Extended metabolic space modeling

Pablo Carbonell, Baudoin Delepine, Jean-Loup Faulon

Extended Metabolic Space Modeling

Pablo Carbonell [1]✉

Email Pablo.carbonell@manchester.ac.uk

Baudoin Delépine [2,3]

Jean-Loup Faulon [1,2,3] **AQ1**

[1] Manchester Centre for Fine and Speciality Chemicals (SYNBIOCHEM), Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN UK

[2] iSSB, Genopole, CNRS, UEVE, Université Paris-Saclay, 91000 Évry, France

[3] MICALIS Institute, INRA, Domaine de Vilvert, 78352 Jouy-en-Josas Cedex, France

Abstract

Determining the fraction of the chemical space that can be processed in vivo by using natural and synthetic biology devices is crucial for the development of advanced synthetic biology ~~devices~~applications. The extended metabolic space is a coding system based on molecular signatures that enables the derivation of reaction rules for metabolic reactions and the enumeration of all possible substrates and products corresponding to the rules. The extended metabolic space expands capabilities for controlling the production, processing, sensing, and the release of specific molecules in chassis organisms.

Key words

Metabolic modeling
Enzyme reactions
Pathways
Products
Chassis

1. Introduction

The set of chemical compounds that organisms can process and synthesize is finite. Such a finite set, however, is not fully known yet. Based on a model that accounts for versatility of enzymatic reactions, we describe here a computational protocol to estimate the extent of such a full metabolic space. The extended metabolic space can be screened to list any possible biological circuit that can be conceived, such as the ones that are used to produce, detect, and process chemicals.

To fully exploit the metabolic space, an essential requirement is having a thorough knowledge of the metabolome associated with any given organism. However, experimental evidences from metabolomics analyses often show that with currently known metabolites one cannot cover the ranges of masses found in actual samples, and consequently there is an impelling need of completing the metabolomes and reactomes of interest for metabolic design [1, 2]. Furthermore, the metabolic phenotype of an organism may vary upon different conditions such as during different growth states

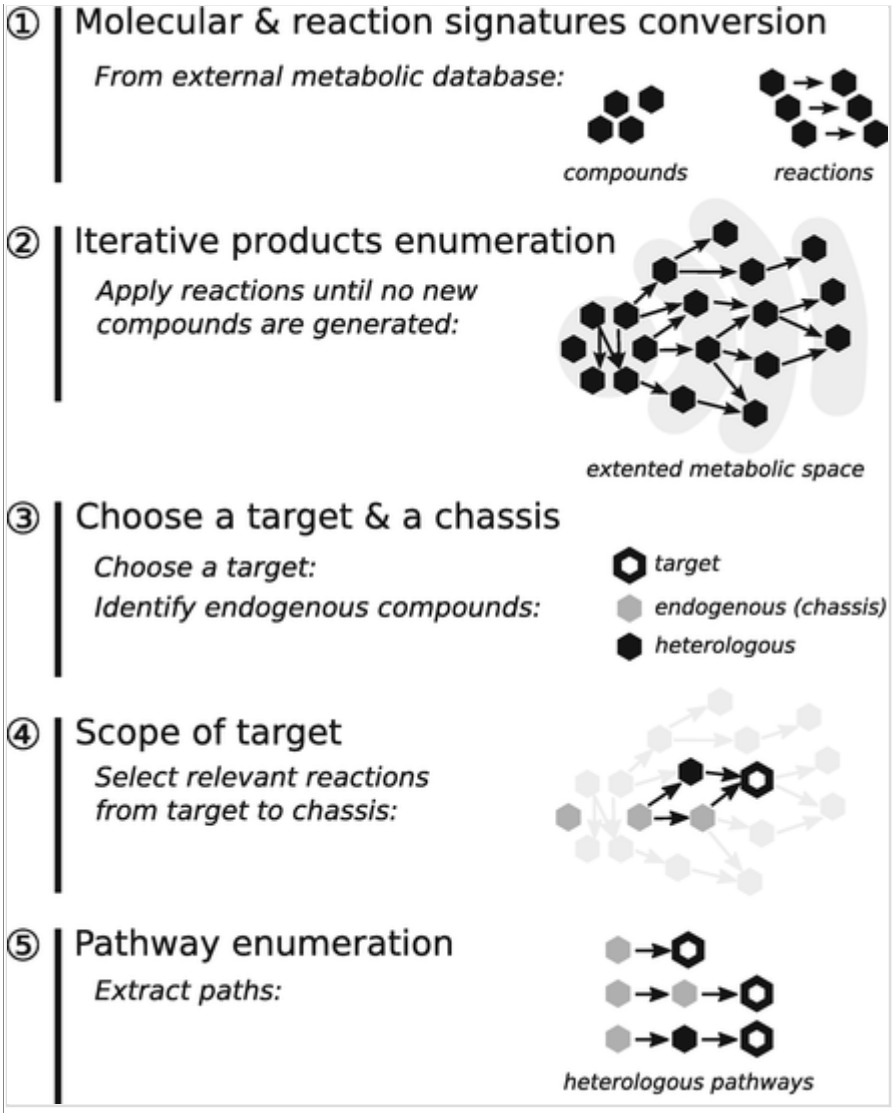
leading to variations in the metabolite profile [3]. Besides such sources of uncertainty in samples, many unassigned peaks should be due to promiscuous activities of enzymes not yet characterized because of the lack of an appropriate description of the mechanisms of enzyme promiscuity.

