A Lattice-Theoretic Framework for Metabolic Pathway Analysis

Yaron A.B. Goldstein and Alexander Bockmayr

DFG Research Center MATHEON, Freie Universität Berlin,
Arnimallee 6, 14195 Berlin, Germany
{yaron.goldstein,alexander.bockmayr}@fu-berlin.de

Abstract. Constraint-based analysis of metabolic networks has become a widely used approach in computational systems biology. In the simplest form, a metabolic network is represented by a stoichiometric matrix and thermodynamic information on the irreversibility of certain reactions. Then one studies the set of all steady-state flux vectors satisfying these stoichiometric and thermodynamic constraints.

We introduce a new lattice-theoretic framework for the computational analysis of metabolic networks, which focuses on the support of the flux vectors, i.e., we consider only the qualitative information whether or not a certain reaction is active, but not its specific flux rate. Our lattice-theoretic view includes classical metabolic pathway analysis as a special case, but turns out to be much more flexible and general, with a wide range of possible applications.

We show how important concepts from metabolic pathway analysis, such as blocked reactions, flux coupling, or elementary modes, can be generalized to arbitrary lattice-based models. We develop corresponding general algorithms and present a number of computational results.

Keywords: metabolic networks, constraint-based analysis, lattices.

1 Introduction

Constraint-based modeling has become a very successful approach for the analysis of genome-scale reconstructions of metabolic networks [1–4]. Given a set of metabolites M and a set of reactions R, the network is represented by its stoichiometric matrix $S \in \mathbb{R}^{\mathsf{M}\times\mathsf{R}}$, and a subset of irreversible reactions $\mathrm{Irrev} \subseteq \mathsf{R}$. The steady-state flux cone $C = \{v \in \mathbb{R}^{\mathsf{R}} \mid Sv = 0, v_{\mathrm{Irrev}} \geq 0\}$ contains all steady-state flux vectors satisfying the stoichiometric and thermodynamic constraints. Based on this cone, many analysis methods have been introduced over the years, among them *Flux Balance Analysis* (FBA) [5, 6], *Elementary Mode Analysis* (EMA) [7–9], and *Flux Coupling Analysis* (FCA) [10, 11].

While these methods are now well-established, various ideas have been explored on how to modify or extend the underlying modelling framework. A lot of research concerns the question of how to include regulatory information

A. Gupta and T.A. Henzinger (Eds.): CMSB 2013, LNBI 8130, pp. 178–191, 2013.

into the metabolic model (e.g. [12]). This has lead to diverse FBA strategies like rFBA [13] or SR-FBA [14]. Elementary mode computation has been extended to include transcriptional regulatory networks in [15]. Further, there has been a discussion on whether stronger thermodynamic constraints should be applied [16,17]. Others combine the idea of FBA to analyse optimal-growth steady-states with the insight that this condition alone does not constrain the system to a single possible state, but to a mathematical space of different (biologically) optimal states [18]. Still other approaches give up the steady-state assumption and use completely different modelling approaches, e.g. hyperpaths that are constructed by ordering the reactions of a network based on their (graph-theoretical) distance to nutrients [19]. So far each modification of the basic modelling approach required a specific reformulation and adaptation of the algorithms and analysis tools.

In this paper, we introduce the algebraic framework of lattices as a unifying approach to metabolic pathway analysis. We will present the necessary concepts that will allow us to adopt a broad range of modelling ideas within a unique generic framework. We have already tested ways to include optimal-growth or thermodynamic constraints as an option into our analysis tools. As a next step, we intend to create a formalism for regulatory constraints, which can be added to lattice-based models. Once implemented and tested, we will be able to perform EMA and especially FCA with regulatory or thermodynamic constraints.

Finite lattices [20, 21] are some of the simplest algebraic structures, but they have proven to be useful in many applications, such as abstract interpretation [22], knowledge representation [23], or distributed computing [24]. As we will see, they can be employed naturally to describe qualitative, pathway-based metabolic models, including the steady-state flux cone and related constraint-based methods. Regarding qualitative modelling, our work is related to [25], who use the concept of abstract interpretation to give knockout predictions in reaction networks.

Here we will introduce lattice-based EMA and a very fast FCA method. Our implementation L4FC (Lattices for Flux Coupling) can be used for traditional, flux-cone-based FCA. But it also allows applying other lattice-based modelling approaches, by simply changing one particular method that looks for pathways through a given reaction in the model.

Lattice-based models are independent from the steady-state assumption. In our models, we can use the flux cone, but we do not have to. The only algebraic requirement a lattice-based model has to fulfill is one that is easily proven for most approaches: any two pathways or states a and b can be combined to a new one that uses together all the reactions of a and b. This already defines a semi-lattice, which in our setting will automatically be a lattice.

Our approach allows for more flexibility in choosing the model constraints and provides general analysis tools that we can immediately use without spending much time on adapting them to our needs. As we will see, lattice-based modelling is fully compatible with the traditional steady-state flux cone and many of its extensions. But, it is also open for completely new ideas.

2 Lattice Theory in Metabolic Pathway Analysis

Many important questions in metabolic pathway analysis involve only qualitative information: Which reactions participate in a pathway? Which are the minimal sets of reactions needed to realize certain biological functions? Which reactions are coupled to each other? To answer these and other questions, we do not need the quantitative information of reaction rates. Instead we can consider a pathway to be simply a subset a of the reaction set R, $a \subseteq \mathsf{R}$, satisfying certain properties. This idea has appeared before in the literature, e.g. as *activity sets* [26] or *flux patterns* [27]. As a unifying framework for various modelling approaches in metabolic pathway analysis, we propose in this paper the algebraic concept of (semi-)lattices.

