AutoPAD Documentation

Working environment: MATLAB
Additional requirements: Cobra Toolbox
Built in: MATLAB 2015a

Authors:
V 1.0 - Magdalena Ribbeck

Introduction

AutoPAD is an automated tool coded in MATLAB for pH adjustment of Genome-Scale Metabolic Models (GSMM). Starting from a charge- and mass-balanced GSMM at a defined pH, AutoPAD modifies the protonation state of the metabolites and appropriately rebalances all the reactions in the model, consistent with a defined pH condition. The latter enables a more realistic representation of the cellular phenotype, overall improving the prediction power of GSMMs.

Syntax

From the Command window in MATLAB, AutoPAD can be called using the following syntax:

```
model_adjusted = autoPAD(model,pHset,pHstart,pKa)
```

Inputs	Description
`model`	COBRA model structure. Requires fields `metFormula` and `metCharges`. Field `comps` (compartments) is desired
`pHset`	list of new pH values for each compartment (one per compartment)
`pHstart`	list of initial pH values for each compartment (one per compartment)
`pKa`	Numeric array. Contains a row-wise list of the pKa values for each metabolite in the model.

Optional Inputs	Description
`modelName`	Name of the COBRA model structure set after adjusting the model (default: `adjustedModel`)
`reportFormat`	Determines the format of the printed document. Options: `0` for disabling the printing option, `1` for an excel file, `2` for a text file archive (default: `1`)
`directionBool`	Boolean array that indicates the transport reactions directed on the opposite direction as defined in the model. Valid for reversible reactions (default: none is modified)

Outputs	Description
`model_adjusted`	pH-adjusted COBRA model structure to the pH values defined in `pHset` defined by the user

AutoPAD algorithm

AutoPAD executes the following two-step procedure for adjusting the pH of GSMMs:

1. Starting from an initial charge- and mass-balanced GSMM defined at a known pH for the internal and external compartments, AutoPAD adjusts the chemical formulae and charges of all metabolites consistent with the set pHs. In this step, a simple algorithm is employed to determine the most abundant protonation state for each metabolite using pKa information estimated with Marvin 18.1 (ChemAxon, http://www.chemaxon.com) for 5,358 commonly-used metabolites.

   Determination of the most abundant protonation state
   Firstly, AutoPAD sorts the list of pKa values associated to a specific metabolite from higher to lower. Then, the algorithm assigns a pKa value from the list to the original pH. The criterion employed for this task consists on choosing the value that is closest to new pH and that is also higher than original pH. The same is done for the set pH. If the same pKa value is selected for both pH values, AutoPAD considers that there is no difference between the protonation state of the metabolite in the two conditions. If there is a nonzero difference, the distance between the pKa values - i.e., the amount of pKa values that are in between the two values – indicates the number of protons lost or gained because of the pH change. One negative unit change indicates dissociation of one proton, whereas a positive unit change represents a proton gain.

2. Once all metabolites have been adjusted to the appropriate pH, the second step involves checking and rebalancing all the reactions in the GSMM that may not be charge- and/or mass-balanced as a consequence of the previous protonation adjustment. As these inconsistencies are related to the addition/removal of hydrogen ions, reactions can be readily balanced by checking and correcting (if necessary) their hydrogen elemental balance. The following steps are carried out in order to rebalance a reaction:

   Ensuring hydrogen ions availability
   Before rebalancing an equation, AutoPAD must verify that there is a source of hydrogen ions for each compartment in the model. This step is critical as it will later enable to rebalance the reactions using protons from the appropriate compartment. For this task, AutoPAD first determined the number of compartments in the model by looking up the `comps` field in the model structure. If the model does not include the `comps` field, AutoPAD creates the compartment list by searching the compartments specified for each metabolite in the `mets` field of the model structure. This assumes that the compartments are indicated in a parenthesis (e.g., glucose[c]) or underscore (e.g., glucose_c) formats. Once the `comps` field is available, AutoPAD searches for metabolites in the `mets` field that represent hydrogen ions and associate each of them to the appropriate compartment. The following names are assumed for identifying the protons: H(+), H+, H, h, h(+) or h+. If AutoPAD is unable to find a

hydrogen ion for a compartment, it will automatically create it under the name “h”, followed by the corresponding compartment identifier. If AutoPAD is unable to find any hydrogen ion, or if it finds more candidates than compartments, this step will fail. The result of this step is a model with hydrogen ions associated to each compartment.

