

BiGG Models ID Specification and Guidelines

Why use BiGG IDs?

- They are human readable
- They are short and memorable
- They are well-defined and compatible with other specifications (e.g. SBML)
- They are available at <http://bigg.ucsd.edu>

Components of a BiGG ID:

For a reaction, universal metabolite, or gene, a BiGG ID has two parts, a prefix and an abbreviation. The prefix exists for compatibility with standards like SBML that disallow a number at the beginning of an identifier. Many tools (BiGG, COBRApy, Escher) only show the abbreviation, and the prefix is added for exporting to the standards where it is necessary.

[prefix]_[abbreviation]

R_GAPD

M_g3p

G_b1779

For compartmentalized metabolites, a compartment code is added:

[prefix]_[abbreviation]_[compartment code]

M_g3p_c

If there is a defined tissue, this can also be appended after an underscore. Tissue codes are differentiated from compartment codes by enforcing that tissue codes start with an *uppercase* letter and compartment codes start with a *lowercase* letter.

[prefix]_[abbreviation]_[compartment code]_[tissue code]

M_g3p_c_T1

Prefixes:

- **R**: reaction
- **M**: metabolite
- **G**: gene
- **/[RMG]/**

Abbreviations:

- Only contain upper and lowercase letters, numbers, and underscores
- `/[a-zA-Z0-9][a-zA-Z0-9_]+[a-zA-Z0-9]/`, only ASCII and do not start or end with underscore(s)
- When converting to BiGG IDs, replace a dash with two underscores. For example, `ala-L` becomes `ala__L`.
- Reactions should be all uppercase. Metabolites should be primarily lowercase, but uppercase letters are allowed (`ala__L` is preferred to `ALA__L`).

Length

Abbreviation length should be between 3 and 60 characters. Short abbreviations are preferred

Compartment codes

- One or two characters in length, and contain only lowercase letters and numbers, and must begin with a lowercase letter.
- `/[a-z][a-z0-9]?/`
- A list of compartments can be found at <http://bigg.ucsd.edu/compartments>

Tissue-type codes

- One or two characters in length, and contain only uppercase letters and numbers, and must begin with an uppercase letter.
- `/[A-Z][A-Z0-9]?/`

Regular expression

The following expression should work for reactions, metabolites, and genes:

```
/^([RMG])_([a-zA-Z][a-zA-Z0-9_]+)(?:_([a-z][a-z0-9]?))?(?:_([A-Z][A-Z0-9]?))?$
```

- group 1: prefix
- group 2: abbreviation
- group 3: compartment code or null
- group 4: tissue code or null

Boundary reactions (pseudoreactions)

The only reactions that are not required to be mass balanced are *boundary reactions*. These reactions must be in one of these categories:

1. Exchange reactions. These have a single metabolite, and their BiGG IDs must begin with **EX_** followed by the abbreviation of the metabolite that is exchanged. E.g. **EX_glc__D_e**.
2. Demand and sink reactions. These have a single metabolite, and their BiGG IDs must begin with **DM_** or **SK_** followed by the abbreviation of the metabolite that is consumed or produced. E.g. **DM_ala__L_c** or **SK_mal_c**.
3. Biomass reactions. These have many metabolites, but all stoichiometric coefficients are negative, and their BiGG IDs must match **/BIOMASS(_.)?/**.
4. ATPM is a special reaction representing the fixed non-growth associated ATP maintenance.

Uniqueness

- It is difficult to provide a precise definition of what makes a unique BiGG metabolite. For instance, there is no distinction in BiGG between metabolites based on their protonation state. However, metabolites with different R-groups are generally separate metabolites.
- Reactions are defined by their reaction stoichiometry. Thus, reactions in different models but with the same stoichiometry will always have the same BiGG abbreviation.
- Similar reactions should have similar IDs. For example, glutamate dehydrogenase using NADPH vs. glutamate dehydrogenase using NADH only differs by one letter: **GLUDy** vs. **GLUDx**.

Old and new IDs

When an existing model is modified to follow these guidelines, the original identifiers should be saved somewhere so that new analyses can be compared to old analyses that used the previous IDs.