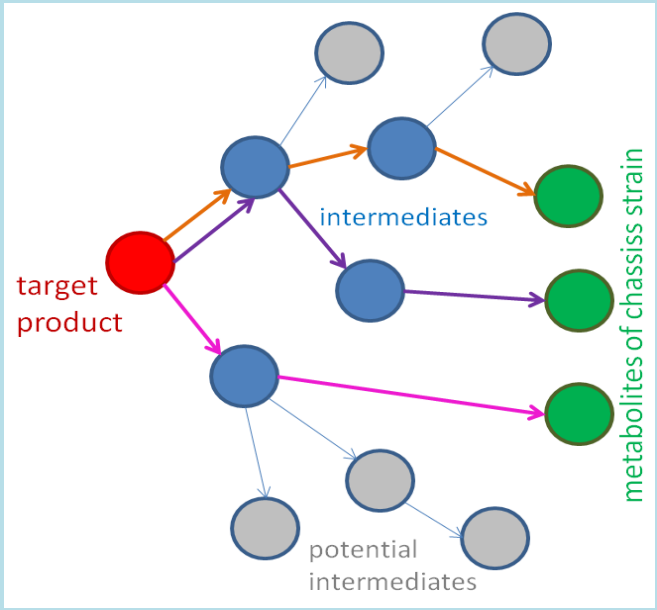
Software package for retrosynthesis-based prediction of metabolic pathways

Baboshin M.[1], Shupletsov M.[1, 2], Golubeva L.[1]

[1] *Joint Stock Company «Ajinomoto-Genetika Research Institute», Moscow, Russia*

[2] *Moscow State University, Moscow, Russia*

Bio-retrosynthesis: powerful method for pathway design

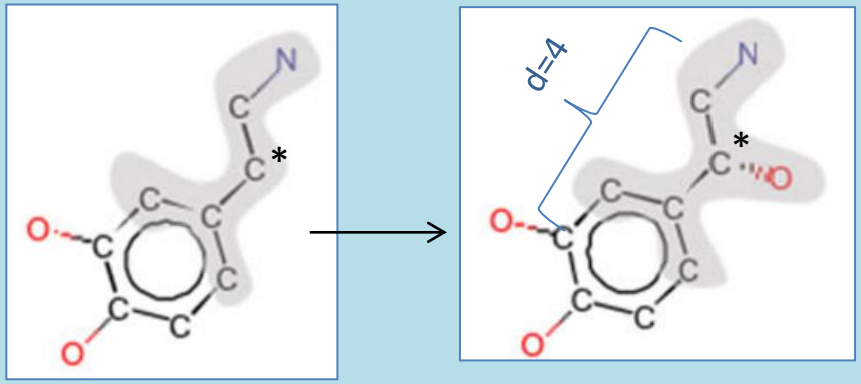


Bio-retrosynthesis:
backward trace from target compound to available metabolites

Popularity of bio-retrosynthesis

GoogleScolar gives about 10 000 references to the query ‘retrosynthesis metabolic’

Generalized reaction rule for modeling of enzyme promiscuity



*reaction center;
shadowed structural pattern corresponds to rule with diameter 4
(i.e. 2 bonds along any chain from reaction center)

Building of retrosynthetic net with generalized reaction rules enables to construct pathways containing hypothetical reactions, e.g. pathways for biosynthesis of **xenobiotics**

The aims of this work:

- develop software package able to find metabolic pathways for biosynthesis of a given target compound and rank constructed pathways to select the best pathway for implementation;
- create an updated database of generalized reaction rules that takes into account new data on biochemical reactions

Testing of updated reaction rule database by generation of metabolic spaces

1) Reproducing of reference reactions, i.e. reactions used for construction of rules – **97 %** of reference reactions reproduced;

2) Generation of metabolic spaces, i.e. chemical spases constructed with the biochemical reaction rules:

