Modeling Signal Transduction in Enzyme Cascades with the Concept of Elementary Flux Modes

JÖRN BEHRE and STEFAN SCHUSTER

ABSTRACT

Concepts such as elementary flux modes (EFMs) and extreme pathways are useful tools in the detection of non-decomposable routes (metabolic pathways) in biochemical networks. These methods are based on the fact that metabolic networks obey a mass balance condition. In signal transduction networks, that condition is of minor importance because it is the flow of information that matters. Nevertheless, it would be interesting to apply pathway detection methods to signaling systems. Here, we present a formalism by which this can be achieved in the case of enzyme cascades operating, for example, by phosphorylation and dephosphorylation. It is based on the ideas that the signal is not diminished along each route and that the system has to return to its original state after each signaling event. We illustrate the method by several simple prototypic single-phosphorylation and double-phosphorylation cascades, including convergent and divergent branching. Moreover, it is applied to a specific example from insulin signaling. (See online Supplementary Material at www.liebertonline.com.)

Key words: double phosphorylation, elementary flux modes, insulin signaling network, kinase cascades, signal transduction networks.

1. INTRODUCTION

Signal transduction is essentially important for all biological functions in living cells. The modeling of signaling networks has recently attracted considerable interest (Steffen et al., 2002; Papin and Palsson, 2004; Xiong et al., 2004; Zevedei-Oancea and Schuster, 2005; Klamt et al., 2006; Sackmann et al., 2007; Arga et al., 2007). In metabolic networks, mass flow is a crucial aspect, while in signal transduction networks, it is the flow of information that matters. Furthermore, metabolic networks usually subsist in steady states, a fact that is used in the modeling of these networks (Savageau, 1976; Heinrich and Schuster, 1996; Schilling et al., 2000; Klamt, 2006). In contrast, in signaling systems, time-dependent signals (e.g., short pulses) occur. Several approaches to model structural properties of signal transduction networks have been presented, for example, by Papin and Palsson (2004) and Xiong et al. (2004). These analyses used the concept of extreme pathways, introduced by Schilling et al. (2000) to model metabolic networks at steady states. This method is based on the earlier concept of extreme currents (Clarke, 1981). A further method for
modeling signaling systems is based on the concept of minimal T-invariants, which had been introduced in the theory of Petri nets, first for non-biological, technological applications (Starke, 1990). Several studies using this concept have been made by Heiner et al. (2004) and Sackmann et al. (2006, 2007). Extreme pathways and minimal T-invariants are similar to the concept of elementary flux modes (Schuster and Hilgetag, 1994; Schuster et al., 1999). The latter concept has widely been used in the structural modeling of metabolic networks (Stelling et al., 2002; Schwender et al., 2004; Schwartz et al., 2007; Behre et al., 2008). A comparison between minimal T-invariants and elementary flux modes and, in general, the modeling by Petri nets with traditional modeling in biochemistry has been given by Zevedei-Oancea and Schuster (2003), and a comparison between extreme pathways and elementary flux modes, by Papin et al. (2004).

Another approach for modeling signal transduction networks are logical interaction hypergraphs, which were introduced by Klamt et al. (2006). This approach reduces the processes on the molecular level to a Boolean perspective. A limitation of that method is that reactions with more than one product are difficult to model.

General difficulties in modeling signal transduction networks arise because the flow of information does not necessarily have the same direction as the mass flow that is required to transport the information. In kinase cascades, for example, the mass flow consists of cycles of enzymes being phosphorylated and dephosphorylated while the information is transduced through the cascade. In some other systems, the phosphate moiety is transferred along the reaction chain, for example, in the phosphotransferase system (PTS). In that case, the flow of information corresponds to the flow of mass, so that methods from metabolic modeling can be used in a straightforward way as shown by Papin and Palsson (2004) using a prototypic example. In the vast majority of signaling systems, however, mass flow does not correspond to information flow. Nevertheless, several authors have applied these methods formally to such systems (Xiong et al., 2004; Heiner et al., 2004; Sackmann et al., 2006, 2007). This has led to meaningful results, although the theoretical foundation for this approach is not really clear.

