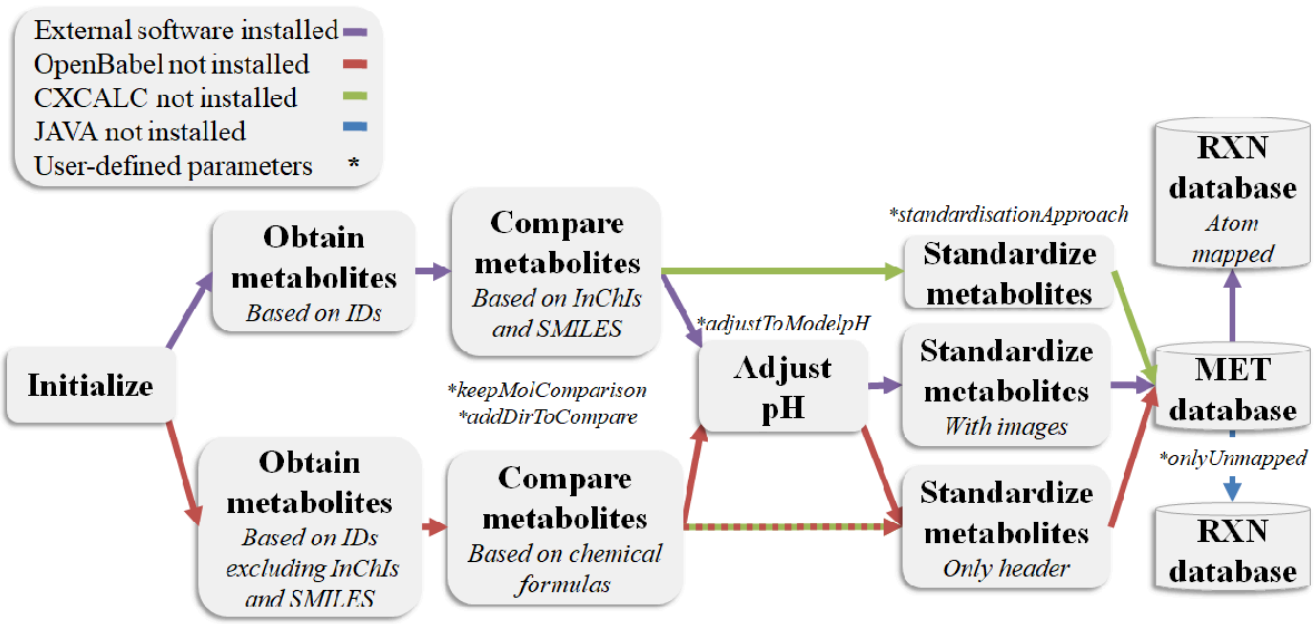


Figure 2. The pipeline can take different paths depending on the system configuration and the user-defined parameters.

The function compares for each metabolite the molecular structure information from the different sources and the molecular structure of the source with the highest score is considered as correct. The disimilarity comparison is done by calculating the euclidean distances of the score of all molecular structures obtained. The score is assigned based on the following criteria:

Comparison	Max score	Description
Similarity with other databases	1	Compares the similarity between sources by dividing the repeated instances of an identifier by the number of unique identifiers. A source without an identifier is not considered in the dissimilarity test.
Chemical formula	10	Compares the chemical formula indicated in the COBRA model and the chemical formula obtained by the source. It gives a score of 10 if match and a score of 8 if match without hydrogen atoms.
InChI layers	11	Compares the layer of information between the InChIs obtained. Start with 4 and each layer gives a score of 1.

The comparison aims at obtaining a greater number of atomically balanced metabolic reactions. The atom mappings of each of the metabolic reactions are obtained with the Reaction Decoder Tool algorithm [8] (**RDT**). Finally, the atom mapping data is used to calculate the number of bonds formed or broken and the energy required to break the bond of a metabolic reaction. All the information obtained is integrated into the COBRA model.

In this tutorial, we will obtain the chemoinformatic data on the Ecoli core model. The identifiers will be added using the funciton `addMetInfoInCBmodel`.

Generate chemoinformatic database

Clean the workspace and set the user directory.

```
clear
if ispc
    userDir = winqueryreg('HKEY_CURRENT_USER',...
        ['Software\Microsoft\Windows\CurrentVersion\' ...
         'Explorer\Shell Folders'],'Personal');
else
    userDir = char(java.lang.System.getProperty('user.home'));
end
```

Add MarvinSuite to the enviroment.

```
if ismac
    setenv('PATH', [getenv('PATH') ':/usr/local/bin:/usr/bin:/bin:/Applications/MarvinS
end
```

Load the ecoli core model.

```
load ecoli_core_model.mat
model.mets = regexprep(model.mets, '\-', '\_');
```

From an external file, add chemoinformatic information in the ecoli core model such as SMILES, InChIs or different database identifiers.

```
dataFile = which('chemoinformaticDatabaseTutorial.mlx')
```

```
dataFile =
'/Users/gpreciat/work/code/cobratoolbox/tutorials/dataIntegration/chemoinformaticDatabase/chemoinformatic
```

```
inputData = regexprep(dataFile, 'chemoinformaticDatabaseTutorial.mlx', 'metaboliteIds.x
replace = false;
[model, hasEffect] = addMetInfoInCBmodel(model, inputData, replace);
```

The user-defiined parameters will activate different processes in the function `generateChemicalDatabase`. Each parameter is described as follows:

- **outputDir**: Path to directory that will contain the RXN files with atom mappings (default: current directory)
- **printlevel**: Verbose level
- **standardisationApproach**: String containing the type of standardisation for the molecules (default: 'explicitH' if openBabel O'Boyle et al. 2011 is installed, otherwise default: 'basic'):

1. explicitH: Normal chemical graphs;
2. implicitH: Hydrogen suppressed chemical graph;
3. neutral: Chemical graphs with protonated molecules;
4. basic: Update the header.