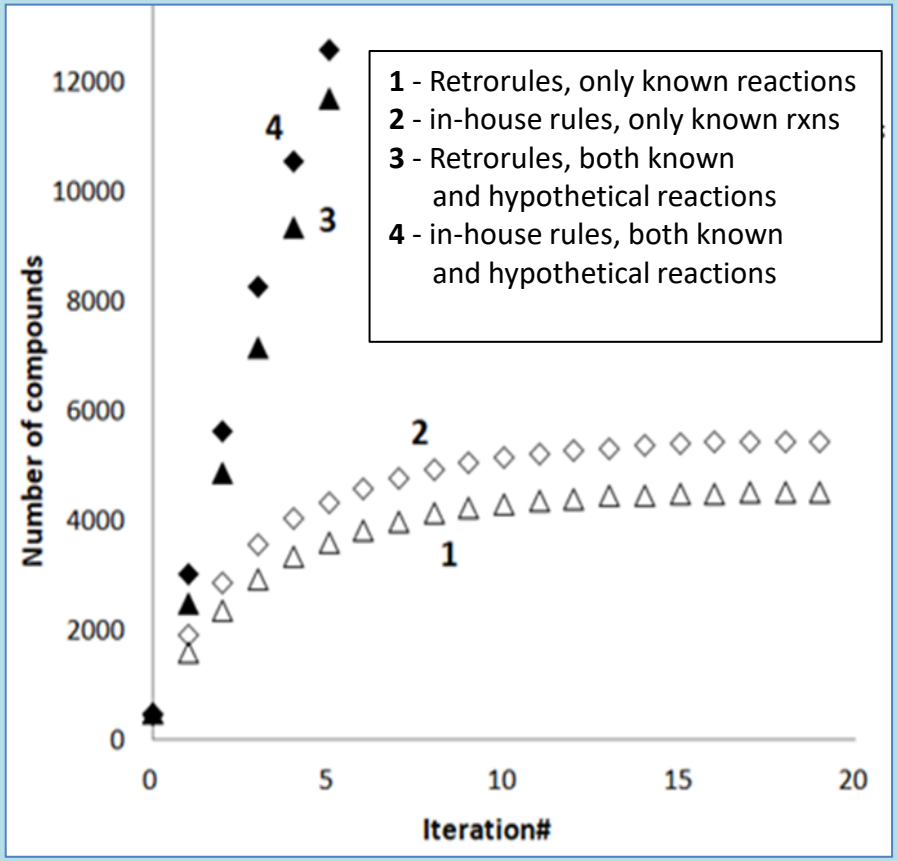


Fig. 1 Generation of metabolic spaces using our reaction rule database (generated on current version of MetaNetX, v.4) and *RetroRules* database (generated on previous version of MetaNetX, v.3). The generation started with set of *E. coli* metabolites, and reaction rules were iteratively applied.

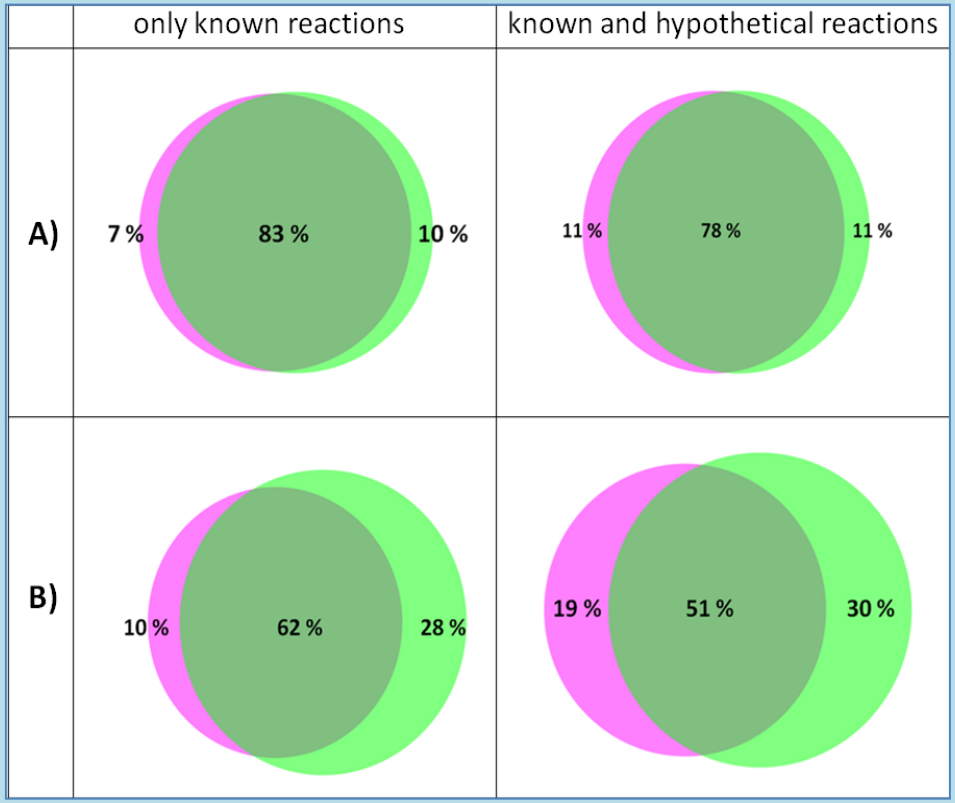


Fig. 2 Venn diagrams for comparison of metabolic spaces constructed using our reaction rule database (**green**) and *RetroRules* database (magenta)
A) common compound and reaction sets were used (for validation of our rule generating method);
B) full compound and reaction sets were used (accounting for differences in MetaNetX versions).

Overview of retrobisynthesis workflow

Generalized reaction rules, metabolite list to constrain the growing net, thermodynamic data, model for FBA analysis.

Target compound

***In house* programs**

1. **Net generation**
(generation of metabolic net containing desired pathways)

2. **Net processing**
(construction of reaction equations, deletion of irrelevant and inappropriate reactions)

3. **Pathway enumeration**
(composing of pathways from the list of reactions)

4. **Pathway ranking**
(sorting of pathways by product yield and number of hypothetical reactions)

Output: Thermodynamically consistent **pathways to target compound** ranked by number of hypothetical reactions and theoretical yield of target compound

Validation of the workflow by prediction pathways for selected target products

Target product	Reproduction of experimental pathway	Number of ranked* pathways	Rank of experimental pathway (total number of ranks)
Acrylic acid	+	15	5(9)***
Benzyl alcohol	+	18	0**(13)***
1,4-Butanediol	+	91	1(15)
Glutaric acid	+	5	1(5)
Mesaconic acid	+	3	0**(2)***
1-Methylpyrrolinium	+	11	1(7)
Muconic acid	+	18	1(10)
1-Propanol	+	146	7(26)***

*Pathways with $\Delta G'^{m}$>-10 kJ/mol or including reactions with $\Delta G'^{m}$>10 kJ/mol were filtered out before the ranking.
**These experimental pathways were reproduced but not selected to ranking due to poor thermodynamic parameters.
***In these cases, pathways with higher product yield and (in some cases) better energy profile were found.

Furthermore, net generating part of the workflow was tested on 20 target products, and experimental pathway was reproduced in all cases.

Conclusion

A software package for retrosynthesis-based prediction of metabolic pathways was created.

Testing on a series of products confirmed its applicability for finding the pathways with promising properties.

Updated database of generalized reaction rules was generated and tested by reproducing of reference reactions and construction of metabolic spaces.

This software, together with the updated database of generalized reaction rules, is an efficient tool for *in silico* prediction of metabolic pathways.