The present article deals with the theoretical justification for applying methods of structural analysis of metabolic systems to signal transduction networks composed of enzyme cascades. We propose an explanation based on the reasoning that (i) such networks usually allow signal amplification, that is, the signal amplitude must at least remain constant, and (ii) that, after the signal has been transmitted, the system must regain its original state so that a steady state can be assumed in the sense of an average over longer time spans. When discussing enzyme cascades, we take phosphorylation as a prototypic covalent modification. Nevertheless, any other form of chemical modification switching on or off the activity of the enzyme—such as methylation, acetylation, adenylylation, or ubiquitination—can be described by the same formalism. Even phosphorylation can occur in various ways depending on which nucleotide phosphate is used. Here, we write NTP and NDP as unspecified symbols for nucleotide triphosphates and diphosphates, with ATP and ADP being the most common representatives.

2. ASSUMING STEADY STATE FOR SIGNAL TRANSDUCTION NETWORKS

As mentioned above, in metabolic networks, mass flow is the important aspect, while in signal transduction networks, the flow of information is more important. Of course, in signaling networks, mass flow exists as well, but its direction is not necessarily the same as that of the information flow. This is particularly obvious in kinase cascades. Upon entry of an initial signal, the first kinase $E_1$ of the cascade is phosphorylated (leading to $E_1P$). Let us assume that it thus becomes active, which is indeed the most common case. Now it catalyzes the phosphorylation of the next kinase and so on. After some time, $E_1P$ will become inactive again by dephosphorylation. Each phosphorylated kinase can catalyze the phosphorylation of the next kinase several times without being consumed. Therefore, there is no mass flow from one kinase to the next while information is transferred. This can even lead to amplification of the signal.

An important assumption made for modeling metabolic networks is the steady-state condition. It is crucial for calculating elementary flux modes (Schuster and Hilgetag, 1994), extreme pathways (Schilling et al., 2000), and minimal T-invariants (Heiner et al., 2004). In signaling systems, by contrast, time-dependent signals occur, for example, short pulses. Nevertheless, there is an important fact that allows applying the steady-state condition also to signaling systems: these systems must be regenerated after each signaling event. Let us consider, for example, the above-mentioned signaling cascade again. Because of the dephosphorylation after the pulse, the concentration of $E_1$ returns to its initial state. Averaged over the time
period that is needed for phosphorylation and dephosphorylation of $E_1$, its concentration stays constant during repeated phosphorylation/dephosphorylation cycles.

On the other hand, the steady state is just an approximation rather than a strict condition. Also in metabolic networks, fluctuations of concentrations occur (e.g., due to diffusion), but they are usually neglected because they average out (Segel, 1993; Heinrich and Schuster, 1996). Therefore, sometimes the term “pseudo-steady-state” is used.

In line of the above reasoning, it can be argued that, for detecting potential routes on which pulses can be propagated across the network, it is sufficient to detect the routes on which a flow can subsist at steady state. This can be understood by a hydrodynamic analog. A flood wave on a river can usually go only along the course on which that river flows at constant water level. If one or more rivers form a network by confluence of rivers or divergence in estuaries, the flood waves can go on the various branches of the network. An exception is when the wave is so high that the water flows over the river bank. Then, new routes can be used. This is, in signaling networks, somewhat analogous to a flow along side branches that are normally at equilibrium.

3. APPLYING THE CONCEPT OF ELEMENTARY FLUX MODES TO ENZYME CASCADES

Motivated by the above reasoning, we here apply the concept of elementary flux modes (EFMs) to enzyme cascades.

A mathematically detailed definition of EFMs was given by Schuster et al. (2002). Here, we briefly recapitulate their most important aspects for the sake of better understanding. An EFM is a minimal set of enzymes that can operate at steady state with all irreversible reactions used in the appropriate direction. All flux distributions in the living cell are non-negative linear combinations of EFMs.