Our group has addressed the issue of complexity by proposing a tradeoff solution based on molecular signatures [4]. Our molecular signature codes for changes in atom bonding environments where the reaction is taking place. The advantage of the signature method is that the reaction rules describe the changes in the environments of the atoms belonging to the catalytic center of the reactions, and the size of the environment (named diameter) can be tuned to control the combinatorial explosion of possible compounds. Moreover, reaction signatures are robust to unbalanced reactions and can be created automatically without the need of any atom-atom mapping. The signature representation has shown itself to be specially well suited for modeling the mechanisms of enzyme promiscuity [5], paving by these means the way toward engineering innovation in metabolic networks. Either through directed evolution [6] or random selection [7], latent capabilities present in enzymes as modeled by the extended metabolic space can be potentially enhanced to optimize the desired activity and eventually implemented as a biological part containing a metabolic circuit.

Here, we describe the necessary steps to generate an extended metabolic space and how to compute all viable routes within the extended space that determine a viable pathway connecting a desired target to the chassis organism (Fig. 1).

Fig. 1

Steps involved in the construction of the extended metabolic space. The first step consists of converting compounds and reactions into molecular signatures. The second step enumerates new products by an iterative algorithm applied to the reaction signatures. The third step consists of choosing a target, i.e., a reaction or a compound, and a chassis organism. The fourth step determines the metabolic scope linking the chassis to the target. Finally, the fifth step enumerates all viable pathways connecting the chassis to the target



2. Materials

Materials for the described computational protocols consist basically of datasets obtained from public databases and processing software.

- A metabolic database of reference covering chemical structures and reactions. Metanetx [8] is a consensus database that reconciliates multiple databases.
- Models of metabolism for chassis organisms. Biomodels [9] and BiGG [2], among others, are databases containing genome-scale models for most commonly used organisms.
- Software to compute molecular signatures, which are a specialized type of topological chemical descriptors. ~~Molsig~~MolSig [4], among others, is an open-source package that provides such capabilities.
- Matrix manipulation software such as octave, ~~matlab~~MATLAB, scipy, R, etc.
- Computation of elementary modes. Efmtool [10] provides both a Java and ~~matlab~~MATLAB-based efficient implementations.

– Software for chemical manipulation. Some of the most popular implementations are RDKit, Marvin, CDK, KNIME (Table 1).

Table 1

A selection of software tools for modeling in the extended metabolic space

Name	Keyword	Comment
Stand-alone software		
Cytoscape	Graph visualization	Cytoscape can be used to manually explore and visualize the EMS [11]
efmtool	Elementary flux	Computation of elementary flux modes [10]
KNIME	Workflow	Knime propose to create automatic processes (“workflow”) through a drag-n-drop interface of small tasks (“node”). It is useful for reproducibility of data analysis [12]
MarvinSketch	Chemical editor	ChemAxon’s chemical editor. Useful to visualize compounds and reactions, especially to manually inspect difficult cases. URL: http://www.chemaxon.com
MolSig	Molecular signatures	Compute molecular signatures from MDL MolFile. URL: http://molsig.sourceforge.net/
Python libraries		
COBRApy	Constraints-based models	A constraint-based steady-state simulation analysis for genome-scale models [13]
NetworkX	Graph exploration	NetworkX has an intuitive interface and an extensive documentation. It is a good solution to handle the conversion of the EMS into standard graphs format, or to programmatically explore the EMS. URL: https://networkx.github.io
RDKit	Chemoinformatic toolbox	RDKit make it very easy to handle chemical structures, especially to standardize compounds. URL: http://www.rdkit.org

AQ2

3. Methods

3.1. Computation of Molecular Signatures

The first step to generate an extended metabolic space is to encode all compounds of a metabolic database in a format that will allow the subsequent encoding of enzymatic reactions. We propose here to showcase the important steps that should be kept in mind through the use of one of the available encoding methods, the molecular signature [4] (*see* **Note 1**).

1. Initially gather compounds from a metabolic database. This database must have structural data for compounds and reactions, and ideally be linked to a whole-cell model (*see* **Note 2**).

