A Lattice-Theoretic Framework for Metabolic Pathway Analysis

Yaron A. B. Goldstein
Alexander Bockmayr

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

April 2013

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, lattice theory

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 have been widely applied, various ideas have been explored of how to modify or extend the underlying model. A lot of research concerns the question of how to include regulatory information into the metabolic model (e.g. [12, 13]). Additionally, there has been a discussion on whether stronger thermodynamical constraints should be applied [14, 15]. Some try to combine the idea of FBA to analyse optimal-yield 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 [16]. Still other approaches give up the steady-state assumption and use completely different modelling approaches, e.g. hyperpaths [17].

In this paper, we want to establish the algebraic framework of lattices as a unifying approach to metabolic pathway analysis. Finite lattices [18, 19] are some of the simplest algebraic structures, but they have proven to be useful in many applications, such as knowledge representation [20], or distributed computing [21]. 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.

Here we will introduce lattice-based Elementary Mode Analysis (EMA) and a very fast Flux Coupling Analysis (FCA) method. Our implementation L4FC (Lattices for Flux Coupling) can be used for traditional, flux-cone-based FCA. But it also allows usage of other lattice-based modelling approaches, just by changing a single 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. We will see that lattice-based modelling is fully compatible with the traditional steady-state flux cone and many of its extensions. However, 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* [22] or *flux patterns* [23]. 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* [19] 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 [24], 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 [19]. 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*. This allows us to identify how pathways can be decomposed into smaller reaction sets. An example for this 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.

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*

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

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 $x^{(\lambda)} = \lambda x^{(i)} + (1 - \lambda)\, x^{(2)}$ 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/maximal reaction rates.

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

A special case of a bounded flux space is the space of all optimal-yield flux vectors, used in FBA and studied e.g. in [16]:

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

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.

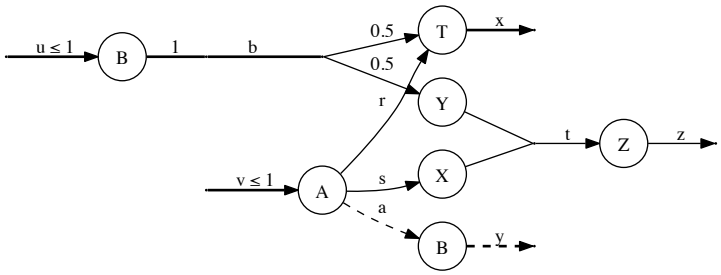

Figure 1: Maximize the flux through x when $u, v \leq 1$.

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].

In general lattices, minimal and irreducible elements have to be distinguished. [19] 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-yield pathways, cf. Fig. 2. Therefore, for general lattices, the two concepts are different.

In [23] the concept 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).

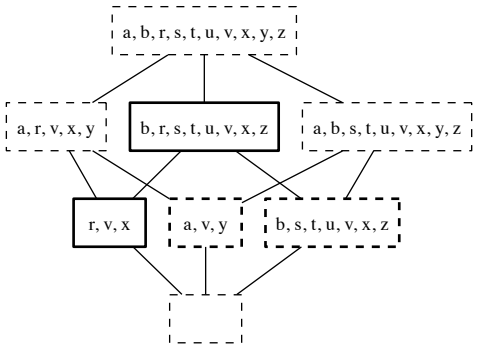

Figure 2: Hasse diagram for the lattice L^C corresponding to the network in Fig. 1. In the space of optimal-yield flux distributions, there is only one minimal element $\mathcal{M}\left(L^C_{\text{opt}}\right) = \{\{r, v, x\}\}$. To describe the whole system L^C_{opt}, we need another, non-minimal irreducible element: $\mathcal{I}\left(L^C_{\text{opt}}\right) = \{\{r, v, x\}, \{b, r, s, tu, v, x\}\} = L^C_{\text{opt}}$

4.2 Lattice Maxima Give a New View on FCA

Flux coupling analysis (FCA) [10, 11, 25] studies blocked and coupled reactions in the steady-state flux cone C. 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* ($r \overset{=0}{\rightarrow} s$) if $v_r = 0$ implies $v_s = 0$, for all $v \in C$, and *partially coupled* ($r \overset{=0}{\leftrightarrow} s$) if both $r \overset{=0}{\rightarrow} s$ and $s \overset{=0}{\rightarrow} r$. If neither $r \overset{=0}{\rightarrow} s$ nor $s \overset{=0}{\rightarrow} r$, then r, s are *uncoupled*. There is also the special case of *fully coupled* reactions, which correspond to *enzyme subsets* [26]. In the case of the flux cone, we can find those pairs using the kernel matrix [25].

Blocked and coupled reactions can be naturally defined in the more general lattice-theoretic framework. A reaction $r \in \mathsf{R}$ is *blocked* in L 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

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

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 $L \subseteq L^{\mathsf{R}}$ 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 & 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 *witnesses*. 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*

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
        $\mathtt{ub} = \mathtt{ub} \cap 1_{\perp\{s\}}$
  $1_{\perp\{r\}} = \mathtt{lb}$

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 $1_{\perp\{r\}}$. 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* $\mathcal{W} \subseteq L$ *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

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

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 progam (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 `C#`. Our program L4FC (Lattices for Flux Coupling) accepts files in METATOOL format [26] 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 [27] 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 under CC BY-NC-SA 3.0 and can be downloaded at https://github.com/goldsteiny/L4FC.git.