Reaction-compartment associations

To determine which hydrogen ion should be used to rebalance a reaction, the reaction itself has to be assigned to a compartment. This is trivial when all the metabolites that participate in a reaction belong to the same compartment. In such case, the hydrogen ions used for rebalancing obviously come from the same compartment of the reaction. Nevertheless, this is not trivial for transport reactions as very often different compartments display different pHs. In these cases, the transport mechanism plays a critical role on how the reaction should be rebalanced. In order to address these cases, different translocation systems were analyzed and a rule was determined to assign a reaction to a compartment. For symport, uniport or diffusion transport, as well as for translocation reactions in which one substrate belongs to a different compartment (e.g., ABC transport, PTS transport, among others), the reaction was assigned to the compartment of the products. For reactions in which one product belongs to a different compartment (e.g., membrane bound reactions), the reaction was assigned to the compartment of the substrates. In reactions in which both substrates and products belong to different compartments (e.g. antiport mechanisms), the reaction was assigned to the compartment with lower pH (i.e., higher proton abundance). The latter works as rule of thumb, as the balance for these reactions is highly dependent on the concentrations of the participating metabolites, and thus, one cannot generalize. Figure 1 illustrates some examples of these translocation reactions and the consumption and production of protons. Additionally, some common shuttle translocation systems were also tested. As their mechanism can be divided into a set of reactions that fit within the previous classifications, AutoPAD is also able to appropriately balance these reactions.