2. Check compounds for incomplete structural data. Some compounds can be defined with incomplete Markush structure or wildcard atoms. Those compounds typically stand to define classes of compounds (e.g., "an alcohol") and should be removed since they cannot be interpreted through the molecular signatures algorithm used in this protocol.

3. Standardize compounds. Molecular signatures encode directly molecular graphs from a MDL MolFile input. Users must ensure that compounds (resp., chemical groups) that should be considered identical have the same molecular graph (resp., subgraph) (*see* **Note 3**).

   (a) Neutralize or remove charges. As much as possible, chemical groups should be represented with the same protonation state to prevent different tautomeric forms. One can either use heuristics to add or remove hydrogen when necessary or simply remove all charges from the compound dataset.

   (b) Choose one conjugated form by compound. This is particularly important for aromatic compounds, which could appear under different kekulé forms in the database. A good solution is to explicitly use aromatic bonds in the molecular graph description.

   (c) Use a consistent hydrogen representation, either implicit or explicit.

4. To compute the signature of a chemical compound, we need initially to consider its molecular graph. Let $G(V, E)$ be the molecular graph associated with some chemical compound C and let $a \in V$ ($b \in E$) be an atom (bond) of G. The atomic signature of atom a of diameter d, ${}^{d}\sigma(a)$, is a canonical representation of the subgraph of G spanned by its vertices at a maximum distance of $d/2$ from a. From a chemical point of view, this corresponds to a circular fragment of the compound centered on a.

5. The molecular signature of a molecular graph G of diameter d associated with C, ${}^{d}\sigma(G)$, is defined as the list of all atomic signatures of diameter d (one by atom). Therefore, a molecular signature is a list of overlapping molecular fragments.

6. Depending on the diameter d, a molecular fingerprint can show degeneracy, i.e., a same molecular signature can represent more than one molecular graph G, much like a chemical formula can correspond to several compounds.

7. Based on previous definitions, the computation of the molecular signature involves two steps:

   (a) Choose a diameter to encode enzymatic promiscuity. To some extent, enzymes have the ability to process additional reactants that are structurally similar to the known ones. In a context where it is important to maximize the number of reactions to get more leads, modeling promiscuity can reveal itself to be a critical feature (*see* **Note 4**). We recommend starting with a diameter of 12 and going lower (down to 4) if no satisfying solution can be found.

   (b) Compute molecular signatures. The MolSig software [4] computes molecular signatures starting with compounds in MDL MolFiles format, which can be easily retrieved from metabolic and chemical databases or converted from other equivalent formats (*see* **Note 5**).

3.2. Computation of Reaction Signature

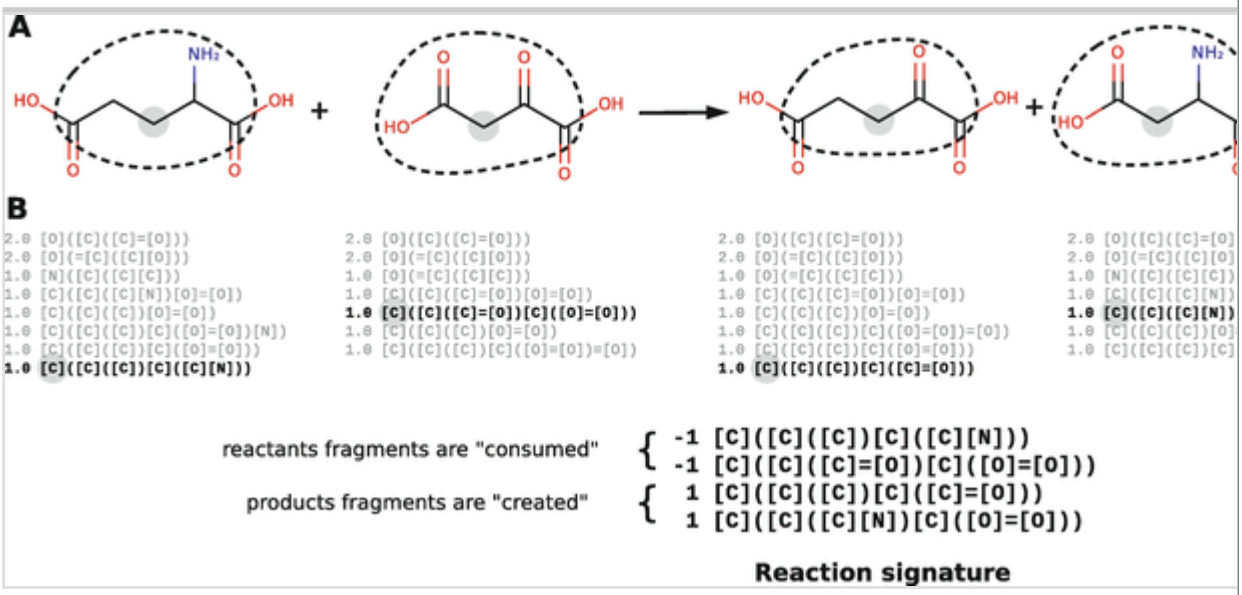
The step following the encoding of compounds is the encoding of reactions into reaction signatures. Reaction signatures should be understood as an exchange of fragments. Unlike other reaction models, reaction signatures do not need any atom-atom mapping to be computed, nor do they need reaction to be balanced (*see* **Note 6**).