- **keepMolComparison**: Logic value to keep the MDL MOL files from the different sources compared (default: FALSE)
- **onlyUnmapped**: Logic value to select create only unmapped MDL RXN files (default: FALSE).
- **adjustToModelpH**: Logic value to select if the pH of a molecule has to be adjusted according the GEM (default: TRUE, requires CXCALC ChemAxon 2015).
- **addDirsToCompare**: Cell(s) with the path to directory to an existing database (default: empty)
- **dirNames**: Cell(s) with the name of the directory(ies) (Default: empty)
- **debug**: Logical value to establish if the results in different points of the function will be saved to debug.

```
options.outputDir = [userDir '/work/code/ctf/databases/ecoliDB'];
options.printlevel = 1;
options.debug = true;
options.standardisationApproach = 'explicitH';
options.adjustToModelpH = true;
options.keepMolComparison = true;
options.dirsToCompare = {'~/work/code/ctf/mets/molFiles'};
options.onlyUnmapped = false;
options.dirNames = {'VMH'};
```

Use the function generateChemicalDatabase

```
info = generateChemicalDatabase(model, options);
```

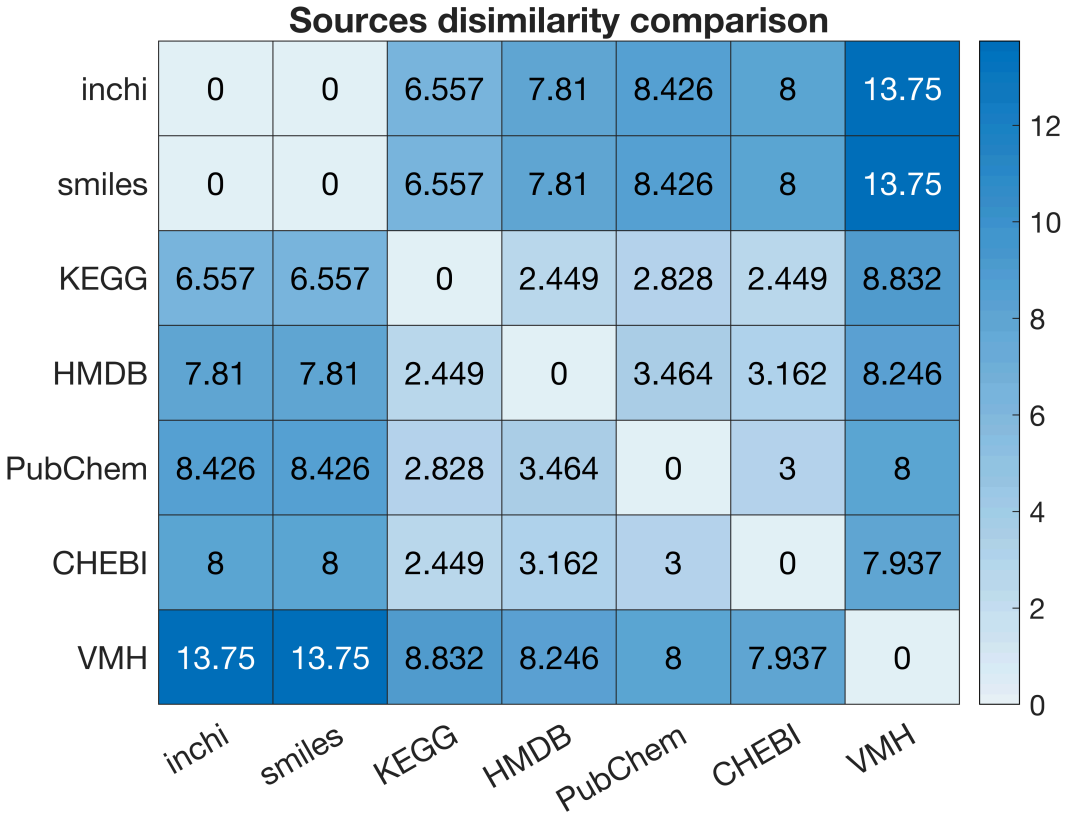
```
------------------------------------------------------------
CHEMICAL DATABASE
------------------------------------------------------------

Generating a chemical database with the following options:

                  outputDir: '/Users/gpreciat/work/code/ctf/databases/ecoliDB'
                 printlevel: 1
                      debug: 1
    standardisationApproach: 'explicitH'
            adjustToModelpH: 1
          keepMolComparison: 1
              dirsToCompare: {'~/work/code/ctf/mets/molFiles/'}
               onlyUnmapped: 0
                   dirNames: {'VMH'}

------------------------------------------------------------
Obtaining MOL files from chemical databases ...

inchi:
molCollectionReport = struct with fields:
              noOfMets: 54
       noOfMetsWithMol: 50
    noOfMetsWithoutMol: 4
              coverage: 92.5926
smiles:
molCollectionReport = struct with fields:
              noOfMets: 54
       noOfMetsWithMol: 50
    noOfMetsWithoutMol: 4
              coverage: 92.5926
```