6 Discussion

We have run our program on seven genome-scale metabolic networks from the BiGG database [28]. The selection is comparable to other FCA benchmarks, e.g. [11, 25]. Table 1 summarizes the results. No calculation took longer than 4 minutes, five of them less than 40 seconds. That puts the new L4FC in the same order of magnitude as F2C2, the fastest tool currently available [25].

Taking a closer look at the result, we see that for all of our models only around 2/3 of the reactions are unblocked (1718/2583 in the case of *E. coli* iJO1366). Calculating these takes around 5-10% of the total running time. Interestingly, about 10% of all LPs are solved during this step, and about 10% of the witnesses are found. This is remarkable, because this first phase calculates only 1 maximum, 1_L, whereas the second phase calculates $|1_L| \sim 2/3\,|\mathsf{R}|$ 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 $\texttt{ub} = \texttt{ub} \cap 1_{\perp\{s\}}$ is an obvious improvement over $\texttt{ub} = \texttt{ub} \setminus \{s\}$ and is quite easy to understand with lattices in mind. Using it, we achieve similar run time improvements as discussed in [25], 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 $|v_i| \leq 10^{-8} \Rightarrow i \notin \operatorname{supp}(v)$.

Model	Value	Number	LPs	$\lvert\mathcal{W}\rvert$	Time (sec)
E. coli iJO1366	Total	2583	11100	4322	242.0
	1_L	1718	1579	469	9.8
	coupled	58613	9521	3853	232.2
E. coli iAF1260	Total	2382	12606	4525	219.5
	1_L	1543	1518	424	8.3
	coupled	39260	11088	4101	211.2
H. pylori iIT341	Total	554	2485	591	6.4
	1_L	436	190	44	0.3
	coupled	62006	2295	547	6.1
M. barkeri iAF692	Total	690	2203	886	8.3
	1_L	483	340	75	0.6
	coupled	76746	1863	811	7.7
M. tuberculosis iNJ661	Total	1025	4141	1699	25.3
	1_L	744	497	158	1.3
	coupled	60750	3644	1541	23.9
S. aureus iSB619	Total	743	4329	741	9.6
	1_L	465	394	65	0.5
	coupled	30160	3935	676	9.0
S. cerevisiae iND750	Total	1266	5189	1483	31.1
	1_L	631	963	129	3.0
	coupled	15511	4226	1354	28.0

Table 1: Run time results for L4FC applied on 7 genome wide networks.

7 Summary

We have shown that the concept of EMs 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-yield 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 tool L4FC is comparable in speed to much more elaborated FCA algorithms.

In a follow-up paper, we will extend our framework to models that do not directly satisfy the lattice theory axioms.

Acknowledgement

This work was funded by the Gerhard C. Starck Stiftung in terms of a PhD stipend.

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) (2004) 400–405

[2] Terzer, M., Maynard, N.D., Covert, M.W., Stelling, J.: Genome-scale metabolic networks. Wiley Interdiscip Rev Syst Biol Med **1**(3) (2009) 285–297

[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.b.: Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nature Protocols **6**(9) (2011) 1290–1307

[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) (2012) 291–305

[5] Varma, A., Palsson, B.: Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. Nature Biotechnology **12**(10) (1994) 994–998

[6] Orth, J.D., Thiele, I., Palsson, B.O.: What is flux balance analysis? Nature Biotechnology **28**(3) (2010) 245–8

[7] Schuster, S., Hilgetag, C.: On elementary flux modes in biochemical reaction systems at steady state. Journal of Biological Systems **2**(2) (1994) 165

[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** (2002) 153–181

[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) (2009) 3158–3165

[10] Burgard, A.P., V., N.E., Schilling, C.H., Maranas, C.D., Nikolaev, E.V.: Flux Coupling Analysis of Genome-Scale Metabolic Network Reconstructions. Genome Research **14**(2) (2004) 301–312

[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) (2011) 236

[12] Covert, M.W., Schilling, C.H., Palsson, B.: Regulation of Gene Expression in Flux Balance Models of Metabolism. J Theoretical Biology **213**(1) (2001) 73–88

[13] Jensen, P., Lutz, K., Papin, J.: TIGER: toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. BMC Syst Biol. **5** (2013) 147

[14] Beard, D.A., Babson, E., Curtis, E., Qian, H.: Thermodynamic constraints for biochemical networks. Journal of Theoretical Biology **228** (2004) 327–333

[15] Müller, A.C., Bockmayr, A.: Fast thermodynamically constrained flux variability analysis. Bioinformatics **29**(7) (2013) 903–909

[16] 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** (2012) 580

[17] 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) (2012) 10

[18] Birkhoff, G.: Lattices and their applications. Bulletin of the American Mathematical Society **44**(12) (1938) 793–801

[19] Davey, B.A., Priestley, H.A.: Introduction to lattices and order. 1 edn. Cambridge University Press (1990)

[20] Oles, F.J.: An application of lattice theory to knowledge representation. Theoretical Computer Science **249**(1) (2000) 163–196

[21] Garg, V., Mittal, N., Sen, A.: Applications of lattice theory to distributed computing. ACM SIGACT Notes (2003)

[22] 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** (1997) 103–11

[23] 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) (2009) 1872–83

[24] Nation, J.B.: Revised Notes on Lattice Theory (2012) `www.math.hawaii.edu/~jb/`.

[25] 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) (2012) 57

[26] Pfeiffer, T., Sanchez-Valdenebro, I., Nuno, J., Montero, F., Schuster, S.: METATOOL: for studying metabolic networks. Bioinformatics **15**(3) (1999) 251–257

[27] Gurobi Optimization Inc: Gurobi 5.1 (2012)

[28] 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) (2010) 213