1. Let R be a reaction for which all substrates $\{S_i, i \in [1,n]\}$ and products $\{P_j, i \in [1,m]\}$ are encoded in molecular signatures, respectively $\{{}^d\sigma(S_i), i \in [1,n]\}$ and $\{{}^d\sigma(P_j), j \in [1,m]\}$. The reaction signature is defined as follows: $${}^d\sigma(R) = \bigcup_{j=1}^{m} {}^d\sigma\left(P_j\right) - \bigcup_{i=1}^{n} {}^d\sigma\left(S_i\right)$$ (*see* **Note 7**).

2. Thus, ${}^d\sigma(R)$ is the difference in terms of atomic signatures (i.e., molecular fragments) occurring reaction; created (resp. consumed or needed) fragments being positives (resp. negatives). In this the diameter d corresponds to the reacting moieties and their neighboring atoms (the environmen the possibility to tune the degree of the enzymatic promiscuity hypothesis by increasing or decre (Fig. 2).

   Fig. 2

   Reaction signature of an aspartate transaminase (EC 2.6.1.1, d = 4). Panel (**a**) shows the stru the compounds involved in the original (aspartate + 2-oxoglutarate → oxaloacetate + glutamate). Fragments (atomic signature) that ar reaction signature are *circled* (*dashed line*) and ~~its~~their ~~center~~centers marked by a *gray dot*. outside of the circle are allowed to vary under an enzymatic promiscuity hypothesis (d = 4). I shows the atomic signatures and resulting reaction signature (d = 4). Fragments involve reaction signature are highlighted (*bold*) in molecular signatures. Note that several fragn compound can end up in the reaction signature, even if that is not the case here



3.3. Products Enumeration

Once reactions have been encoded into reaction signatures, they can be applied to compounds to predict potential products under the enzymatic promiscuity hypothesis.

1. Let *DB* be a database binding compounds signatures to their respective molecular graphs.

2. Let ${}^{d}\sigma(R)$ be the molecular signature associated with a reaction R, and $\{S_i^{'}, i \in [1,n]\}$ a set of candidate substrates potentially reacting together.

3. Under the enzymatic promiscuity hypothesis determined by d, we predict that R can process any candidate substrate $\{S_i^{'}, i \in [1,n]\}$ if:

   (a) $\bigcup_{i=1}^{n} {}^{d}\sigma\left(S_i'\right) \supseteq \left\{x \in {}^{d}\sigma(R), x < 0\right\}$, i.e., if the signatures of candidate substrates include all fragments consumed by R,

   (b) and the predicted product(s) signature(s) ${}^{d}\sigma(P^{'})$ correspond to some previously known compound(s) in *DB*, with ${}^{d}\sigma\left(P'\right) = \bigcup_{i=1}^{n} {}^{d}\sigma\left(S_i'\right) + {}^{d}\sigma(R)$ (*see* **Note 8**).

4. Being able to model enzymatic promiscuity assumes that reaction signatures can be used with other substrates than the ones in the native reaction. In turn, alternative substrates produce new products. Those compounds may be absent from the metabolic space, i.e., the set of known metabolites. Therefore, reaction signatures extend the metabolic space by linking potentially new compounds to the metabolism (*see* **Note 9**).

3.4. Chassis Modeling in the Extended Metabolic Space

In the previous sections we have described the protocol that allows extending the metabolic space. When the extension is applied to a metabolic network consisting of all known metabolic reactions, we arrive at the full description of all available metabolic capabilities. Some of these capabilities are going to be common to several groups of organisms, such as reactions in the central metabolism, while others like secondary metabolism will be specific to some groups. In applications such a biotechnology, the organism that is engineered is known as the chassis organism and often the objective will be to expand the natural capabilities of the chassis by introducing heterologous enzymes. In this section, we will describe how to model the chassis organism as a subset of the extended metabolic space.

1. The extended metabolic space of diameter d, denoted by M_d, represents all the possible compounds C and allowed transformations (reactions) R between compound as spanned by the enumerated reactions computed by following the method described.

2. A chassis is a subset of the extended metabolic space $O_d \subset M_d$ that corresponds to the extended metabolic network of an organism at signature diameter d. A chassis is defined by the set of nominal reactions annotated for the enzymes present in the organism.

