Which genes can potentially catalyze the biotransformation of novel/orphan reactions?

Thousands of biochemical reactions with characterized activities are orphan, meaning they cannot be assigned to a specific enzyme, leaving gaps in metabolic pathways. Novel reactions predicted by pathway-generation tools also lack associated sequences, limiting protein engineering applications. Associating orphan and novel reactions with known biochemistry and suggesting enzymes to catalyze them is a daunting problem. We propose a new method, BridgIT, to identify candidate genes and catalyzing proteins for these reactions. This method introduces, for the first time, information about the *enzyme binding pocket* into reaction similarity comparisons. BridgIT assesses the similarity of two reactions, one orphan and one well-characterized, non-orphan reaction, using their substrate reactive sites, their surrounding structures, and the structures of the generated products to suggest enzymes that catalyze the most similar non-orphan reactions as candidates for also catalyzing the orphan ones.

We performed two large-scale validation studies to test BridgIT predictions against experimental biochemical evidence. For the 234 orphan reactions from KEGG 2011 (a comprehensive enzymatic reaction database) that became non-orphan in KEGG 2018, BridgIT predicted the exact or a highly related enzyme for 211 of them. Moreover, for 334 out of 379 novel reactions in 2014 that were later catalogued in KEGG 2018, BridgIT predicted the exact or highly similar enzymes.

BridgIT requires knowledge about only four connecting bonds around the atoms of the reactive sites to correctly annotate proteins for 93% of analyzed enzymatic reactions. Increasing to seven connecting bonds allowed for the accurate identification of a sequence for nearly all known enzymatic reactions.

The proposed candidate enzymes by BridgIT, are either capable of catalyzing these reactions or they can serve as good initial sequences for the enzyme engineering.

https://doi.org/10.1073/pnas.1818877116

User Manual

Step1.

Create a user account

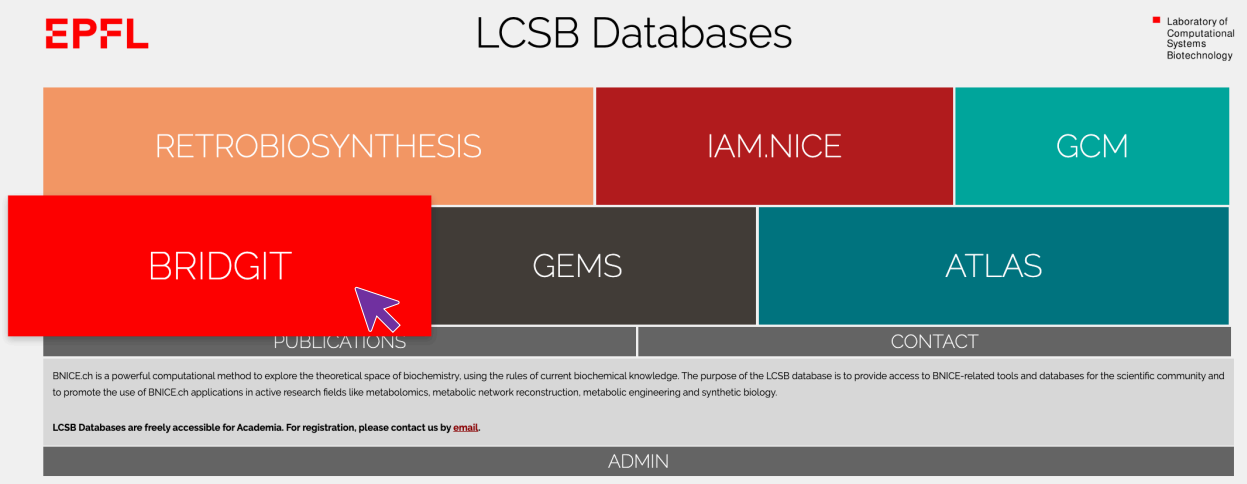
Please fill out the online request form:

https://lcsb-databases.epfl.ch/pathways/database_requests

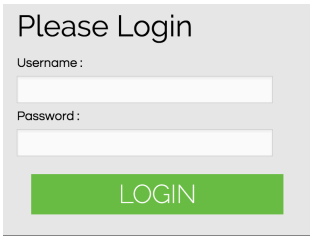
To apply for an academic license, please use your academic email address when registering.

We will send you a license agreement to be signed, and provide your username & password

Go to http://lcsb-databases.epfl.ch/pathways/Bridgit :



and login:



Please note, BridgIT tool is linked to ATLASx. In ATLASx page, there is an option to directly run BridgIT for reactions in ATALSx (no need to prepare input files).
To avoid errors in the preparation of input files, we recommend first searching your input reaction in ATLASx.
You can find more information about ATLASx here: https://lcsb-databases.epfl.ch/pathways/Atlas2

Step2.

Go to analyze tab.
You must provide a zip file with one file containing in its filename the string "systemfile", and one folder named molfiles with all the associated mol files.

The content of the ZIP file must be the following:

1) myproject_systemfile.txt
2) molfiles/1.mol molfiles/2.mol etc...