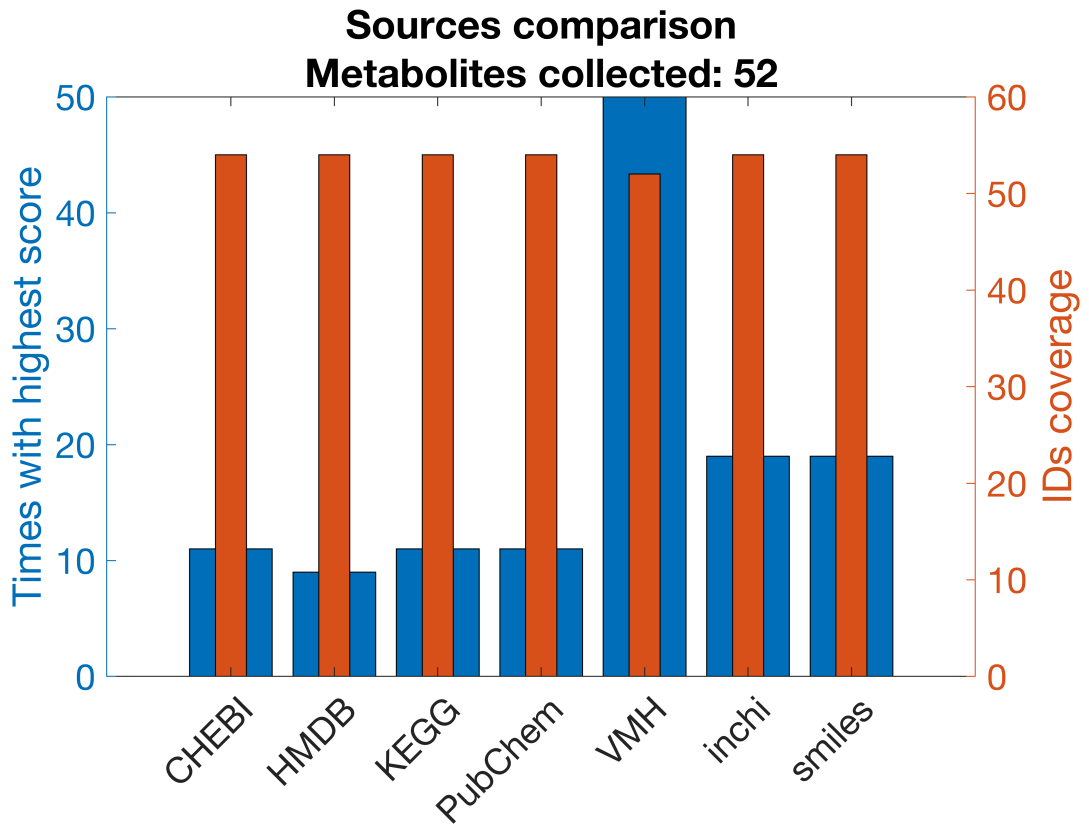
```
KEGG:
molCollectionReport = struct with fields:
              noOfMets: 54
       noOfMetsWithMol: 49
    noOfMetsWithoutMol: 5
              coverage: 90.7407
HMDB:
molCollectionReport = struct with fields:
              noOfMets: 54
       noOfMetsWithMol: 51
    noOfMetsWithoutMol: 3
              coverage: 94.4444
PubChem:
molCollectionReport = struct with fields:
              noOfMets: 54
       noOfMetsWithMol: 50
    noOfMetsWithoutMol: 4
              coverage: 92.5926
CHEBI:
molCollectionReport = struct with fields:
              noOfMets: 54
       noOfMetsWithMol: 48
    noOfMetsWithoutMol: 6
              coverage: 88.8889
```

```
Comparing information from sources ...
  52×8 table
```

mets	sourceWithHighestScore	metNames
"13dpg"	"VMH"	"3-Phospho-D-glyceroyl phosphate"
"2pg"	"VMH"	"D-Glycerate 2-phosphate"
"3pg"	"VMH"	"3-Phospho-D-glycerate"
"6pgc"	"VMH"	"6-Phospho-D-gluconate"
"6pgl"	"VMH"	"6-phospho-D-glucono-1,5-lactone"
"ac"	"inchi smiles VMH"	"Acetate"
"acald"	"inchi smiles KEGG HMDB PubChem CHEBI VMH"	"Acetaldehyde"
"accoa"	"VMH"	"Acetyl-CoA"
"adp"	"VMH"	"ADP"
"akg"	"inchi smiles VMH"	"2-Oxoglutarate"
"amp"	"VMH"	"AMP"
"atp"	"VMH"	"ATP"
"cit"	"inchi smiles VMH"	"Citrate"
"co2"	"inchi smiles KEGG HMDB PubChem CHEBI VMH"	"CO2"
"coa"	"VMH"	"Coenzyme A"
"dhap"	"inchi smiles VMH"	"Dihydroxyacetone phosphate"
"e4p"	"VMH"	"D-Erythrose 4-phosphate"
"etoh"	"inchi smiles KEGG HMDB PubChem CHEBI VMH"	"Ethanol"
"f6p"	"VMH"	"D-Fructose 6-phosphate"
"fdp"	"VMH"	"D-Fructose 1,6-bisphosphate"
"for"	"inchi smiles VMH"	"Formate"
"fru"	"KEGG HMDB PubChem CHEBI VMH"	"D-Fructose"
"fum"	"VMH"	"Fumarate"
"g3p"	"VMH"	"Glyceraldehyde 3-phosphate"
"g6p"	"VMH"	"D-Glucose 6-phosphate"
"glc_D"	"KEGG PubChem CHEBI VMH"	"D-Glucose"
"gln_L"	"KEGG HMDB PubChem CHEBI VMH"	"L-Glutamine"
"glu_L"	"VMH"	"L-Glutamate"
"glx"	"inchi smiles VMH"	"Glyoxylate"
"h"	"inchi smiles KEGG HMDB PubChem CHEBI VMH"	"H+"
"h2o"	"inchi smiles KEGG HMDB PubChem CHEBI VMH"	"H2O"