3. The list of nominal metabolic reactions for a given organism can be compiled from databases such as KEGG [14], MetaCyc [15], BiGG [2], BRENDA [16], etc. The choice of one database over the others depends on several factors:

   (a) The degree of curation of the model.

   (b) The free and open availability of the model.

   (c) The way the model is going to be analyzed, i.e., network analysis, steady-state

simulation or simply as a reference list of metabolites and reactions (*see* **Note 10**).

4. In silico organism models showing a good degree of accuracy and reproducibility are currently available for many industrial strains, including *Escherichia coli*, *Saccharomyces cerevisiae*, or *Bacillus subtilis*. They can be generally downloaded in a SBML format [17].

5. To determine O_d, each reaction in the reference model is augmented with the set of enumerated reactions of the chassis in the extended metabolic space, resulting in an extended model (*see* **Note 11**).

3.5. Computing the Scope

The next step in modeling in the extended metabolic space is to have an understanding of the design space for a given target metabolic activity. In other words, we want to compute the metabolic scope connecting some target reaction to the chassis. To that end, we provide in this section some relevant definitions and a two-step procedure that allows the determination of the metabolic scope.

1. A minimal pathway is defined as any set of reactions connecting the chassis to the target that are minimal:
   (a) They form a viable production pathway in terms of precursors availability.
   (b) All reactions are essential, i.e., the removal of any reaction renders nonviable the pathway (*see* **Note 12**).

2. Based on that definition, the metabolic scope is defined as follows: given an initial set of source metabolites **S** (the chassis) and a final set of target metabolites **T**, the scope is the set of enzymes that are at least involved in one minimal pathway connecting elements of **T** to the source **S**, i.e., the scope should contain only enzymes that are at least essential for establishing one of the metabolic pathways. To compute the scope for a given compound, a two-step procedure can be applied, as described in the following.

3. Reduction of the extended metabolic space to the reachable space of reactions. It consists of the following steps:
   (a) A compound is defined as reachable if there exists a reachable reaction that can produce it, i.e., a reaction for which all substrates are available.
   (b) Start from the set of initial compounds **S** and iteratively find newly reachable compounds.
   (c) The process stops when no new reachable compounds are found.
   (d) Build a graph to keep track of which reactions produced each compound.

4. Backward determination of the scope. It consists of the following steps:
   (a) Start from the target compound(s) **T**. For each reaction that can produce the target compound(s), add it to the scope.

(b) Recursively apply the same procedure on each substrate of the reaction.

(c) The recursion stops when initial compounds **S** are reached.

3.6. Enumerating Pathways

Once the extended metabolic scope has been determined, we should be interested in enumerating all viable metabolic pathways connecting the source to the target. This turns out to be a computationally complex problem that can be solved through several approaches [18]. We describe here a solution based on the computation of elementary flux modes [19] (*see* **Note 13**). EFMs are the set of minimal pathways that are nontrivial solutions to the steady-state equation whose combination can describe any possible path in the network (*see* **Note 14**).

1. Define the augmented metabolic space formed by the union of the reactions in the chassis and in the scope (Fig. 3a).

2. Construct a stoichiometric matrix **S** where each row corresponds to a compound and each column to a reaction of the previous augmented metabolic spaces and the value of each cell is the stoichiometric coefficient (Fig. 3b).

3. Remove all rows representing initial compounds (*see* **Note 15**).

4. Remove all rows representing compounds that are produced by a reaction but never used in any other.

5. Merge identical columns by deleting redundant columns and renaming the remaining column with the names of all reactions (*see* **Note 16**).