A *semi-lattice* [21] is an algebraic structure $(L, \circ)$ consisting of a set L and a binary operation $\circ$ which satisfy the following axioms:

- L is $\circ$-closed, i.e., if $a, b \in L$ then $a \circ b \in L$.
- $\circ$ is associative and commutative, i.e., $a \circ (b \circ c) = (a \circ b) \circ c$ and $a \circ b = b \circ a$.
- $\circ$ is idempotent, i.e., $a \circ a = a$.

A *lattice* can be defined as an algebraic structure $(L, \vee, \wedge)$ such that $(L, \vee)$ and $(L, \wedge)$ are semi-lattices and in addition for any $a, b \in L$, we have $a \wedge (a \vee b) = a$, and $a \vee (a \wedge b) = a$. An example is the lattice $(2^X, \cup, \cap)$ of all subsets of a set X, together with the usual set operations of union and intersection.

In the context of metabolic pathway analysis, we will look at semi-lattices $(L, \cup)$, where $L \subseteq 2^{\mathsf{R}}$ and R is the finite set of reactions in the metabolic network. As we will see, many metabolic models are indeed union-closed, which simply means that the union of two pathways is a pathway again. As noted in [28], such a finite semi-lattice is already a lattice if there exists a neutral element $0 \in L$, with $0 \cup a = a$, for all $a \in L$. This holds if $\emptyset \in L$. Thus for any $L \subseteq 2^{\mathsf{R}}$, we can obtain a *lattice* $(L, \cup, \wedge)$ if the following two axioms are satisfied:

- L is $\cup$-closed, i.e., if $a, b \in L$ then $a \cup b \in L$.
- There is an element $0 \in L$ such that $0 \cup a = a$, for all $a \in L$.

With these two axioms, we can define a second operation $\wedge$ on L, so that $(L, \cup, \wedge)$ becomes a lattice:

$$a \wedge b := \bigcup_{c \subseteq a, c \subseteq b} c \ . \tag{1}$$

The operation $\wedge$ is well-defined because $0 \subseteq a$, for all $a \in L$.

Similarly to this construction, we can prove that every finite lattice L has a unique maximum 1_L:

$$1_L = \bigcup_{a \in L} a \ . \tag{2}$$

Since $a \subseteq 1_L$, for all $a \in L$, we call 1_L the *maximum* of L. In Sect. 4, we will use the maximum to reformulate the concept of blocked reactions and flux coupling in metabolic network analysis.

Additionally, there are ways to describe finite lattices based on special sets of elements, the so-called *minimal* and *irreducible elements*, discussed e.g. in [21]. As we will see, these correspond exactly to the concept of elementary modes in the steady-state flux cone.

Lattices are sometimes also introduced as specially ordered sets. A *partial ordering* on pathways can naturally be defined by $a \leq b \Leftrightarrow a \subseteq b$. This reflects the idea that a pathway that is contained in another should be considered smaller in some sense. Because of their order-theoretical roots, many concepts in lattice theory should be understood in this context, e.g. the minimal elements, or the maximum.

The order-theoretical point of view also provides an interesting way of visualizing the relationship of different pathways via the so-called *Hasse diagram*. A Hasse diagram represents a finite, partially ordered set in a compact way. It can be seen as a directed graph with the elements of the set as nodes, and certain pairs of elements as edges. An element a_1 is connected to another element a_2 by an edge iff a_1 *is covered by* a_2, i.e., if $a_1 < a_2$ and there is no other element a with $a_1 < a < a_2$. All edges are implicitly oriented from bottom to top. In lattice-based metabolic models, we can draw a Hasse diagram where the elements are the reaction sets in our model. Two sets a_1, a_2 are connected if $a_1 \subset a_2$ and there is no other set from the model in between. The Hasse diagram of a lattice provides a lot of useful information. An element is irreducible iff it covers only one other element, i.e., there is only one edge going downwards. A reducible reaction set always covers at least two different reaction sets. Since our lattices are $\cup$-closed, it is easy to see that each reaction set that covers three or more other sets can always be written as the union of any two of those reaction sets that it covers. This allows us to identify how pathways can be decomposed into smaller reaction sets. An example is given in Fig. 2.

3 Steady-State Flux Spaces Can Be Modeled as Lattices

Constraint-based analysis of metabolic networks is based on the steady-state flux cone $C = \left\{v \in \mathbb{R}^{\mathsf{R}} \mid Sv = 0, v_{\text{Irrev}} \geq 0\right\}$, where $S \in \mathbb{R}^{\mathsf{M}\times\mathsf{R}}$ is the stoichiometric matrix over the set of metabolites M and $\text{Irrev} \subseteq \mathsf{R}$ is the set of irreversible reactions. Constraint-based methods include Flux Balance Analysis (FBA), Elementary Mode Analysis (EMA), or Flux Coupling Analysis (FCA), which allow for growth prediction, structural understanding, or target prediction in metabolic engineering [5–11].

We will show here how two of these approaches, namely EMA and FCA, may be reformulated in lattice-theoretic terms. Proving that we can work on a lattice L^C induced by the flux cone C, will allow us to use the general framework of lattice theory, which simplifies the development of optimized and unified algorithms. As a first step, we prove that any polyhedron $P \subseteq \mathbb{R}^{\mathsf{R}}$ induces a lattice. For this we look at the support of the vectors.