"icit"	"inchi smiles VMH"	"Isocitrate"
"lac_D"	"VMH"	"D-Lactate"
"mal_L"	"VMH"	"L-Malate"
"nad"	"KEGG PubChem CHEBI"	"Nicotinamide adenine dinucleotide"
"nadh"	"VMH"	"Nicotinamide adenine dinucleotide - reduce
"nadp"	"KEGG PubChem CHEBI"	"Nicotinamide adenine dinucleotide phosphat
"nadph"	"VMH"	"Nicotinamide adenine dinucleotide phosphat
"nh4"	"inchi smiles VMH"	"Ammonium"
"o2"	"inchi smiles KEGG HMDB PubChem CHEBI VMH"	"O2"
"oaa"	"inchi smiles VMH"	"Oxaloacetate"
"pep"	"inchi smiles VMH"	"Phosphoenolpyruvate"
"pi"	"inchi smiles VMH"	"Phosphate"
"pyr"	"inchi smiles VMH"	"Pyruvate"
"r5p"	"VMH"	"alpha-D-Ribose 5-phosphate"
"ru5p_D"	"VMH"	"D-Ribulose 5-phosphate"
"s7p"	"VMH"	"Sedoheptulose 7-phosphate"
"succ"	"inchi smiles VMH"	"Succinate"
"succoa"	"VMH"	"Succinyl-CoA"
"xu5p_D"	"VMH"	"D-Xylulose 5-phosphate"
"actp"	"VMH"	"Acetyl phosphate"
"q8h2"	"HMDB VMH"	"Ubiquinol-8"






```
Adjusting pH based on the model's chemical formula ...

adjustedpH:
  52×13 table
```

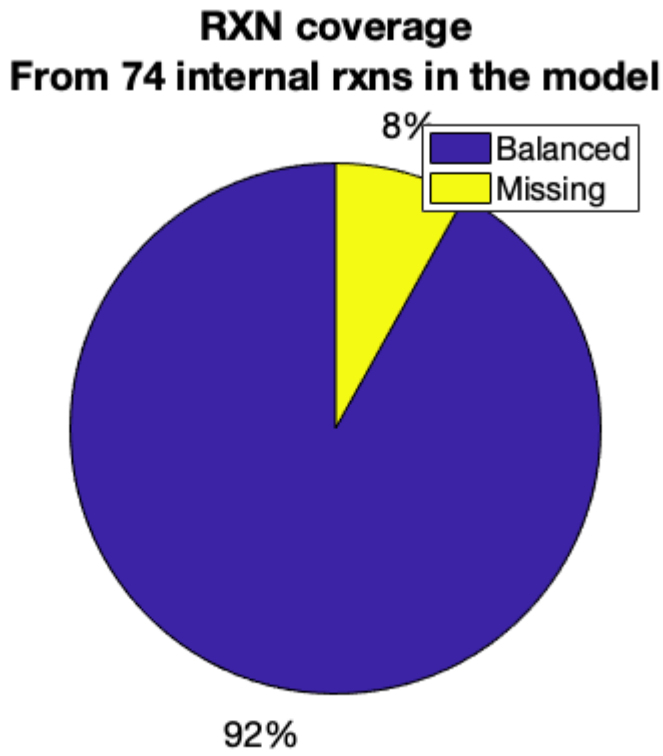
mets	sourceWithHighestScore	metNames
"13dpg"	"VMH"	"3-Phospho-D-glyceroyl phosphate"
"2pg"	"VMH"	"D-Glycerate 2-phosphate"
"3pg"	"VMH"	"3-Phospho-D-glycerate"
"6pgc"	"VMH"	"6-Phospho-D-gluconate"
"6pgl"	"VMH"	"6-phospho-D-glucono-1,5-lactone"
"ac"	"inchi smiles VMH"	"Acetate"
"acald"	"inchi smiles KEGG HMDB PubChem CHEBI VMH"	"Acetaldehyde"
"accoa"	"VMH"	"Acetyl-CoA"
"adp"	"VMH"	"ADP"
"akg"	"inchi smiles VMH"	"2-Oxoglutarate"
"amp"	"VMH"	"AMP"
"atp"	"VMH"	"ATP"
"cit"	"inchi smiles VMH"	"Citrate"
"co2"	"inchi smiles KEGG HMDB PubChem CHEBI VMH"	"CO2"
"coa"	"VMH"	"Coenzyme A"
"dhap"	"inchi smiles VMH"	"Dihydroxyacetone phosphate"
"e4p"	"VMH"	"D-Erythrose 4-phosphate"
"etoh"	"inchi smiles KEGG HMDB PubChem CHEBI VMH"	"Ethanol"
"f6p"	"VMH"	"D-Fructose 6-phosphate"
"fdp"	"VMH"	"D-Fructose 1,6-bisphosphate"
"for"	"inchi smiles VMH"	"Formate"
"fru"	"KEGG HMDB PubChem CHEBI VMH"	"D-Fructose"
"fum"	"VMH"	"Fumarate"
"g3p"	"VMH"	"Glyceraldehyde 3-phosphate"
"g6p"	"VMH"	"D-Glucose 6-phosphate"
"glc_D"	"KEGG PubChem CHEBI VMH"	"D-Glucose"
"gln_L"	"KEGG HMDB PubChem CHEBI VMH"	"L-Glutamine"