The steady-state condition is given by the equation system

$$NV(S) = 0$$ (1)

with $N$ being the stoichiometric matrix and $V(S)$ being the vector containing all fluxes in the system. Together, with the sign restriction for irreversible fluxes

$$V^{irr} \geq 0$$ (2)

this gives a linear equation/inequality system. For any flux vector $V$ with elements $v_i$, let

$$S(V) = \{i : v_i = 0\}$$ (3)

Each EFM $V^*$ fullfills conditions (1) and (2) as well as the following non-decomposability condition. There exists no vector $V^{**}$ (unequal to the null vector) that obeys conditions (1) and (2), and contains zero components wherever $V^*$ does and in at least one additional position,

$$S(V^*) \subset S(V^{**})$$ (4)

The solution of equation system (1, 2) is a convex polyhedral cone in flux space. The edges of this cone correspond to EFMs. Additional EFMs may lie inside this cone (Schuster et al., 2002).

For computing EFMs, we use the program Metatool, developed earlier in our group (von Kamp and Schuster, 2006).

3.1. Single-phosphorylation cascades

When the phosphorylated form of some enzyme, $E_i P$, is to convey a signal to the next level, it must catalyze the phosphorylation of at least one molecule of $E_{i+1}$, and can then be dephosphorylated. Of course, it can also catalyze the phosphorylation of more than one molecule of $E_{i+1}$. Then, amplification occurs. The reaction equation can be written as

$$E_i P + n E_{i+1} + n \text{NTP} = E_i + P + n E_{i+1} P + n \text{NDP}$$ (5)
Here, we have combined the phosphorylation of \( E_{i+1} \) and the subsequent dephosphorylation of \( E_i \) into one reaction. We have \( n \geq 1 \), because diminution of the signal is irrelevant. The minimum is phosphorylation of one molecule. Considering, for the moment, only this lower limit, we can write

\[
E_i P + E_{i+1} + NTP = E_i + P + E_{i+1}P + NDP \tag{6}
\]

This reaction equation applies to all levels \( i \) of the cascade. The application of this reaction equation to a single-phosphorylation cascade comprising the kinases \( E_1 \)–\( E_5 \) gives rise to one EFM, as expected. Its overall reaction is:

\[
E_1 P + E_5 + 4 NTP = E_1 + E_3 P + 4 NDP + 4 P \tag{7}
\]

The corresponding Metatool input and output is given in the Supplementary Material (see online supplementary material at www.liebertonline.com).

An important case is where some enzyme \( E_i \) can phosphorylate two different kinases, \( F_1 \) and \( G_1 \), so that a branching in the cascade occurs. Then, there are two minimum situations: either \( E_i P \) activates exactly one \( F_1 \) or exactly one \( G_1 \) before being dephosphorylated. Thus, reactions (8a,b) occur:

\[
E_i P + F_1 + NTP = E_i + P + F_1P + NDP \tag{8a}
\]

\[
E_i P + G_1 + NTP = E_i + P + G_1P + NDP \tag{8b}
\]

We have modeled a diverging cascade comprising the kinases \( E_1\)–\( E_3 \), \( F_1\)–\( F_3 \), and \( G_1\)–\( G_3 \). The cascade is shown in Figure 1 with reactions (8a,b) depicted by dash-dotted arrows. The corresponding Metatool input is given in Table 1. This cascade contains an OR-connection at the branching point because the phosphorylations of \( F_1 \) and \( G_1 \) are not necessarily coupled (see eqs. 8a,b).

As one would expect for an OR-connection at the branching point, this model comprises two EFMs. The first one consists of the reactions \( R_1, R_2, R_3, R_4, \) and \( R_5 \), the second one of \( R_1, R_2, R_6, R_7, \) and \( R_8 \). These reaction lists already indicate the routes of information flow. The corresponding overall reactions are

\[
E_1 P + F_3 + 5 NTP = E_1 + F_3P + 5 NDP + 5 P \tag{9a}
\]

for the first EFM, and for the second one

\[
E_1 P + G_3 + 5 NTP = E_1 + G_3P + 5 NDP + 5 P \tag{9b}
\]