Proposition 1. *Given* $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$, *with* $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$, *let*

$$L^P := \{\operatorname{supp} x \mid x \in P\}$$

with $\operatorname{supp} x = \{r \in \mathbb{R} \mid x_r \neq 0\}$. *Then* $(L^P, \cup)$ *is a finite lattice.*

Proof. Let $a_1, a_2 \in L^P$ with $a_i = \operatorname{supp}\left(x^{(i)}\right)$. Define $x^{(\lambda)} = \lambda x^{(1)} + (1-\lambda)\, x^{(2)}$ for $\lambda \in [0,1]$. P is a polyhedron, thus $x^{(\lambda)} \in P$ and $\operatorname{supp}\left(x^{(\lambda)}\right) \subseteq a_1 \cup a_2$. Now we only have to show that there is $\lambda^* \in [0,1]$ with $\operatorname{supp}\left(x^{(\lambda^*)}\right) = a_1 \cup a_2$. So let us look at the cases where this equality does not hold. We have $x_i^{(\lambda)} = 0$ if and only if $\lambda x_i^{(1)} + (1-\lambda)\, x_i^{(2)} = 0$. So for each $i \in a_1 \cup a_2$ there is at most one λ such that $i \notin \operatorname{supp}\left(x^{(\lambda)}\right)$. Because there are less than $|\mathsf{R}| + 1$ values for λ with $\operatorname{supp}\left(x^{(\lambda)}\right) \subsetneq a_1 \cup a_2$, we know that the desired $\lambda^* \in [0,1]$ must exist. □

So we know that the flux cone C induces a lattice:

$$L^C := \{\operatorname{supp} v \mid Sv = 0, v_{\text{Irrev}} \geq 0\} \quad . \tag{3}$$

But we can also work on bounded flux vectors, where we assume minimal and maximal reaction rates $l, u \in \mathbb{R}^{\mathsf{R}}$:

$$L^C_{l \leq v \leq u} := \{\operatorname{supp} v \mid Sv = 0, l \leq v \leq u\} \quad . \tag{4}$$

A special case of a bounded flux space is the space of all optimal-growth flux vectors, used in FBA and studied e.g. in [18]:

$$L^C_{\text{opt}} := \{\operatorname{supp} v \mid Sv = 0, l \leq v \leq u, v_{\text{Biomass}} = \max\} \quad . \tag{5}$$

Fig. 1 shows an example network for this case. As we will see in Sect. 4, lattice theory allows us to define concepts equivalent to EFMs and FCA on these bounded flux spaces, too.
Finally, given a lattice $L \subseteq 2^{\mathsf{R}}$ and a subset $Q \subseteq \mathsf{R}$, we define

$$L_{\perp Q} := \{a \in L \mid a \cap Q = \emptyset\}\,, \tag{6}$$

$$L_Q := \{a \cap Q \mid a \in L\} \quad . \tag{7}$$

Clearly, $(L_{\perp Q}, \cup)$ resp. $(L_Q, \cup)$ satisfy the two lattice axioms from Sect. 2. Therefore, we get two new lattices, which we call L *without* Q resp. L *projected on* Q.

4 Methods

4.1 Elementary Modes in Lattices

An *elementary mode* [7] is a steady-state flux vector $v \in C$ that is *irreducible* in the sense that it cannot be written in the form $v = v^1 + v^2$, with $v^1, v^2 \in C, \operatorname{supp} v^1, \operatorname{supp} v^2 \subsetneq \operatorname{supp} v$. As proven in [8], a flux vector $v \in C \setminus \{0\}$ is irreducible if and only if $\operatorname{supp} v$ is *minimal* (w.r.t. $\subseteq$). In the context of this paper, it is interesting to note that an elementary mode is uniquely determined by its support, i.e., given two elementary modes $v, v' \in C$ with $\operatorname{supp} v = \operatorname{supp} v'$, there exists $\lambda \neq 0$ such that $v = \lambda v'$ [8].

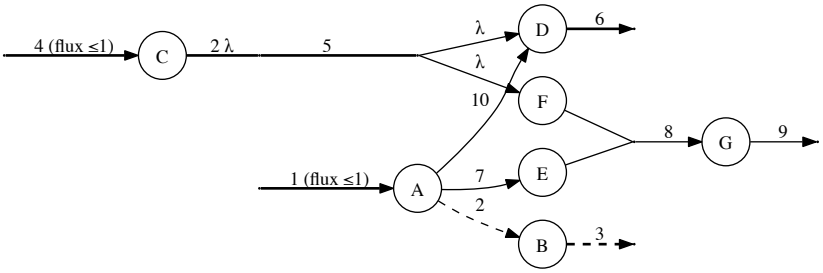

Fig. 1. Example network with metabolites $A, \ldots, G$ and reactions $1, \ldots, 10$. For example, reaction 5 corresponds to the chemical reaction $2C \to D + F$. Without constraints on the input reactions 1 and 4, none of the reactions is blocked and flux through reaction 6 is unbounded. However, if we include bounds on the input fluxes $v_1, v_4 \leq 1$, then we obtain a maximal flux of $v_6 = 1$. The corresponding optimal solution space is given by $v_1 = v_6 = 1, v_2 = v_3 = 0, v_4 = 2\lambda, v_5 = v_7 = v_8 = v_9 = \lambda, v_{10} = 1 - \lambda$ with $\lambda \in [0, 0.5]$. In particular, reactions 2 and 3 become blocked.