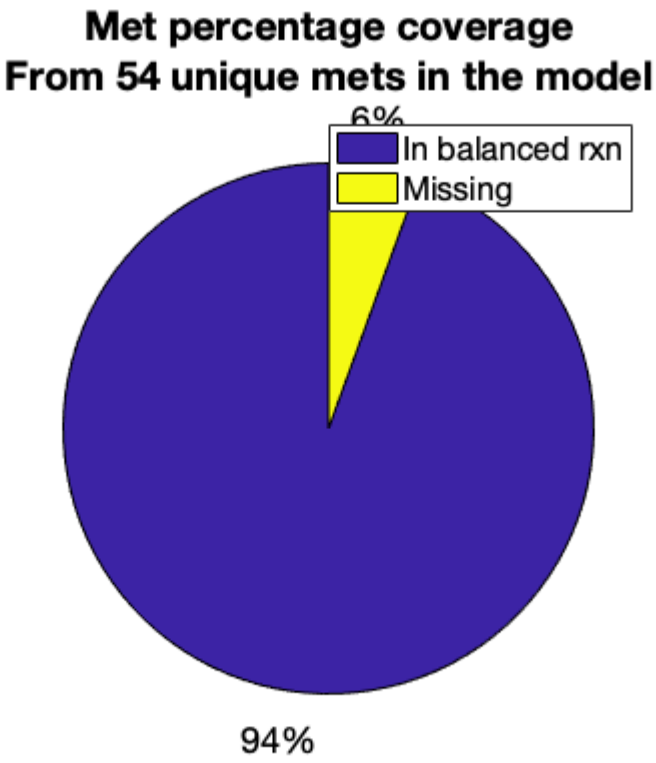
"glu_L"	"VMH"	"L-Glutamate"
"glx"	"inchi smiles VMH"	"Glyoxylate"
"h"	"inchi smiles KEGG HMDB PubChem CHEBI VMH"	"H+"
"h2o"	"inchi smiles KEGG HMDB PubChem CHEBI VMH"	"H2O"
"icit"	"inchi smiles VMH"	"Isocitrate"
"lac_D"	"VMH"	"D-Lactate"
"mal_L"	"VMH"	"L-Malate"
"nad"	"KEGG PubChem CHEBI"	"Nicotinamide adenine dinucleotide"
"nadh"	"VMH"	"Nicotinamide adenine dinucleotide - reduce
"nadp"	"KEGG PubChem CHEBI"	"Nicotinamide adenine dinucleotide phosphat
"nadph"	"VMH"	"Nicotinamide adenine dinucleotide phosphat
"nh4"	"inchi smiles VMH"	"Ammonium"
"o2"	"inchi smiles KEGG HMDB PubChem CHEBI VMH"	"O2"
"oaa"	"inchi smiles VMH"	"Oxaloacetate"
"pep"	"inchi smiles VMH"	"Phosphoenolpyruvate"
"pi"	"inchi smiles VMH"	"Phosphate"
"pyr"	"inchi smiles VMH"	"Pyruvate"
"r5p"	"VMH"	"alpha-D-Ribose 5-phosphate"
"ru5p_D"	"VMH"	"D-Ribulose 5-phosphate"
"s7p"	"VMH"	"Sedoheptulose 7-phosphate"
"succ"	"inchi smiles VMH"	"Succinate"
"succoa"	"VMH"	"Succinyl-CoA"
"xu5p_D"	"VMH"	"D-Xylulose 5-phosphate"
"actp"	"VMH"	"Acetyl phosphate"
"q8h2"	"HMDB VMH"	"Ubiquinol-8"