### Table 1. Metatool Input for the OR-Connected Diverging Cascade Shown in Figure 1

| -ENZREV |
| -ENZREV |
| R1 R2 R3 R4 R5 R6 R7 R8 |
| -METINT |
| E2 E2P E3 E3P F1 F1P F2 F2P G1 G1P G2 G2P |
| -METEXT |
| NDP NTP E1 E1P F3 F3P G3 G3P P |
| -CAT |
| R1 : 1 NTP + 1 E2 + 1 E1P = 1 E2P + 1 NDP + 1 E1 + 1 P |
| R2 : 1 NTP + 1 E3 + 1 E2P = 1 E3P + 1 NDP + 1 E2 + 1 P |
| R3 : 1 NTP + 1 F1 + 1 E3P = 1 F1P + 1 NDP + 1 E3 + 1 P |
| R4 : 1 NTP + 1 F2 + 1 F1P = 1 F2P + 1 NDP + 1 F1 + 1 P |
| R5 : 1 NTP + 1 F3 + 1 F2P = 1 F3P + 1 NDP + 1 F2 + 1 P |
| R6 : 1 NTP + 1 G1 + 1 E3P = 1 G1P + 1 NDP + 1 E3 + 1 P |
| R7 : 1 NTP + 1 G2 + 1 G1P = 1 G2P + 1 NDP + 1 G1 + 1 P |
| R8 : 1 NTP + 1 G3 + 1 G2P = 1 G3P + 1 NDP + 1 G2 + 1 P |
The same routes are relevant when signal amplification occurs, that is, when \( n > 1 \). Even if one molecule of \( E_1 \) catalyzes the phosphorylation of more than one molecule of \( F_1 \) and of more than one molecule of \( G_1 \), then the signal flow from \( E_1 \) to \( F_3 \) and \( G_3 \) can be decomposed into two elementary routes: from \( E_1 \) to \( F_3 \) and from \( E_1 \) to \( G_3 \).

The complete Metatool output is given in the Supplementary Material (see online supplementary material at www.liebertonline.com). In the following, we will show some selected Metatool inputs in the text. All other inputs and outputs are given in the Supplementary Material (see online supplementary material at www.liebertonline.com).

In the case that \( E_3P \) can catalyze the phosphorylation of \( F_1 \) only when \( G_1 \) is also phosphorylated, the diverging cascade is AND-connected. For this case, we modified the model above by merging the reactions (8a) and (8b) to the new reaction:

\[
2 \text{NTP} + F_1 + G_1 + E_3P = F_1P + G_1P + 2 \text{NDP} + E_3 + P
\]  

(10)

The cascade is shown in Figure 1 with reaction (10) depicted by a dotted arrow.

As expected, only one EFM is calculated. It consists of all reactions in the model, and its overall reaction reads as follows:

\[
E_1P + F_3 + G_3 + 8 \text{NTP} = E_1 + F_3P + G_3P + 8 \text{NDP} + 7 \text{P}
\]  

(11)

Also, for converging cascades, one can distinguish between OR-connection and AND-connection. Thus, we also model these two variants.

The converging cascade with OR-connection is shown in Figure 2 with reactions (12a,b) depicted by dash-dotted arrows. The corresponding Metatool input is given in Table 2. The OR-connection is given by the two reactions:

\[
\text{NTP} + G_1 + E_3P = G_1P + \text{NDP} + E_3 + P \quad (12a)
\]

\[
\text{NTP} + G_1 + F_3P = G_1P + \text{NDP} + F_3 + P \quad (12b)
\]
Again, the OR-connection gives rise to two EFM s, as expected. The first EFM consists of the reactions R\textsubscript{1}, R\textsubscript{2}, R\textsubscript{5}, R\textsubscript{7}, and R\textsubscript{8}, and the second one of R\textsubscript{3}, R\textsubscript{4}, R\textsubscript{6}, R\textsubscript{7}, and R\textsubscript{8}. The corresponding overall reactions are

\[
E_1P + G_3 + 5 \text{NTP} = E_1 + G_3P + 5 \text{NDP} + 5 \text{P}
\]

for the first EFM, and for the second one

\[
F_1P + G_3 + 5 \text{NTP} = F_1 + G_3P + 5 \text{NDP} + 5 \text{P}
\]

The converging cascade with AND-connection is shown in Figure 2 with reaction (14) depicted by a dotted arrow. Analogously to the diverging cascade with AND-connection, here also the branching point is represented by one reaction:

\[
\text{NTP} + G_1 + E_3P + F_3P = G_1P + \text{NDP} + E_3 + F_3 + 2 \text{P}
\]