In general lattices, minimal and irreducible elements have to be distinguished. [21] defines two sets of lattice elements, which we write as $\mathcal{M}(L)$ and $\mathcal{I}(L)$:

$$\mathcal{M}(L) := \{e \in L \mid \forall a \in L : a \subsetneq e \Rightarrow a = 0\} \ ,$$
$$\mathcal{I}(L) := \{b \in L \mid \forall A \subseteq L : b = \bigcup_{a \in A} a \Rightarrow b \in A\} \ .$$

We call $\mathcal{M}(L)$ the set of (non-trivial) *minimal elements* of L and $\mathcal{I}(L)$ the set of *irreducible elements* of L. The irreducible elements generate the lattice, i.e., for all $a \in L$ there exist $b_1, \ldots, b_t \in \mathcal{I}(L)$ such that $a = \bigcup_{i=1}^{t} b_i$. Clearly, all minimal elements are irreducible, i.e., $\mathcal{M}(L) \subseteq \mathcal{I}(L)$. Lattices where both sets are the same are called *atomic*. While the lattice L^C is atomic, this does not hold for the lattice L^C_{opt} of all optimal-growth pathways, cf. Fig. 2. Therefore, for general lattices, the two concepts are different.

In [27] the notion of *elementary flux patterns* was introduced to describe the generating pathways through subsystems $Q \subseteq \mathsf{R}$ of a metabolic network. These may be interpreted as the set of irreducible, but not as the set of minimal elements, in a suitably defined lattice $L_Q := \{a \cap Q \mid a \in L\}$ (cf. (6)).

4.2 Lattice Maxima Give a New View on FCA

Flux coupling analysis (FCA) [10, 11, 29] studies blocked and coupled reactions in the steady-state flux cone C. It has been used for exploring a wide range of biological questions such as network evolution, gene essentiality, or gene regulation [30–35]. Here we offer an extended lattice-theoretic view of FCA, which

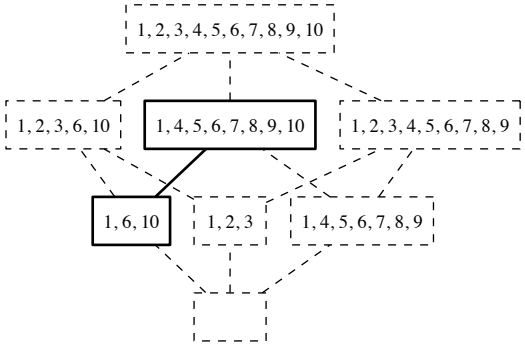

Fig. 2. Hasse diagram for the lattice L^C corresponding to the network in Fig. 1. Each possible support of a flux vector is represented by a box, the empty reaction set (zero flux) as an empty box. Reaction sets that do not represent optimal-growth flux vectors are contained in dashed boxes. In the space of optimal-growth flux vectors, there is only one minimal element: $\mathcal{M}\left(L^C_{\text{opt}}\right) = \{\{1, 6, 10\}\}$. To describe the whole lattice L^C_{opt}, we need another (non-minimal) irreducible element: $\mathcal{I}\left(L^C_{\text{opt}}\right) = \{\{1, 6, 10\}, \{1, 4, 5, 6, 7, 8, 9, 10\}\}$.

allows us to apply this tool not only on the classical flux lattice L^C (cf. (3)), but also on many other structures, such as the lattices defined in (4)-(7).

A reaction $r \in \mathsf{R}$ is *blocked*, if $v_r = 0$, for all $v \in C$. Two unblocked reactions r, s are *directionally coupled* ($r \overset{=0}{\to} s$) if $v_r = 0$ implies $v_s = 0$, for all $v \in C$, and *partially coupled* ($r \overset{=0}{\leftrightarrow} s$) if both $r \overset{=0}{\to} s$ and $s \overset{=0}{\to} r$ [10,11]. If neither $r \overset{=0}{\to} s$ nor $s \overset{=0}{\to} r$, then r, s are *uncoupled*. There is also the special case of *fully coupled* reactions, which correspond to *enzyme subsets* [36]. In the case of the flux cone, we can find those pairs using the kernel matrix [29].

Blocked and coupled reactions can be naturally defined in the more general lattice-theoretic framework. A reaction $r \in \mathsf{R}$ is *blocked* in a lattice $L \subseteq 2^{\mathsf{R}}$ if and only if $r \notin a$, for all $a \in L$. For unblocked reactions $r, s \in \mathsf{R}$, we define the *coupling* relations in L:

$$r \to s :\Leftrightarrow \forall a \in L \ : \ (r \notin a \Rightarrow s \notin a) \ ,$$
$$r \leftrightarrow s :\Leftrightarrow \forall a \in L \ : \ (r \in a \Leftrightarrow s \in a) \ .$$

Now we come back to the unique maximum 1_L in a lattice L. From (2), we know that a reaction $r \in \mathsf{R}$ is blocked in L if and only if $r \notin 1_L$. Next we look at the lattice $L_{\perp\{r\}} := \{a \in L \mid r \notin a\}$, cf. (6). Using again (2), we see that two unblocked reactions $r, s \in 1_L$ are directionally coupled if and only if s is blocked in $1_{\perp\{r\}} := 1_{L_{\perp\{r\}}}$. Therefore, we get:

Proposition 2. *Given a lattice $L \subseteq 2^{\mathsf{R}}$ and a reaction $r \in \mathsf{R}$, we have:*

$$r \text{ is blocked in } L \Leftrightarrow r \notin 1_L \ . \tag{8}$$

For two unblocked reactions $r, s \in \mathsf{R}$*, we have:*

$$r \to s \Leftrightarrow s \notin 1_{\perp\{r\}} \ . \tag{9}$$

In Sect. 5, we will give a fast algorithm for determining 1_L and $1_{\perp\{r\}}$, which will allow us to perform FCA in a simplified way.