Standardizing 52 MOL files ...
52×4 table

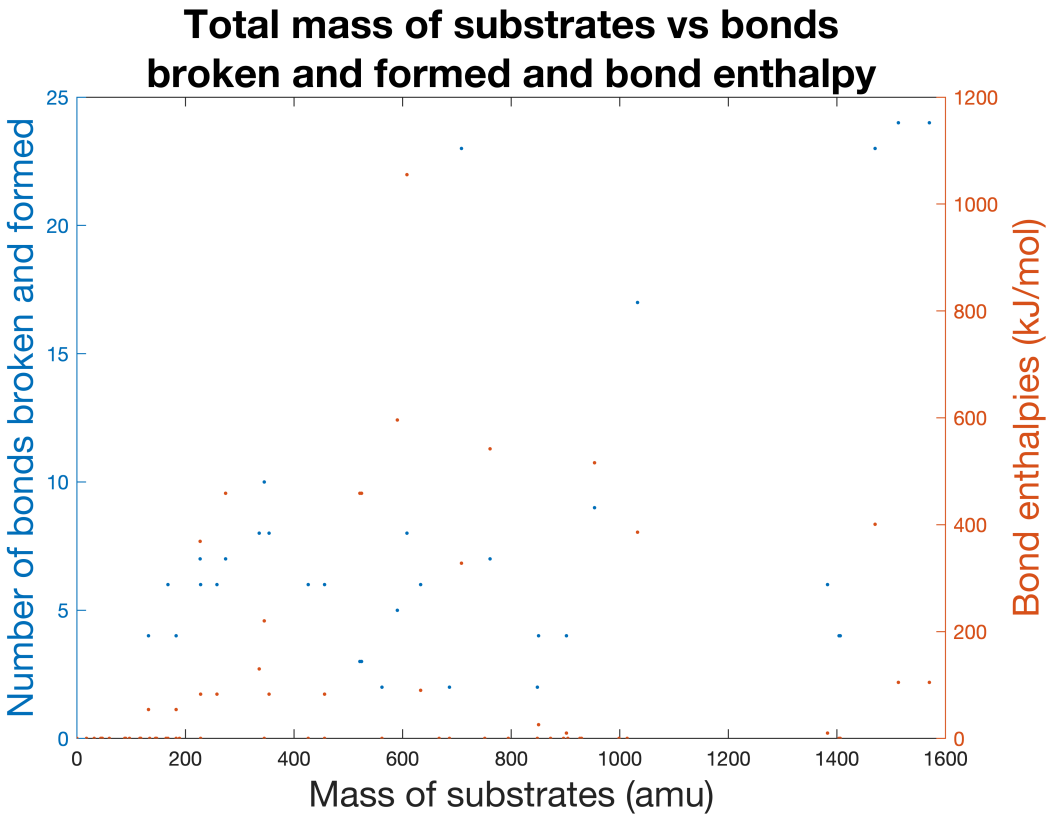
mets	InChIKeys	
"13dpg"	"InChIKey=LJQLQCAXBUHEAZ-UWTATZPHSA-J"	"InChI=1S/C3H8O10P2/c4-2(1-12-14(6,7)8)3(5)13-1
"2pg"	"InChIKey=GXIURPTVHJPJLF-UWTATZPHSA-K"	"InChI=1S/C3H7O7P/c4-1-2(3(5)6)10-11(7,8)9/h2,4
"3pg"	"InChIKey=OSJPPGNTCRNQQC-UWTATZPHSA-K"	"InChI=1S/C3H7O7P/c4-2(3(5)6)1-10-11(7,8)9/h2,4
"6pgc"	"InChIKey=BIRSGZKFKXLSJQ-SQOUGZDYSA-K"	"InChI=1S/C6H13O10P/c7-2(1-16-17(13,14)15)3(8)4
"6pgl"	"InChIKey=IJOJIVNDFQSGAB-SQOUGZDYSA-L"	"InChI=1S/C6H11O9P/c7-3-2(1-14-16(11,12)13)15-6
"ac"	"InChIKey=QTBSBXVTEAMEQO-UHFFFAOYSA-M"	"InChI=1S/C2H4O2/c1-2(3)4/h1H3,(H,3,4)/p-1"
"acald"	"InChIKey=IKHGUXGNUITLKF-UHFFFAOYSA-N"	"InChI=1S/C2H4O/c1-2-3/h2H,1H3"
"accoa"	"InChIKey=ZSLZBFCDCINBPY-ZSJPKINUSA-J"	"InChI=1S/C23H38N7O17P3S/c1-12(31)51-7-6-25-14(
"actp"	"InChIKey=LIPOUNRJVLNBCD-UHFFFAOYSA-L"	"InChI=1S/C2H5O5P/c1-2(3)7-8(4,5)6/h1H3,(H2,4,5
"adp"	"InChIKey=XTWYTFMLZFPYCI-KQYNXXCUSA-K"	"InChI=1S/C10H15N5O10P2/c11-8-5-9(13-2-12-8)15(
"akg"	"InChIKey=KPGXRSRHYNQIFN-UHFFFAOYSA-L"	"InChI=1S/C5H6O5/c6-3(5(9)10)1-2-4(7)8/h1-2H2,(
"amp"	"InChIKey=UDMBCSSLTHHNCD-KQYNXXCUSA-L"	"InChI=1S/C10H14N5O7P/c11-8-5-9(13-2-12-8)15(3-
"atp"	"InChIKey=ZKHQWZAMYRWXGA-KQYNXXCUSA-J"	"InChI=1S/C10H16N5O13P3/c11-8-5-9(13-2-12-8)15(
"cit"	"InChIKey=KRKNYBCHXYNGOX-UHFFFAOYSA-K"	"InChI=1S/C6H8O7/c7-3(8)1-6(13,5(11)12)2-4(9)10
"co2"	"InChIKey=CURLTUGMZLYLDI-UHFFFAOYSA-N"	"InChI=1S/CO2/c2-1-3"
"coa"	"InChIKey=RGJOEKWQDUBAIZ-IBOSZNHHSA-J"	"InChI=1S/C21H36N7O16P3S/c1-21(2,16(31)19(32)24
"dhap"	"InChIKey=GNGACRATGGDKBX-UHFFFAOYSA-L"	"InChI=1S/C3H7O6P/c4-1-3(5)2-9-10(6,7)8/h4H,1-2
"e4p"	"InChIKey=NGHMDNPXVRFFGS-IUYQGCFVSA-L"	"InChI=1S/C4H9O7P/c5-1-3(6)4(7)2-11-12(8,9)10/h
"etoh"	"InChIKey=LFQSCWFLJHTTHZ-UHFFFAOYSA-N"	"InChI=1S/C2H6O/c1-2-3/h3H,2H2,1H3"
"f6p"	"InChIKey=GSXOAOHZAIYLCY-HSUXUTPPSA-L"	"InChI=1S/C6H13O9P/c7-1-3(8)5(10)6(11)4(9)2-15-
"fdp"	"InChIKey=RNBGYGVWRKECFJ-VRPWFDPXSA-J"	"InChI=1S/C6H14O12P2/c7-4-3(1-16-19(10,11)12)18
"for"	"InChIKey=BDAGIHXWWSANSR-UHFFFAOYSA-M"	"InChI=1S/CH2O2/c2-1-3/h1H,(H,2,3)/p-1"
"fru"	"InChIKey=RFSUNEUAIZKAJO-ARQDHWQXSA-N"	"InChI=1S/C6H12O6/c7-1-3-4(9)5(10)6(11,2-8)12-3
"fum"	"InChIKey=VZCYOOQTPOCHFL-OWOJBTEDSA-L"	"InChI=1S/C4H4O4/c5-3(6)1-2-4(7)8/h1-2H,(H,5,6)
"g3p"	"InChIKey=LXJXRIRHZLFYRP-VKHMYHEASA-L"	"InChI=1S/C3H7O6P/c4-1-3(5)2-9-10(6,7)8/h1,3,5H
"g6p"	"InChIKey=NBSCHQHZLSJFNQ-GASJEMHNSA-L"	"InChI=1S/C6H13O9P/c7-3-2(1-14-16(11,12)13)15-6
"glc_D"	"InChIKey=WQZGKKKJIJFFOK-GASJEMHNSA-N"	"InChI=1S/C6H12O6/c7-1-2-3(8)4(9)5(10)6(11)12-2
"gln_L"	"InChIKey=ZDXPYRJPNDTMRX-VKHMYHEASA-N"	"InChI=1S/C5H10N2O3/c6-3(5(9)10)1-2-4(7)8/h3H,1
"glu_L"	"InChIKey=WHUUTDBJXJRKMK-VKHMYHEASA-M"	"InChI=1S/C5H9NO4/c6-3(5(9)10)1-2-4(7)8/h3H,1-2
"glx"	"InChIKey=HHLFWLYXYJOTON-UHFFFAOYSA-M"	"InChI=1S/C2H2O3/c3-1-2(4)5/h1H,(H,4,5)/p-1"
"h"	"InChIKey=GPRLSGONYQIRFK-UHFFFAOYSA-N"	"InChI=1S/p+1"
"h2o"	"InChIKey=XLYOFNOQVPJJNP-UHFFFAOYSA-N"	"InChI=1S/H2O/h1H2"