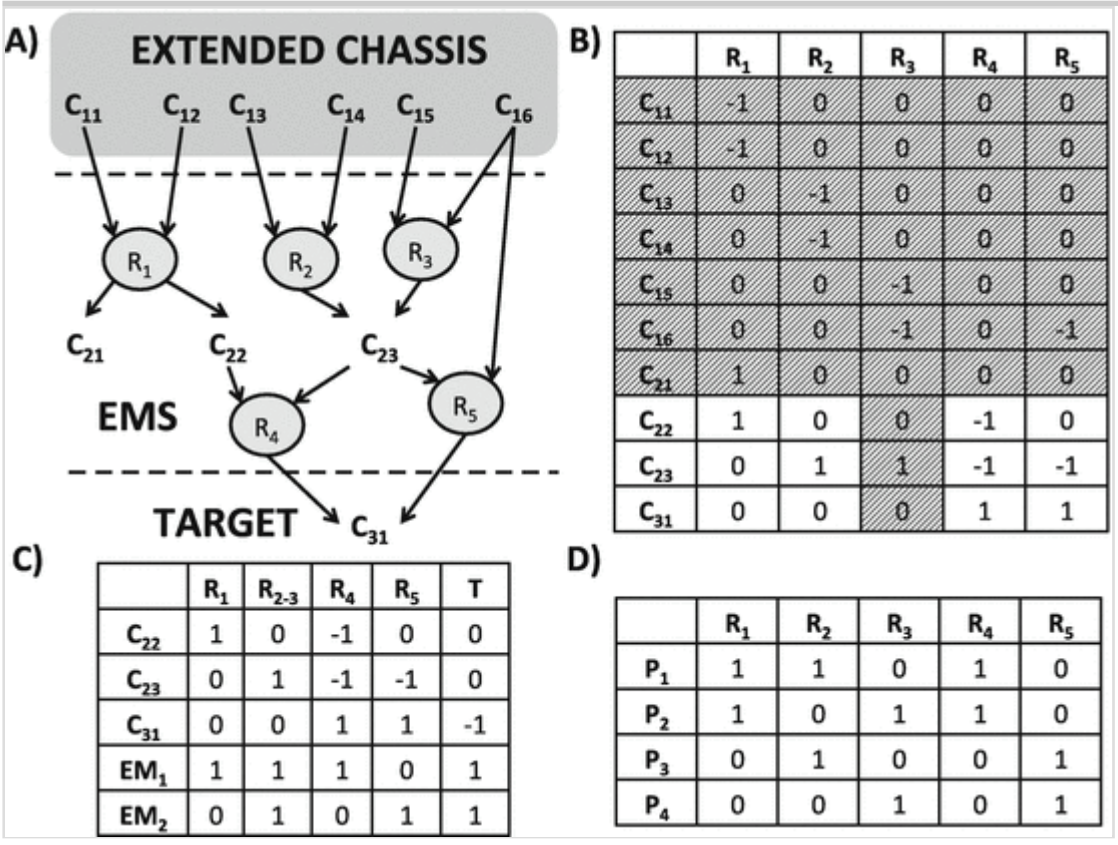
6. Add an additional column to create a flux out for the target compound.

7. Several toolboxes exist that allow efficiently computing the elementary modes (Fig. 3c). For instance, efmtool [10] provides an efficient implementation that can either run in ~~matlab~~MATLAB or in Java.

8. Expand resulting elementary modes into the pathway solutions by enumerating all combinations of merged reactions in each elementary mode (Fig. 3d).

Fig. 3

Example of pathway enumeration in the extended metabolic space. Panel (**a**) shows the scope graph connecting compounds in the extended chassis (C_{11}, C_{12}, C_{13}, C_{14}, C_{15}, C_{16}) to target compound C_{31} through reactions R_1, R_2, R_3, R_4 and intermediate compounds C_{21}, C_{22}, C_{23} in the extended metabolic space (EMS). Panel (**b**) displays the equivalent stoichiometric matrix. *Grayed columns* and *rows* are discarded in the enumeration, as described in the enumeration protocol. Panel (**c**) shows the reduced matrix used for enumeration, containing an additional reaction T for the selected target compound. The enumeration algorithm found two elementary modes EM_1 and EM_2. Panel (**d**) shows the resulting four pathways solution P_1–P_4 after expansion of topological equivalent reactions. Pathways P_1 and P_2 involve three reactions, while pathways P_3 and P_4 involve two reactions



B)

	R_1	R_2	R_3	R_4	R_5
C_{11}	-1	0	0	0	0
C_{12}	-1	0	0	0	0
C_{13}	0	-1	0	0	0
C_{14}	0	-1	0	0	0
C_{15}	0	0	-1	0	0
C_{16}	0	0	-1	0	-1
C_{21}	1	0	0	0	0
C_{22}	1	0	0	-1	0
C_{23}	0	1	1	-1	-1
C_{31}	0	0	0	1	1

C)

	R_1	$R_{2\text{-}3}$	R_4	R_5	T
C_{22}	1	0	-1	0	0
C_{23}	0	1	-1	-1	0
C_{31}	0	0	1	1	-1
EM_1	1	1	1	0	1
EM_2	0	1	0	1	1

D)

	R_1	R_2	R_3	R_4	R_5
P_1	1	1	0	1	0
P_2	1	0	1	1	0
P_3	0	1	0	0	1
P_4	0	0	1	0	1

3.7. Design in the Extended Metabolic Space

We have described in previous sections step-by-step methods that generate extended metabolic spaces for (a) global metabolic capabilities; (b) chassis organisms; (c) organisms augmented with desired target activities. From here, resulting extended models can be used in multiple engineering biology applications, from production of chemicals to their sensing and regulation. Some of the main applications developed to date in extended metabolic spaces include the following:

1. Engineering of heterologous pathways for the production of a desired chemical in a chassis organism. To select enzyme sequences for each enzymatic step in the pathway for the most promising routes in the extended metabolic space, a pathway ranking function needs to be defined. The approach is described in detail in the retrosynthetic RetroPath protocol [20] and a

demonstration of the application of such a protocol is shown in the XTMS web service [21].