5 Algorithms and Implementation

5.1 Finding Maxima in General Lattices

We first present an algorithm that can be used to perform FCA in any lattice-based model. It is designed in a way that it is easily adaptable to all kinds of models and still very fast. We achieve this by re-using intermediate results $a \in L_{\perp\{r\}}$, which we call collect in a set of *witnesses* $\mathcal{W}$. Using those witnesses, we search a maximum via nested intervals.

At the beginning, we do not know anything, so we assume $\mathtt{lb} = \emptyset \subseteq 1_{\perp\{r\}} \subseteq \mathtt{ub} = \mathsf{R}$ with lower and upper bounds `lb` and `ub`. Each element $a \in L_{\perp\{r\}}$ that we obtain improves the lower bound. Every time we find that there is no $a \in L_{\perp\{r\}}$ with $s \in a$, we can decrease `ub` by removing s. Finally, we get $\mathtt{lb} = \mathtt{ub}$, which is then our maximum $1_{\perp\{r\}}$.

Algorithm 1 *FCA*

$\mathcal{W} = \emptyset$
for $r \in \mathsf{R}$ ***do***
  $1_{\perp\{r\}} = \mathsf{R} \setminus \{r\}$
for $r \in \mathsf{R}$ ***do***
  $\mathcal{W}_{\perp r} = \{a \in \mathcal{W} \mid r \notin a\}$
  $\mathtt{lb} = \bigcup_{a \in \mathcal{W}_{\perp r}} a$, $\mathtt{ub} = \mathsf{R}$
  for $s \in \mathsf{R}$ ***do***
    if $s \in \mathtt{ub} \setminus \mathtt{lb}$ ***then***
      $a = \mathtt{Test}(r, s)$
      if $s \in a$ ***then***
        $\mathtt{lb} = a \cup \mathtt{lb}$, $\mathcal{W} = \{a\} \cup \mathcal{W}$
      else
        $\mathtt{ub} = \mathtt{ub} \cap 1_{\perp\{s\}}$
  $1_{\perp\{r\}} = \mathtt{ub}$

Alg. 1 uses the fact that lattices are $\cup$-closed. Therefore, we can combine each pair of already known pathways to create a new, larger feasible solution. This gives us a lower bound for the maxima $1_{\perp\{r\}}$. By keeping already calculated pathways as witnesses in $\mathcal{W}$, we get a major improvement in running time.

The algorithm does not use any specific properties of the flux cone. It is defined for any lattice-based model. To use it, we include the method `Test(r,s)` that

returns a lattice element $a \in L$ with $r \notin a \ni s$, if such an element exists, and $\emptyset$ otherwise. This method is the only part of the code depending on model-specific information or constraints.

To implement this method for traditional FCA, we can solve in `Test` the following linear program (LP) (with a trivial objective function):

$$\min \{0 \cdot v \mid Sv = 0, v_{\text{Irrev}} \geq 0, v_r = 0, v_s = \sigma\} \quad . \tag{10}$$

For reversible reactions $s \in \mathsf{R} \setminus \text{Irrev}$, this linear program has to be solved twice, i.e., $\sigma \in \Omega_s = \{1, -1\}$, for irreversible reactions $s \in \text{Irrev}$, we use $\Omega_s = \{1\}$. If we find a feasible solution $v \in L_{\perp\{r\}}$, the method `Test` returns $a = \{r \in \mathsf{R} \mid v_r \neq 0\}$, otherwise it returns $\emptyset$.

Lemma 1. *The LP* (10) *is infeasible for all* $\sigma \in \Omega_s$ *if and only if* $r \overset{=0}{\to} s$.

Proof. $\Rightarrow$**:** If r is not directionally coupled to s, there exists $a \in L^C$ s.t. $r \notin a \ni s$. Because of the definition of L^C there exists $v \in C$ with $a = \{i \in \mathsf{R} \mid v_i \neq 0\}$. Thus, $v_r = 0 \neq v_s$. Because C is a cone, v is scalable by positive scalars $\lambda > 0$. Thus, there exists a feasible solution of LP (10).

$\Leftarrow$**:** If LP (10) is feasible, it follows that we have found a flux vector v with support $a = \{i \in \mathsf{R} \mid v_i \neq 0\}$. We further know that $s \in a$, but $r \notin a$, thus r is not directionally coupled to s. □

Theorem 1. *Let* $L \subseteq 2^{\mathsf{R}}$ *be a lattice and* $\emptyset \subseteq \mathcal{W} \subseteq L$ *a list of known lattice elements (witnesses). Then Alg. 1 computes the maxima* $1_{\perp\{r\}}$ *needed for FCA (cf. Prop. 2).*

Proof. Given a reaction $r \in \mathsf{R}$, we show that Alg. 1 computes $1_{\perp\{r\}}$. Since $\mathcal{W} \subseteq L$ is a set of lattice elements, we have

$$\texttt{lb} = \bigcup_{a \in \mathcal{W}_{\perp r}} a \quad \subseteq \quad \bigcup_{a \in L_{\perp\{r\}}} a = 1_{\perp\{r\}} \quad . \tag{11}$$