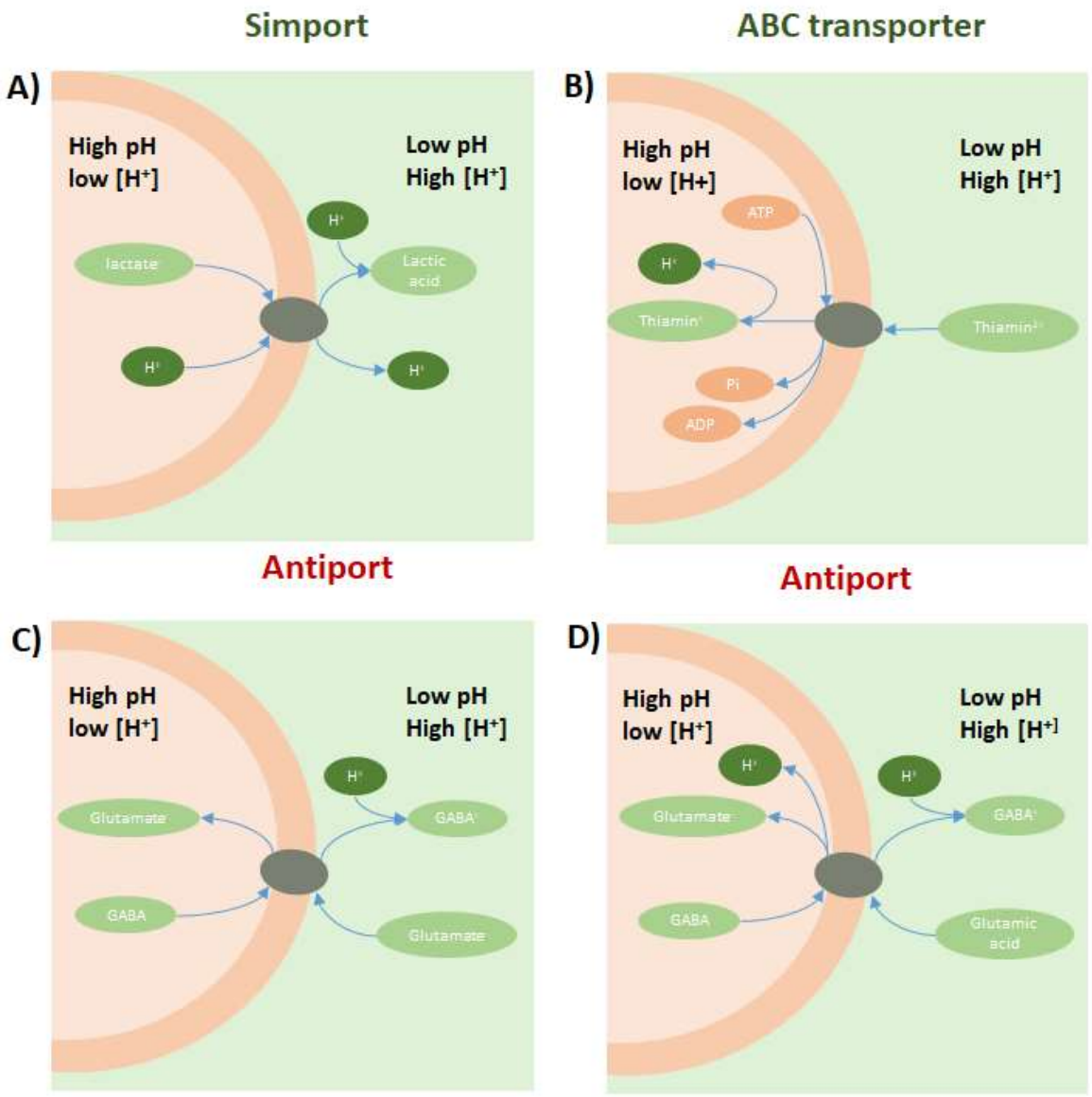

Figure 1. Proton transport depends on the translocation mechanisms and pH conditions
This figure exemplifies how the pH difference between compartments (e.g., cytoplasmic and extracellular) affects the dissociation state of weak acids/bases compounds, which in turn results in the import/release of a proton(s) from/into the environment. This import or release depends on: the type of transport, its direction (import or export), the pKa of the metabolites, and the acidity of each compartment. Examples are shown in an acidic environment. **A)** *Lactobacillus plantarum's* simport reaction for lactate export. If external pH is lower than 3.8, extracellular lactate exported from the cytoplasm turns into lactic acid, and thus gains a proton from the extracellular environment. **B)** *Oenococcus oeni*'s ABC transporter for thiamin uptake. If external pH is lower than 5.5, thiamin^{2+} loses a proton when transported into the internal environment with higher pH. **C)** *Oenococcus oeni*'s GABA-Glutamate antiport. If external pH is between 4.3 and 4.5 (pKa of glutamate and GABA, respectively), glutamate does not change its dissociation state but GABA does, and thus a proton is taken from the external environment. **D)** *Oenococcus oeni*'s GABA-Glutamate antiport. If external pH is lower than 4.3, glutamate is then present as glutamic acid, and thus it dissociates as it enters the cell.

Considerations

Metabolites

- AutoPAD assigns to each metabolite the chemical formula of the most abundant protonation state at a specific pH.
- AutoPAD considers that if a chemical formula is present in the given model, this is the correct protonation state of the metabolite at the initial pH. **AutoPAD will not check if the chemical formula is correct at the given pH.**
- When AutoPAD adjusts a metabolite, it modifies its chemical formula and charge. For this task, AutoPAD separates the string that contains the chemical formula present in the `metFormulas` field of the model structure into a matrix of *n* x 25, where *n* is the quantity of metabolites in the model and 25 indicates the following elements:

  ```
  Elements =
  {'H','C','O','P','S','N','Mg','X','Fe','Zn','Co','R','K','Cl','Cd','
  Na','Ni','Mn','Cu','Ca','Y','I','F','Ag','FULLR'}
  ```

  The latter represent the most common elements in metabolic models. If a metabolite formula does not contain any of these elements, the formula is maintained as the original. **WARNING:** If you have a metabolite with a chemical formula that contains some of these elements but also others that are not listed, you should manually add them to the `Elements` vector listed above. Otherwise, when the chemical formula is reassigned to the metabolite after the pH adjustment, it will lack the said element.

Reactions:

- AutoPAD will check and correct the proton imbalance of reactions, with the exception of: exchange reactions, reactions that contain metabolites that lack chemical formulas, and reactions that display mass imbalances not related to hydrogen ions.
- Reactions in which both substrates and products belong to different compartments (e.g., antiport reactions) depend on the pH of both compartments and on the pKa of the metabolites, and thus it is recommended that they are manually checked and – if necessary – curated.
- Translocation reactions are rebalanced assuming they take place from the left- to the right-hand side of the stoichiometric reaction. If the reaction occurs in the opposite direction, it will affect the balance as the proton flux should act on the opposite direction (see Figure 2). Reverse directionalities can be indicated using the `DirectionBool` by setting to `1` the position of the reaction in the reverse direction.