2. Development of novel biosensors based of metabolic pathways. Metabolic pathways that transform a target compound into a detectable compound allow the expansion of the observable extended metabolic space [22]. Such an application has been demonstrated through the SensiPath web service [23].

4. Notes

1. Molecular signatures are an efficient and intuitive way to model metabolites. They are similar to the well-known Extended Connectivity FingerPrint (ECFP) topological fingerprint, which summarizes compounds in lists of circular molecular fragments.

2. Chemical structures and reactions can be found in multiple formats. Reactions are often defined in a database-specific flat-file where reactants are referenced by their compound identifier. Most of the time, you will find a file in MDL SDF or MOL format binding the compound identifiers to their respective structures. Other interchangeable formats are usually available such as SMILES and InChI. Inter-conversion between formats using standard software such as Open Babel [24] yields to equivalent representations of the compound. A sanity check can help to ensure that they all refer to the same compound. This will eventually filter out wrong annotations.

3. Before being converted into molecular signatures, molecular graphs do not need to represent chemically valid compounds in terms of valence, charges, etc. The important point is that compounds (moieties) that should be considered identical according to the final application share the same molecular graph (subgraph). Of course, those simplifications introduced at the compound encoding step must be kept in mind while interpreting the results.

4. Putative enzymes promiscuity can be modeled through molecular signatures given an appropriate diameter. Obviously, as we lower the diameter, the stronger is the promiscuity hypothesis and the riskier are the predictions.

5. Molecular signatures can take into account stereo-chemistry, which is particularly appealing when working with enzymes. Nonetheless, if stereo information is considered, it is important to ensure that it is available (and valid) for most of the compounds; otherwise, compounds with and without stereo information will be perceived differently through signatures.

6. Metabolic databases contain generally a substantial portion of reactions that are not stoichiometrically balanced. Reaction signatures can be computed for reactions that do not need strictly balanced input reaction. Nonetheless, working with balanced reactions is always recommended and is a sign of a well-curated database.

7. This mathematical expression simply states that the reaction signature is the set formed by the difference between product signatures and reactions signatures. Intuitively, it can be understood as the chemical groups that are transferred or transformed through the reaction.

8. Multi-substrate reactions are difficult to handle with the proposed equation. Indeed, testing all compounds with a reaction would take N^m tries, where N is the total number of compounds in the database and m the number of substrates anticipated for that reaction. A more practical

option is to allow promiscuity for only one substrate at a time, therefore limiting the number of trials ~~to~~to $N^{*}m$. A complementary approach is to allow promiscuity only for non-cofactors compounds.

9. This feature is particularly desirable to untap enzymes full potential in metabolic engineering applications since it can find an unexpected synthesis route.

10. There is a basic difference between the information that is required in the model to design heterologous metabolic pathways and to estimate steady-state fluxes. In the former case, the most essential information is the knowledge about the metabolites that are endogenous to the organism and therefore can be used as precursors in the heterologous pathway. In the latter case, the accuracy of the stoichiometric relationship between those reactions that directly influence the pathway is required, while partial knowledge about upstream reactions with low influence into the pathway can be tolerated.

11. The extended metabolic space of the model of an organism provides useful information to discover previously unidentified routes and to fill gaps in present models.

12. Pathway minimality is a heuristic condition based on reducing metabolic burden in the cell (a pathway with a less number of enzymes should be more tolerated by the cell because it potentially imposes less stress).

13. Metabolic networks are formally modeled as hypergraphs for pathway enumeration. Basically, the availability of each substrate is required in the reaction to produce the product. That creates some level of complexity higher than in the classical graph pathway enumeration algorithm. Moreover, standard graph approaches do not consider stoichiometry. The stoichiometric approach, in turn, based on linear algebraic decomposition provides an easier analytic approach.

14. Pathway enumeration based on elementary flux modes can become computationally intractable for highly connected networks such as central metabolism. However, in cases where we want to produce some heterologous compound in a chassis organism, pathways are generally almost linear and the elementary flux mode enumeration remains tractable. The enumeration of elementary flux modes can be also expressed as a dual problem using minimal cut sets.

15. We remove all the initial compounds in the chassis, as we already know that they are available. Products of reactions in the scope consuming the initial compounds will be kept for the enumeration.

16. Identical columns represent routes that are topologically equivalent. In order to make the enumeration algorithm more efficient, we remove duplicated columns. However, for the final enumeration we should list each topologically equivalent reaction as an alternative pathway.