Therefore, `lb` is a lower bound for $1_{\perp\{r\}}$ before we enter the inner loop. Since $L_{\perp\{r\}} \subseteq L$, we know $1_{\perp\{r\}} \subseteq 1_L$. Thus, `ub` is an upper bound before we enter the inner loop. Let $s \in \texttt{ub} \setminus \texttt{lb}$ be minimal. Let a be the result of $\texttt{Test}(r, s)$ in the inner loop. By the definition of $\texttt{Test}(r, s)$, we know that $a \in L_{\perp\{r\}}$ and, if $a \neq \emptyset$, then $s \in a$. Assume $s \in a$. Then the new $\texttt{lb} = a \cup \texttt{lb}$ is an element of $L_{\perp\{r\}}$, with $s \in \texttt{lb}$. Thus, for the next iteration, it holds that $s \notin \texttt{ub} \setminus \texttt{lb}$. Now assume $s \notin a$. This means that $s \notin 1_{\perp\{r\}}$. It follows $1_{\perp\{r\}} \subseteq 1_{\perp\{s\}}$. Since $\texttt{ub} \supseteq 1_{\perp\{r\}}$, it follows $\texttt{ub} \cap 1_{\perp\{s\}} \supseteq 1_{\perp\{r\}}$ is an upper bound. Because of the first loop in Alg. 1, we know $s \notin 1_{\perp\{s\}}$. Thus, in the next iteration, we have $s \notin \texttt{ub} \setminus \texttt{lb}$. □

Remark 1. We can accelerate Alg. 1 by replacing loops over R with loops over the set of all unblocked reactions 1_L.

Remark 2. Obviously, we can also modify the algorithm and the LP (10) to calculate this maximum 1_L of the lattice. For that, we have to replace $\texttt{Test}(r, s)$ with a method $\texttt{Test}(s)$ that does not use the constraint $v_r = 0$ in (10).

5.2 FCA in n Steps

Alg. 1 provides a method that can be used for any lattice-based model for which we can implement the method $\texttt{Test}(r, s)$. The constraints on this method are as simple as they could be: *find a pathway that goes through s but not through r, if possible.* Any $a \in L_{\perp\{r\}}$ with $s \in a$ is suitable. This simplicity is one of the many reasons why this algorithm is so easily adaptable to other lattice-based models. But, there may be cases where we can go even simpler. If there is a direct way to find the lattice maxima $1_{\perp\{r\}}$, we may compare this with Alg. 1. We will do this for classical FCA defined on the flux cone C. According to Prop. 1, the set L^C of all supports of flux vectors is indeed a lattice. That means there is a feasible flux vector $v^* \in C$ with $1_{L^C} = \{r \in \mathsf{R} \mid v^*_r \neq 0\}$. Obviously, the support of this flux vector has maximal cardinality.

Figueiredo et al. [9] introduce a mixed-integer linear program (MILP) that enumerates the (cardinality) shortest elementary modes. To achieve this, they add binary variables $a_i = 1 \Leftrightarrow v_i \neq 0$ to the LP (10). A slight variation of their MILP already provides the solution to find the lattice maximum 1_{L^C} in one single step. Since [9] is interested in finding elements of small cardinality, their objective function is $\min \sum_{i \in \mathsf{R}} a_i$. Here, we want to find an element of maximal cardinality. So we change the function to $\max \sum_{i \in \mathsf{R}} a_i$. Doing that we find the unique $a \in L$ with $a = 1_{L^C}$. For finding the maxima $1_{\perp\{r\}}$, we just have to (re-)add the single constraint $v_s = 0$ or alternatively $a_s = 0$.

5.3 Implementation

We have implemented the algorithm for general lattices in the language `C#`. Our program L4FC (Lattices for Flux Coupling) accepts files in METATOOL format [36] or separate files for stoichiometric information and irreversibility constraints. The implementation makes full use of the flexibility of lattices: The main program first computes the set of (un-)blocked reactions, before it calculates the FCA-relevant maxima $1_{\perp\{r\}}$. The calculation of those $|\mathsf{R}| + 1$ maxima is encapsulated into a separate calculator class. Our current version uses the idea of nested intervals introduced in Alg. 1. The model-specific method $\texttt{Test}(r, s)$ is implemented in form of a Gurobi model [37] that solves the LPs (10). This design allows us to include other modelling approaches in an easy and elegant way by implementing new calculator classes. The source code is available at GitHub https://github.com/goldsteiny/L4FC and is licensed under CC BY-NC-SA 3.0. The projects history and future updates will be linked to www.hoverboard.io/L4FC.

6 Discussion

We have run our program on seven widely studied genome-scale metabolic networks from the BiGG database [38] as well as the more recent reconstruction *E. coli* iJO1366 [39]. This selection is comparable to other FCA benchmarks,

Table 1. Runtime behavior of L4FC applied on 7 genome-scale metabolic networks. In addition, we report on the number of LPs solved and the number of pathways found. The computation was done into two steps: First we calculate the set of blocked reactions, then we search for the pairs of unblocked reactions that are coupled.