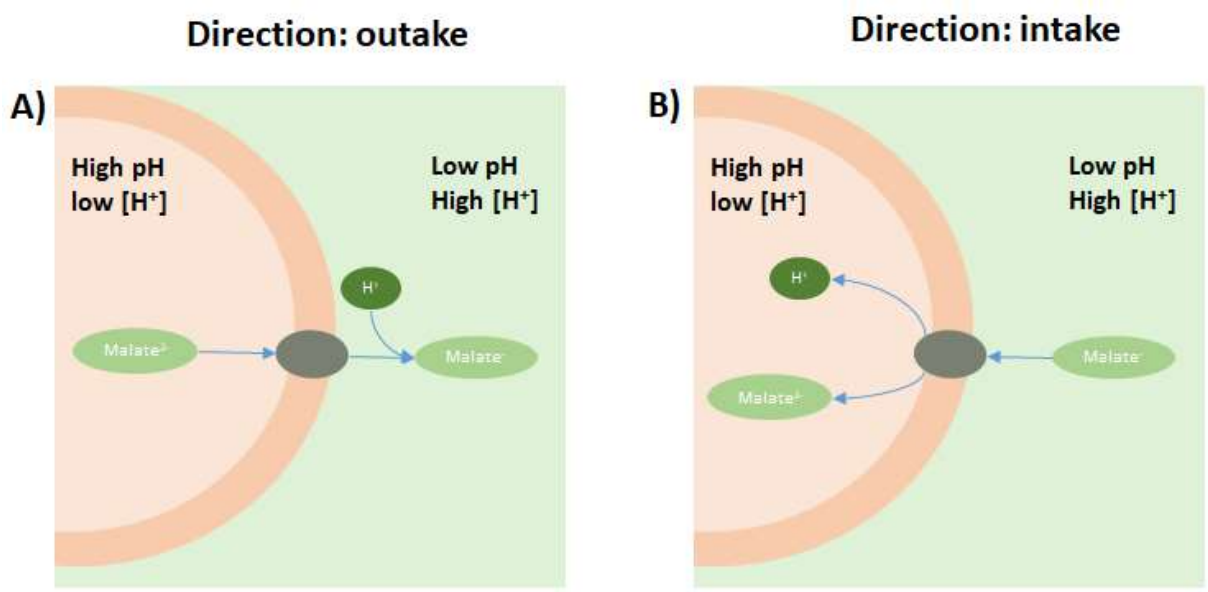

Figure 2. Effect of directionality in the final balance of a transport reaction.
This figure illustrates the effect of the directionality on the malate uptake reaction in *Oenococcus oeni* at identical pH conditions. In this example, the transport of malate is mainly dictated by malate concentration gradient. **A)** When malate is exported from the cell, the difference in the malate dissociation states causes an extracellular proton to be incorporated to $malate^{2-}$ because of the external pH. **B)** When malate is imported into the cell, $malate^{-}$ dissociates in the cytoplasm and consequently a proton is released.

Additional material and scripts

To facilitate the application of AutoPAD to any model, the following scripts and material are included:

Metabolite pKa list
This corresponds to a list of pKa values for 5,358 metabolites. This list includes the following fields: Name, KeggID, ChEBI ID, PubChem ID, SMILES, chemical formula and charge (as reported in the KEGG Database).

constructpKaTableForAModel
MATLAB script that generates a model-specific pKa table based on the aforementioned metabolite pKa list. For this function to work, the COBRA model structure has to include the `KeggID` field.

obtainChebiKeggIDForABiGGModel
MATLAB script that constructs or completes the `KeggID` and `ChEBIID` fields of a COBRA model structure. This function requires a model that in the `mets` field possesses the BiGG ID of the metabolite, which is used to search the ID in the BiGG Database.

compareTwoGSMM

MATLAB script that compares the stoichiometric coefficients of two GSMMs that contain the same reactions. This is an easy way to determine the variation between the original model and the pH-adjusted version of the same model.

Examples of use

Here we will illustrate the application of AutoPAD starting from a model obtained from the BiGG Database, the *Escherichia coli* K-12 model iML1515[1]. This model does not possess the metKeggID vector, which is needed to build the pKa table, and thus the first step corresponds to obtaining these values. A model that already possess this vector may not need to perform the first step. The supplementary MATLAB files required to follow this example are contained in a zip file called `autoPAD example`, which is available along with the AutoPAD package.