"icit"	"InChIKey=ODBLHEXUDAPZAU-UHFFFAOYSA-K"	"InChI=1S/C6H8O7/c7-3(8)1-2(5(10)11)4(9)6(12)13
"lac_D"	"InChIKey=JVTAAEKCZFNVCJ-REOHCLBHSA-M"	"InChI=1S/C3H6O3/c1-2(4)3(5)6/h2,4H,1H3,(H,5,6)
"mal_L"	"InChIKey=BJEPYKJPYRNKOW-REOHCLBHSA-L"	"InChI=1S/C4H6O5/c5-2(4(8)9)1-3(6)7/h2,5H,1H2,(
"nad"	"InChIKey=BAWFJGJZGIEFAR-NNYOXOHSSA-M"	"InChI=1S/C21H27N7O14P2/c22-17-12-19(25-7-24-17
"nadh"	"InChIKey=BOPGDPNILDQYTO-NNYOXOHSSA-L"	"InChI=1S/C21H29N7O14P2/c22-17-12-19(25-7-24-17
"nadp"	"InChIKey=XJLXINKUBYWONI-NNYOXOHSSA-K"	"InChI=1S/C21H28N7O17P3/c22-17-12-19(25-7-24-17
"nadph"	"InChIKey=ACFIXJIJDZMPPO-NNYOXOHSSA-J"	"InChI=1S/C21H30N7O17P3/c22-17-12-19(25-7-24-17
"nh4"	"InChIKey=QGZKDVFQNNGYKY-UHFFFAOYSA-O"	"InChI=1S/H3N/h1H3/p+1"
"o2"	"InChIKey=MYMOFIZGZYHOMD-UHFFFAOYSA-N"	"InChI=1S/O2/c1-2"
"oaa"	"InChIKey=KHPXUQMNIQBQEV-UHFFFAOYSA-L"	"InChI=1S/C4H4O5/c5-2(4(8)9)1-3(6)7/h1H2,(H,6,7
"pep"	"InChIKey=DTBNBXWJWCWCIK-UHFFFAOYSA-K"	"InChI=1S/C3H5O6P/c1-2(3(4)5)9-10(6,7)8/h1H2,(H
"pi"	"InChIKey=NBIIXXVUZAFLBC-UHFFFAOYSA-L"	"InChI=1S/H3O4P/c1-5(2,3)4/h(H3,1,2,3,4)/p-2"
"pyr"	"InChIKey=LCTONWCANYUPML-UHFFFAOYSA-M"	"InChI=1S/C3H4O3/c1-2(4)3(5)6/h1H3,(H,5,6)/p-1"
"q8h2"	"InChIKey=LOJUQFSPYHMHEO-SGHXUWJISA-N"	"InChI=1S/C49H76O4/c1-36(2)20-13-21-37(3)22-14-






```
Obtaining RInChIes and reaction SMILES ...
Found biomass reaction: Biomass_Ecoli_core_N(w/GAM)-Nmet2
ATP maintenance reaction is not considered an exchange reaction by default. It should be mass balanced:
ATPM atp[c] + h2o[c]  -> adp[c] + h[c] + pi[c]
```




```
95×4 table
```