Model	Step	Solution size	# LPs	$\lvert\mathcal{W}\rvert$	Time (sec)
	Total		11100	4322	242.0
E. coli iJO1366	find unblocked	1718 reactions	1579	469	9.8
2583 reactions	find couples	58613 couples	9521	3853	232.2
	Total		12606	4525	219.5
E. coli iAF1260	find unblocked	1543 reactions	1518	424	8.3
2382 reactions	find couples	39260 couples	11088	4101	211.2
	Total		2485	591	6.4
H. pylori iIT341	find unblocked	436 reactions	190	44	0.3
554 reactions	find couples	62006 couples	2295	547	6.1
	Total		2203	886	8.3
M. barkeri iAF692	find unblocked	483 reactions	340	75	0.6
690 reactions	find couples	76746 couples	1863	811	7.7
	Total		4141	1699	25.3
M. tuberculosis iNJ661	find unblocked	744 reactions	497	158	1.3
1025 reactions	find couples	60750 couples	3644	1541	23.9
	Total		4329	741	9.6
S. aureus iSB619	find unblocked	465 reactions	394	65	0.5
743 reactions	find couples	30160 couples	3935	676	9.0
	Total		5189	1483	31.1
S. cerevisiae iND750	find unblocked	631 reactions	963	129	3.0
1266 reactions	find couples	15511 couples	4226	1354	28.0

e.g. [11, 29]. Table 1 summarizes the results. No calculation took longer than 4 minutes, five of them less than 40 seconds. Given these results we can conclude that the new generic algorithm L4FC has a runtime in the same order of magnitude as F2C2, the fastest dedicated tool currently available [29].

Taking a closer look at the results, we see that the calculation of the blocked reactions takes around $5 - 20\%$ of the total running time. Similar observations can be made about the number of LPs to be solved and the number of feasible reaction sets found during this first step of the program. This is remarkable, because this first phase calculates only 1 maximum, 1_L, whereas the second phase calculates $|1_L| \sim |\mathsf{R}|$ maxima. This large disproportion is a direct consequence of our use of nested intervals, where we 1) re-use all elements found in phase 1 to get better lower bounds and 2) directly apply earlier found upper bounds $1_{\perp\{s\}}$ to improve our approximation of $1_{\perp\{r\}}$ for $s < r$. Doing the iteration $\mathtt{ub} = \mathtt{ub} \cap 1_{\perp\{s\}}$ is an obvious improvement over $\mathtt{ub} = \mathtt{ub} \setminus \{s\}$, and is quite easy to understand with lattices in mind. Using this, we achieve similar run time

improvements as discussed in [29], where transitivity tables are analysed and proven.

We ran our algorithm on a machine with Intel Core i7-2600 (3.4 GHz, 4 cores, hyperthreading) and 4GB RAM. We used Gurobi 5.1 with Windows 7 Professional, Service Pack 1 (64-bit), .NET Framework 4.0.30319. As tolerance values for zero flux, we used $|v_i| \leq 10^{-8} \Rightarrow i \notin \operatorname{supp}(v)$.

7 Summary

We have shown that the concept of EFMs and FCA can be extended to general lattice-based models. Using this algebraic framework, we can now apply these methods to new classes of models. For example, we can run FCA on the space of all optimal-growth flux vectors.

We have introduced a new algorithm for computing the set of unblocked reactions 1_L and performing FCA, using only lattice properties. This allows an easy adaptation to any lattice-based model. We have further implemented the algorithm for traditional FCA of the flux cone and shown on a benchmark set of genome-scale metabolic networks like *E. coli* iJO1366 that our generic tool L4FC is comparable in speed to dedicated FCA algorithms.

Acknowledgement. This work was funded by the Gerhard C. Starck Stiftung in terms of a PhD stipend. We thank Alexandra Grigore and Arne C. Müller for their comments on this paper and many helpful discussions.

References

1. Papin, J.A., Stelling, J., Price, N.D., Klamt, S., Schuster, S., Palsson, B.O.: Comparison of network-based pathway analysis methods. Trends in Biotechnology 22(8), 400–405 (2004)
2. Terzer, M., Maynard, N.D., Covert, M.W., Stelling, J.: Genome-scale metabolic networks. Wiley Interdiscip. Rev. Syst. Biol. Med. 1(3), 285–297 (2009)
3. Schellenberger, J., Que, R., Fleming, R.M.T., Thiele, I., Orth, J.D., Feist, A.M., Zielinski, D.C., Bordbar, A., Lewis, N.E., Rahmanian, S., Kang, J., Hyduke, D.R., Palsson, B.O.: Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0. Nature Protocols 6(9), 1290–1307 (2011)
4. Lewis, N.E., Nagarajan, H., Palsson, B.: Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nat. Rev. Microbiol. 10(4), 291–305 (2012)
5. Varma, A., Palsson, B.: Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. Nature Biotechnology 12(10), 994–998 (1994)
6. Orth, J.D., Thiele, I., Palsson, B.O.: What is flux balance analysis? Nature Biotechnology 28(3), 245–248 (2010)
7. Schuster, S., Hilgetag, C.: On elementary flux modes in biochemical reaction systems at steady state. Journal of Biological Systems 2(2), 165–182 (1994)
8. Schuster, S., Hilgetag, C., Woods, J.H., Fell, D.A.: Reaction routes in biochemical reaction systems: Algebraic properties, validated calculation procedure and example from nucleotide metabolism. J. Math. Biol. 45, 153–181 (2002)