Association of the vector metKeggID to the model

To obtain the `metKeggID` vector for `iML1515`, the function `obtainChebiKeggIDForABiGGModel` is used. This function assigns a Kegg ID to each metabolite. The following commands perform the aforementioned tasks:

```
load('iML1515');

%get metKeggID to build the pKa table
iML1515_m=obtainChebiKeggIDForABiGGModel(iML1515,'iML1515_m', 'iML1515');
```

This step takes around 30 min for iML1515, although it mostly depends on the size of the model. The state of progress of the above function can be seen in the console. An example of this output is given in Figure 3.

[1] Monk, J.M., *et al.* iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat Biotechnol* 2017;35:904.

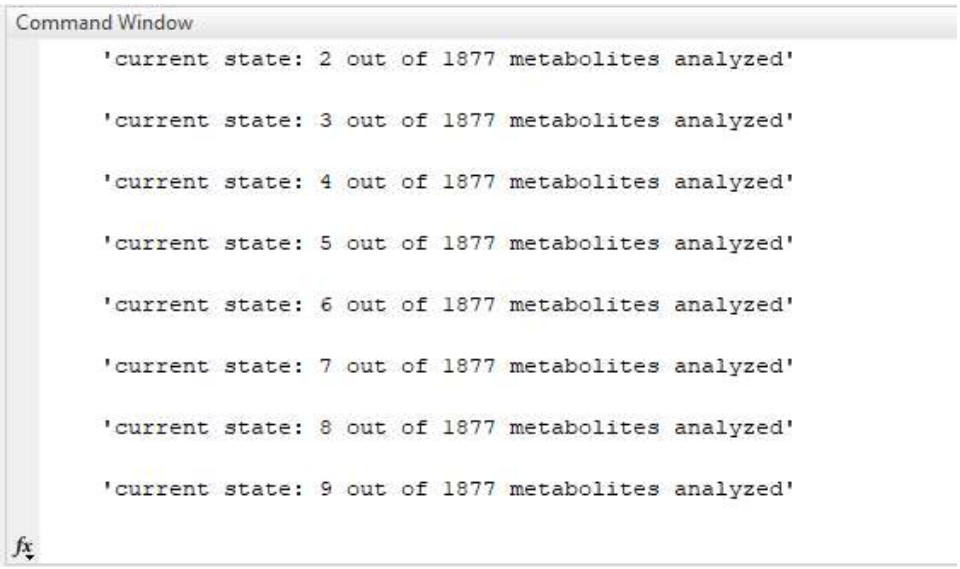

Figure 3. Progress information during the obtention of the Kegg IDs.

The result of the above function is a new .mat file named `iML1515_m` containing a model structure variable called `iML1515`. The model is kept as a variable in the MATLAB environment.

Construction of the pKa table

The pKa table can be built now using the function `constructpKaTableForAModel` as follows:

```
%get pKa table for iML1515 from metKeggIDs
[pKaTable,report]=constructpKaTableForAModel(iML1515_m,'iML1515 pKa table',1);
```

This generates an xls file named `iML1515 pKa table` that contains the pKa list and a report of the metabolites mapped onto the AutoPAD knowledge base. The latter outputs are also given as variables. An extract of the pKa table is shown in Figure 4.

1877x8 double

	1	2	3	4	5	6	7	8
1	NaN	NaN	NaN	NaN	NaN	6	3.4800	0.8100
2	NaN	NaN	NaN	11.9600	6.6400	6.0400	1.6100	1.0100
3	NaN	12.9000	12.2800	6.5500	5.9500	1.6400	1.1500	0.6500
4	NaN	NaN	NaN	NaN	NaN	7.4200	3.2000	1.7700
5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
6	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
7	NaN	NaN	NaN	NaN	NaN	NaN	13.9900	0.3100
8	NaN	NaN	NaN	NaN	NaN	NaN	4.4100	3.5500
9	NaN	NaN	NaN	NaN	NaN	9.7900	9.1900	2
10	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
11	NaN	NaN	NaN	NaN	NaN	9.3100	4.3700	2.2900

Figure 4. Extract of the pKa table generated for iML1515.
Each row corresponds to the pKa information of the ith metabolite in the model.