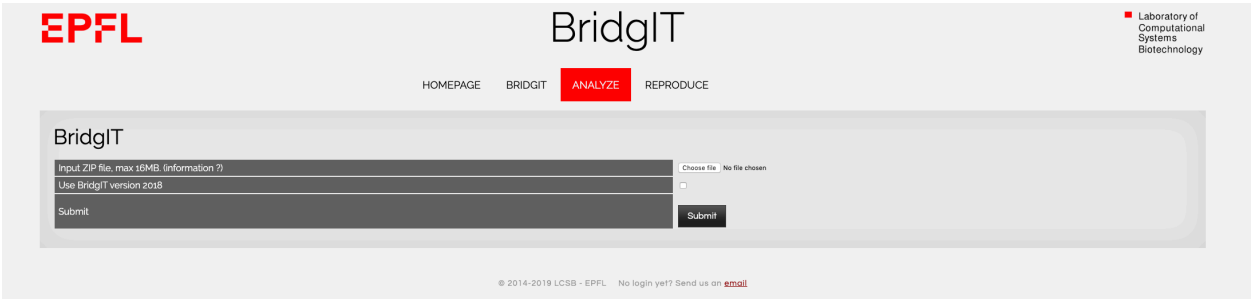
The template for myproject_systemfile.txt:

COMPOUNDS
ENTRY;GENERATION;KEGG;NAME;STRINGCODE;SPECIES;INCHIKEY;SMILES;ISBIO;FORMULA;CARBON;CHARGE;FILENAME;COFACTOR;DATABASE_LINKS;ENERGY;ERROR
reactionsS
ENTRY;KEGG;EQUATION;OPERATORS
246; ;25+58<=>26+175;

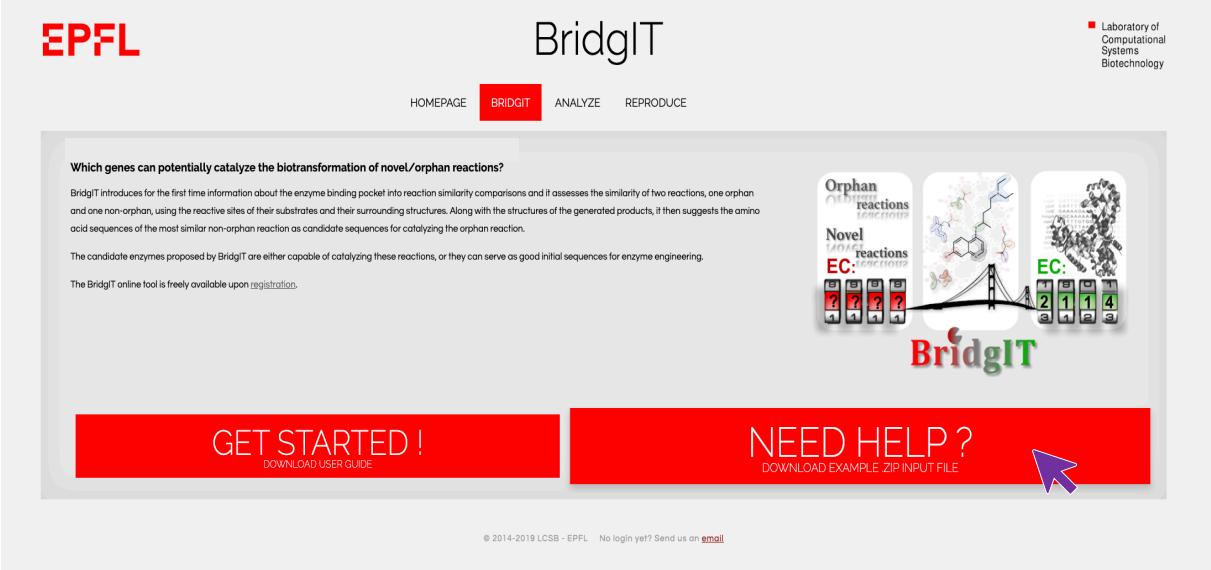
The blue texts are defining the headers for Compounds and reactions section.
Only Entry and EQUATION fields in reaction section are mandatory. You can leave the other fields empty.

If the systemfile is in BNICE.ch format, the OPERATORS field should be filled, otherwise BridgIT scans molecules with all the enzymatic rules
EQUATIONs should be in the following format:
23+24<=>25+26
25+(2)6<=>27+31

With one reaction per line and the stoichiometry defined in parentheses. The compounds are characterized by their mol file. For example, Compound 23 is defined by a file named 23.mol in the molfiles/ directory.



If you have any doubt about the formulation of Input file use the example test input and change the equation and molfiles according to your reaction.



You can have access to the published version of 2018, by clicking on “use BridgIT version 2018”, otherwise, by default the latest version of BridgIT will be launched.

Then press submit. A new page provides the URL of the results; you can access results as soon as the analysis finished. By clicking on the link and downloading a new zip file.

Zip file includes the text files of BridgIT analysis per reaction.
Output (headers and descriptions) of the **output file** are the followings:

- reactionsB/ECB: input reaction along with its enzymatic BNICE rule
- reactionsA/ECA: Similar KEGG reactions to input reaction along with its enzymatic BNICE rule and EC number
- Tanimoto_FBI_Scores: overall BridgIT score
- TL0 to TL7: similarity score in level 0 to level 7 of BridgIT fingerprint

- (r): reverse of reaction is used in similarity evaluations. Note: BridgIT considers all reactions reversible and bidirectional.
  As most of enzymes are bidirectional in vitro, in vivo directionality is controlled by co-factor affinity, cellular redox status, pH and etc.