Acknowledgment

The authors acknowledge funding from the BBSRC under grant BB/M017702/1, “Centre for synthetic biology of fine and speciality chemicals.” J.L.F. acknowledges funding provided by the ANR under grant ANR-15-CE21-008. B.D. is affiliated to the Doctoral School Structure et Dynamique des Systèmes Vivant, Université Paris-Saclay.

References

1. Wishart DS, Jewison T, Guo AC et al (2013) HMDB 3.0—the human metabolome database in 2013. Nucleic Acids Res 41:D801–D807. doi: 10.1093/nar/gks1065

2. Schellenberger J, Park J, Conrad T, Palsson B (2010) BiGG: a biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions. BMC Bioinformatics 11:213. doi: 10.1186/1471-2105-11-213

3. Tu BP, Mohler RE, Liu JC et al (2007) Cyclic changes in metabolic state during the life of a yeast cell. Proc Natl Acad Sci U S A 104:16886–16891. doi: 10.1073/pnas.0708365104

4. Carbonell P, Carlsson L, Faulon J-L (2013) Stereo signature molecular descriptor. J Chem Inf Model 53:887–897. doi: 10.1021/ci300584r

5. Carbonell P, Faulon J-L (2010) Molecular signatures-based prediction of enzyme promiscuity. Bioinformatics 26:2012–2019. doi: 10.1093/bioinformatics/btq317

6. Dougherty MJ, Arnold FH (2009) Directed evolution: new parts and optimized function. Curr Opin Biotechnol 20:486–491. doi: 10.1016/j.copbio.2009.08.005

7. Kim J, Kershner JP, Novikov Y et al (2010) Three serendipitous pathways in E. coli can bypass a block in pyridoxal-5-phosphate synthesis. Mol Syst Biol. doi: 10.1038/msb.2010.88

8. Moretti S, Martin O, Van Du Tran T et al (2016) MetaNetX/MNXref reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks. Nucleic Acids Res 44:D523–D526. doi: 10.1093/nar/gkv1117

9. Chelliah V, Juty N, Ajmera I et al (2014) BioModels: ten-year anniversary. Nucleic Acids Res 43:D542–D548. doi: 10.1093/nar/gku1181

10. Terzer M, Stelling J (2008) Large-scale computation of elementary flux modes with bit pattern trees. Bioinformatics 24:2229–2235. doi: 10.1093/bioinformatics/btn401

11. Shannon P, Markiel A, Ozier O et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. doi: 10.1101/gr.1239303

12. Berthold MR, Cebron N, Dill F et al (2009) KNIME – the Konstanz information miner. SIGKDD Explor 11:26–31. doi: 10.1145/1656274.1656280

13. Ebrahim A, Lerman JAJ, Palsson BO, Hyduke DR (2013) COBRApy: COnstraints-Based Reconstruction and Analysis for python. BMC Syst Biol 7:74. doi: 10.1186/1752-0509-7-74

14. Kanehisa M, Goto S, Sato Y et al (2012) KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40:D109–D114. doi: 10.1093/nar/gkr988

15. Caspi R, Altman T, Dreher K et al (2012) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res 40:D742–D753. doi: 10.1093/nar/gkr1014

16. Chang A, Schomburg I, Placzek S et al (2014) BRENDA in 2015: exciting developments in its 25th year of existence. Nucleic Acids Res. doi: 10.1093/nar/gku1068

17. Hucka M, Finney A, Sauro HM et al (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19:524–531

18. Carbonell P, Fichera D, Pandit S, Faulon JL (2012) Enumerating metabolic pathways for the production of heterologous target chemicals in chassis organisms. BMC Syst Biol 6:10. doi: 10.1186/1752-0509-6-10

19. Zanghellini J, Ruckerbauer DE, Hanscho M, Jungreuthmayer C (2013) Elementary flux modes in a nutshell: properties, calculation and applications. Biotechnol J 8:1009–1016. doi: 10.1002/biot.201200269

20. Carbonell P, Planson A-GG, Faulon J-LL (2013) Retrosynthetic design of heterologous pathways. Methods Mol Biol 985:149–173. doi: 10.1007/978-1-62703-299-5_9

21. Carbonell P, Parutto P, Herisson J et al (2014) XTMS: pathway design in an eXTended metabolic space. Nucleic Acids Res:W389–W394. doi: 10.1093/nar/gku362

22. Carbonell P, Parutto P, Baudier C et al (2014) Retropath: automated pipeline for embedded metabolic circuits. ACS Synth Biol 3:565–577. doi: 10.1021/sb4001273

23. Delpine B, Libis V, Carbonell P, Faulon J-L (2016) SensiPath: computer-aided design of sensing-enabling metabolic pathways. Nucleic Acids Res. doi: 10.1093/nar/gkw305

24. O'Boyle NM, Banck M, James CA et al (2011) Open babel: an open chemical toolbox. J Cheminform 3:33. doi: 10.1186/1758-2946-3-33