Generation of a new model using AutoPAD

Once the pKa table is constructed, AutoPAD can now be applied to generate the new pH-adjusted model. As indicated in Monk et al., the metabolites in iML1515 are present in the dissociation state that is the most abundant at pH 7.2, so this value is used as reference for all the compartments. For illustration purposes, the new model is adjusted to the following pH values: external pH, 5.5; cytoplasmic pH, 7.6; periplasmic pH, 6. We note that, as `iML1515_m` does not possess a `comps` vector with model compartments, this will be automatically generated by AutoPAD as shown in the MATLAB console below:

```
Command Window
> In autoPAD>assignCompartments (line 328)
  In autoPAD (line 43)
  In iML1515test (line 13)

comps =

    'c'
    'e'
    'p'
```

Figure 5. Extract of the information printed by AutoPAD when determining the compartments of iML1515.

The above figure also indicates the order of the `comps` vector, in this case: [`'c', 'e', 'p'`]. The same order is then used to write the pH values (see below).

```
%create a new model, adjusted to external pH 5.5, cytoplasmic pH 7.6,
%periplasmic pH 6.
adjusted_iML1515_m=autoPAD(iML1515_m,[7.6 5.5 6],[7.2 7.2 7.2], pKaTable,'adjusted_iML1515_m');
```

Finally, the output of AutoPAD are a report of the modified reactions, and a .mat file containing the new model named `adjusted_iML1515_m`. The AutoPAD report is shown in figure 6.

6-Jun-18			
adjusted model = adjusted_iML1515_m			
PH VALUES			
compartm	c	e	p
pH values	7.6	5.5	6
REACTIONS			
2709 out of 2712 reactions were analyzed and balanced if needed			
the following 3 reactions were not, due to lack of chemical formula of one of the metabolites that participate in the reaction,			
or due to complex mass imbalances:			
PUACGAMS			
BIOMASS_Ec_iML1515_core_75p37M			
BIOMASS_Ec_iML1515_WT_75p37M			

Figure 6. Report generated by AutoPAD for the generation of a pH-adjusted version of the iML1515 model.

Tracking stoichiometric modifications due to the pH

To facilitate determining the differences between the original and the pH-adjusted model, the function `compareTwoGSMM` can be employed. This function generates an xls file named in this case `'iML1515 comparison'`, which shows the name of the modified reactions, and the respective stoichiometric equations.

```
%compare the changes in both models
compareTwoGSMM(iML1515_m,adjusted_iML1515_m,1,'iML1515 comparison');
```

In this example, 296 reactions required modification. Figure 7 shows an extract of the xls file.

	A	B	C
1	name of the reaction	model1	model2
2	DMATT	Isopentenyl diphosp	H+ + Isopentenyl diphosphate + Dimethylallyl diphosphate -> Geranyl diphosphate + Diphosphate
3	GRTT	Geranyl diphosphate	Geranyl diphosphate + H+ + Isopentenyl diphosphate -> Diphosphate + Farnesyl diphosphate
4	GLUTRS	ATP C10H12N5O13P3	H+ + ATP C10H12N5O13P3 + L-Glutamate + TRNA (Glu) -> L-Glutamyl-tRNA(Glu) + Diphosphate + AMP C10H12N5O7P
5	TMPPP	H+ + 2-Methyl-4-ami	2 H+ + 2-Methyl-4-amino-5-hydroxymethylpyrimidine diphosphate + 4-Methyl-5-(2-phosphoethyl)-thiazole -> Diphosphate + Thiamin monophosphate
6	XPPT	Xanthine + 5-Phosph	H+ + Xanthine + 5-Phospho-alpha-D-ribose 1-diphosphate -> Diphosphate + Xanthosine 5'-phosphate
7	HXPRT	5-Phospho-alpha-D-	H+ + 5-Phospho-alpha-D-ribose 1-diphosphate + Hypoxanthine -> IMP C10H11N4O8P + Diphosphate
8	ACS	Acetate + ATP C10H1	H+ + Acetate + ATP C10H12N5O13P3 + Coenzyme A -> Acetyl-CoA + Diphosphate + AMP C10H12N5O7P
9	GLGC	H+ + ATP C10H12N5C	2 H+ + ATP C10H12N5O13P3 + D-Glucose 1-phosphate -> ADPglucose C16H23N5O15P2 + Diphosphate
10	ANPRT	Anthranilate + 5-Phc	H+ + Anthranilate + 5-Phospho-alpha-D-ribose 1-diphosphate -> N-(5-Phospho-D-ribosyl)anthranilate + Diphosphate

Figure 7. Extract of the xls file containing the differences between the models