rxns	rxnNames
'ACALD'	'acetaldehyde dehydrogenase (acetylating)'
'ACALDt'	'acetaldehyde reversible transport'
'ACKr'	'acetate kinase'
'ACONTa'	'aconitase (half-reaction A, Citrate hydro-lyase)'
'ACONTb'	'aconitase (half-reaction B, Isocitrate hydro-lyase)'
'ACt2r'	'acetate reversible transport via proton symport'
'ADK1'	'adenylate kinase'
'AKGDH'	' 2-Oxogluterate dehydrogenase'
'AKGt2r'	' 2-oxoglutarate reversible transport via symport'
'ALCD2x'	'alcohol dehydrogenase (ethanol)'
'ATPM'	'ATP maintenance requirement'
'ATPS4r'	'ATP synthase (four protons for one ATP)'
'Biomass_Ecoli_core_N(w/GAM)-Nmet2'	'core E. coli biomass equation (Neidhardt Based with GAM, N me
'CO2t'	'CO2 transporter via diffusion'
'CS'	'citrate synthase'
'CYTBD'	'cytochrome oxidase bd (ubiquinol-8: 2 protons)'
'D-LACt2'	'D-lactate transport via proton symport'
'ENO'	'enolase'
'ETOHt2r'	'ethanol reversible transport via proton symport'
'EX_ac(e)'	'Acetate exchange'
'EX_acald(e)'	'Acetaldehyde exchange'
'EX_akg(e)'	' 2-Oxoglutarate exchange'
'EX_co2(e)'	'CO2 exchange'
'EX_etoh(e)'	'Ethanol exchange'
'EX_for(e)'	'Formate exchange'
'EX_fru(e)'	'D-Fructose exchange'
'EX_fum(e)'	'Fumarate exchange'
'EX_glc(e)'	'D-Glucose exchange'
'EX_gln-L(e)'	'L-Glutamine exchange'
'EX_glu-L(e)'	'L-Glutamate exchange'
'EX_h2o(e)'	'H2O exchange'
'EX_h(e)'	'H+ exchange'
'EX_lac-D(e)'	'D-lactate exchange'
'EX_mal-L(e)'	'L-Malate exchange'
'EX_nh4(e)'	'Ammonia exchange'
'EX_o2(e)'	'O2 exchange'
'EX_pi(e)'	'Phosphate exchange'
'EX_pyr(e)'	'Pyruvate exchange'
'EX_succ(e)'	'Succinate exchange'
'FBA'	'fructose-bisphosphate aldolase'
'FBP'	'fructose-bisphosphatase'
'FORt2'	'formate transport in via proton symport'
'FORti'	'formate transport via diffusion'
'FRD7'	'fumarate reductase'
'FRUpts2'	'Fructose transport via PEP:Pyr PTS (f6p generating)'
'FUM'	'fumarase'
'FUMt2_2'	'Fumarate transport via proton symport (2 H)'
'G6PDH2r'	'glucose 6-phosphate dehydrogenase'
'GAPD'	'glyceraldehyde-3-phosphate dehydrogenase'
'GLCpts'	'D-glucose transport via PEP:Pyr PTS'
'GLNS'	'glutamine synthetase'
'GLNabc'	'L-glutamine transport via ABC system'
'GLUDy'	'glutamate dehydrogenase (NADP)'
'GLUN'	'glutaminase'
'GLUSy'	'glutamate synthase (NADPH)'
'GLUt2r'	'L-glutamate transport via proton symport, reversible'
'GND'	'phosphogluconate dehydrogenase'
'H2Ot'	'H2O transport via diffusion'
'ICDHyr'	'isocitrate dehydrogenase (NADP)'
'ICL'	'Isocitrate lyase'
'LDH_D'	'D-lactate dehydrogenase'

```
    'MALS'                                  'malate synthase'
    'MALt2_2'                               'Malate transport via proton symport (2 H)'
    'MDH'                                   'malate dehydrogenase'
    'ME1'                                   'malic enzyme (NAD)'
    'ME2'                                   'malic enzyme (NADP)'
    'NADH16'                                'NADH dehydrogenase (ubiquinone-8 & 3 protons)'
    'NADTRHD'                               'NAD transhydrogenase'
    'NH4t'                                  'ammonia reversible transport'
    'O2t'                                   'o2 transport (diffusion)'
    'PDH'                                   'pyruvate dehydrogenase'
    'PFK'                                   'phosphofructokinase'
    'PFL'                                   'pyruvate formate lyase'
    'PGI'                                   'glucose-6-phosphate isomerase'
    'PGK'                                   'phosphoglycerate kinase'
    'PGL'                                   ' 6-phosphogluconolactonase'
    'PGM'                                   'phosphoglycerate mutase'
    'PIt2r'                                 'phosphate reversible transport via symport'
    'PPC'                                   'phosphoenolpyruvate carboxylase'
    'PPCK'                                  'phosphoenolpyruvate carboxykinase'
    'PPS'                                   'phosphoenolpyruvate synthase'
    'PTAr'                                  'phosphotransacetylase'
    'PYK'                                   'pyruvate kinase'
    'PYRt2r'                                'pyruvate reversible transport via proton symport'
    'RPE'                                   'ribulose 5-phosphate 3-epimerase'
    'RPI'                                   'ribose-5-phosphate isomerase'
    'SUCCt2_2'                              'succinate transport via proton symport (2 H)'
    'SUCCt3'                                'succinate transport out via proton antiport'
    'SUCDi'                                 'succinate dehydrogenase (irreversible)'
    'SUCOAS'                                'succinyl-CoA synthetase (ADP-forming)'
    'TALA'                                  'transaldolase'
    'THD2'                                  'NAD(P) transhydrogenase'
    'TKT1'                                  'transketolase'
    'TKT2'                                  'transketolase'
    'TPI'                                   'triose-phosphate isomerase'

Diary written to: /Users/gpreciat/work/code/ctf/databases/ecoliDB
generateChemicalDatabase run is complete.
```

Bibliography

1. Noronha et al., "The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease", *Nucleic acids research* **(2018).**
2. Wishart et al., "HMDB 4.0 — The Human Metabolome Database for 2018*". Nucleic acids research* **(2018).**
3. Sunghwan et al. “PubChem in 2021: new data content and improved web interfaces.” *Nucleic acids research* **(2021).**
4. Kanehisa, and Goto. "KEGG: Kyoto Encyclopedia of Genes and Genomes". *Nucleic acids research* **(2000).**
5. Hastings et al,. "ChEBI in 2016: Improved services and an expanding collection of metabolites". *Nucleic acids research* **(2016).**
6. O'Boyle et al,. "Open Babel: An open chemical toolbox." *Journal of Cheminformatics* **(2011).**
7. "Marvin was used for drawing, displaying and characterizing chemical structures, substructures and reactions, ChemAxon (http://www.chemaxon.com)"
8. Rahman et al,. "Reaction Decoder Tool (RDT): Extracting Features from Chemical Reactions", Bioinformatics **(2016).**