9. de Figueiredo, L.F., Podhorski, A., Rubio, A., Kaleta, C., Beasley, J.E., Schuster, S., Planes, F.J.: Computing the shortest elementary flux modes in genome-scale metabolic networks. Bioinformatics 25(23), 3158–3165 (2009)
10. Burgard, A.P., Nikolaev, E.V., Schilling, C.H., Maranas, C.D.: Flux Coupling Analysis of Genome-Scale Metabolic Network Reconstructions. Genome Research 14(2), 301–312 (2004)
11. David, L., Marashi, S.A., Larhlimi, A., Mieth, B., Bockmayr, A.: FFCA: a feasibility-based method for flux coupling analysis of metabolic networks. BMC Bioinformatics 12(1), 236 (2011)
12. Jensen, P., Lutz, K., Papin, J.: TIGER: toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. BMC Syst. Biol. 5, 147 (2013)
13. Covert, M.W., Schilling, C.H., Palsson, B.: Regulation of Gene Expression in Flux Balance Models of Metabolism. J. Theoretical Biology 213(1), 73–88 (2001)
14. Shlomi, T., Eisenberg, Y., Sharan, R., Ruppin, E.: A genome-scale computational study of the interplay between transcriptional regulation and metabolism. Molecular Systems Biology 3, 101 (2007)
15. Jungreuthmayer, C., Ruckerbauer, D.E., Zanghellini, J.: regEfmtool: Speeding up elementary flux mode calculation using transcriptional regulatory rules in the form of three-state logic. BioSystems 113(1), 37–39 (2013)
16. Beard, D.A., Babson, E., Curtis, E., Qian, H.: Thermodynamic constraints for biochemical networks. Journal of Theoretical Biology 228, 327–333 (2004)
17. Müller, A.C., Bockmayr, A.: Fast thermodynamically constrained flux variability analysis. Bioinformatics 29(7), 903–909 (2013)
18. Kelk, S.M., Olivier, B.G., Stougie, L., Bruggeman, F.J.: Optimal flux spaces of genome-scale stoichiometric models are determined by a few subnetworks. Scientific Reports 2, 580 (2012)
19. Carbonell, P., Fichera, D., Pandit, S.B., Faulon, J.L.: Enumerating metabolic pathways for the production of heterologous target chemicals in chassis organisms. BMC Systems Biology 6(1), 10 (2012)
20. Birkhoff, G.: Lattices and their applications. Bulletin of the American Mathematical Society 44(12), 793–801 (1938)
21. Davey, B.A., Priestley, H.A.: Introduction to lattices and order. Cambridge University Press (1990)
22. Cousot, P., Cousot, R.: Abstract interpretation: a unified lattice model for static analysis of programs by construction or approximation of fixpoints. In: Fourth Annual ACM Symposium on Principles of Programming Languages, Los Angeles, pp. 238–252. ACM Press (1977)
23. Oles, F.J.: An application of lattice theory to knowledge representation. Theoretical Computer Science 249(1), 163–196 (2000)
24. Garg, V., Mittal, N., Sen, A.: Applications of lattice theory to distributed computing. ACM SIGACT Notes (2003)
25. John, M., Nebut, M., Niehren, J.: Knockout Prediction for Reaction Networks with Partial Kinetic Information. In: Giacobazzi, R., Berdine, J., Mastroeni, I. (eds.) VMCAI 2013. LNCS, vol. 7737, pp. 355–374. Springer, Heidelberg (2013)
26. Nuño, J.C., Sánchez-Valdenebro, I., Pérez-Iratxeta, C., Meléndez-Hevia, E., Montero, F.: Network organization of cell metabolism: monosaccharide interconversion. The Biochemical Journal 324, 103–111 (1997)
27. Kaleta, C., de Figueiredo, L.F., Schuster, S.: Can the whole be less than the sum of its parts? Pathway analysis in genome-scale metabolic networks using elementary flux patterns. Genome Research 19(10), 1872–1883 (2009)

28. Nation, J.B.: Revised Notes on Lattice Theory (2012), http://www.math.hawaii.edu/~jb/
29. Larhlimi, A., David, L., Selbig, J., Bockmayr, A.: F2C2: a fast tool for the computation of flux coupling in genome-scale metabolic networks. BMC Bioinformatics 13(75), 57 (2012)
30. Notebaart, R.A., Teusink, B., Siezen, R.J., Papp, B.: Co-regulation of metabolic genes is better explained by flux coupling than by network distance. PLoS Comput. Biol. 4, e26 (2008)
31. Notebaart, R.A., Kensche, P.R., Huynen, M.A., Dutilh, B.E.: Asymmetric relationships between proteins shape genome evolution. Genome Biol. 10, R19 (2009)
32. Pál, C., Papp, B., Lercher, M.J.: Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. Nat. Genet. 37, 1372–1375 (2005)
33. Yizhak, K., Tuller, T., Papp, B., Ruppin, E.: Metabolic modeling of endosymbiont genome reduction on a temporal scale. Mol. Syst. Biol. 7, 479 (2011)
34. Montagud, A., Zelezniak, A., Navarro, E., de Córdoba, P.F., Urchueguía, J.F., Patil, K.R.: Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium *Synechocystis* sp. PCC6803. Biotechnol. J. 6, 330–342 (2011)
35. Szappanos, B., Kovács, K., Szamecz, B., Honti, F., Costanzo, M., Baryshnikova, A., Gelius-Dietrich, G., Lercher, M., Jelasity, M., Myers, C., Andrews, B., Boone, C., Oliver, S., Pál, C., Papp, B.: An integrated approach to characterize genetic interaction networks in yeast metabolism. Nat. Genet. 43(7), 656–662 (2011)
36. Pfeiffer, T., Sanchez-Valdenebro, I., Nuno, J., Montero, F., Schuster, S.: METATOOL: for studying metabolic networks. Bioinformatics 15(3), 251–257 (1999)
37. Gurobi Optimization Inc: Gurobi 5.1 (2012)
38. Schellenberger, J., Park, J.O., Conrad, T.M., Palsson, B.O.: BiGG: A Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. BMC Bioinformatics 11(213), 213 (2010)
39. Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., Palsson, B.: A comprehensive genome-scale reconstruction of Escherichia coli metabolism-2011. Molecular Systems Biology 7(535) (2011)