Again, the AND-connection leads to only one EFM, as expected. The EFM consists of all reactions in the model, and its overall reaction looks as follows:

\[
E_1P + F_1P + G_3 + 7 \text{NTP} = E_1 + F_1 + G_3P + 7 \text{NDP} + 8 \text{P}
\]

3.2. Double-phosphorylation cascades

A further motif in signal transduction networks are double-phosphorylation cascades. In such cascades, an enzyme E\textsubscript{i+1} needs to be phosphorylated twice to become active.

\[
E_{i+1} + \text{NTP} = E_{i+1}P + \text{NDP}
\]

\[
E_{i+1}P + \text{NTP} = E_{i+1}PP + \text{NDP}
\]

These two phosphorylation steps are often catalyzed by the same preceding enzyme, just as the two subsequent dephosphorylation steps are usually catalyzed by the same phosphatase. Without any kinetic
constants, the exact chronology of phosphorylation and dephosphorylation, and thus the amplification of the signal along the cascade, cannot be determined. But again a minimal precondition to transduce a signal can be identified. That is, we can assume that an enzyme $E_i$ is double-phosphorylated and stays in this active state until the next enzyme $E_{i+1}$ is phosphorylated twice. Therefore, we can merge the two phosphorylation steps and set up eq. (17) analogously to eq. (5).

$$E_{i\text{PP}} + n E_{i+1} + 2n \text{NTP} = E_i + 2P + n E_{i+1\text{PP}} + 2n \text{NDP} \quad (17)$$

Again, we have combined the phosphorylation of $E_{i+1}$ and the subsequent dephosphorylation of $E_i$ into one reaction. In the minimum case of no signal amplification, $n = 1$, we get:

$$E_{i\text{PP}} + E_{i+1} + 2 \text{NTP} = E_i + 2P + E_{i+1\text{PP}} + 2 \text{NDP} \quad (18)$$

This reaction equation applies to all levels $i$ of the cascade.

The double-phosphorylation cascade and its simplified form with merged phosphorylation steps are shown in Figures 3 and 4, respectively. The latter scheme gives rise to one EFM. The corresponding overall reaction reads as follows:

$$E_i \text{PP} + E_4 + 6 \text{NTP} = E_i + E_4\text{PP} + 6 \text{NDP} + 6P \quad (19)$$

The complete Metatool input is given in Table 3.

Sometimes a single-phosphorylated kinase $E_{i+1}\text{P}$ shows a different substrate specificity than the double-phosphorylated $E_{i+1}\text{PP}$. In such cases, the $E_{i+1}\text{P}$ can catalyze the phosphorylation of a kinase $F_1$ while $E_{i+1}\text{PP}$ phosphorylates $G_1$, and thus a diverging double-phosphorylation cascade with a branching point occurs. But again we do not know any kinetics and thus can model neither the amplification of the signal along the cascade nor the exact interplay of all phosphorylation and dephosphorylation steps of $E_{i+1}$. Nevertheless, the following two minimal preconditions must be valid: To transduce the signal along the two possible routes of the diverging cascade, it is necessary that $F_1$ is double-phosphorylated by $E_{i+1}\text{P}$ and $G_1$ by $E_{i+1}\text{PP}$. So we reduce the branching point to the following four reactions:

$$E_{i\text{PP}} + n E_{i+1} + n \text{NTP} = E_i + 2P + n E_{i+1\text{P}} + n \text{NDP} \quad (20a)$$
$$E_{i\text{PP}} + n E_{i+1} + 2n \text{NTP} = E_i + 2P + n E_{i+1\text{PP}} + 2n \text{NDP} \quad (20b)$$
$$E_{i+1}\text{P} + n F_1 + 2n \text{NTP} = E_{i+1} + P + n F_1\text{PP} + 2n \text{NDP} \quad (20c)$$
$$E_{i+1}\text{PP} + n G_1 + 2n \text{NTP} = E_{i+1} + 2P + n G_1\text{PP} + 2n \text{NDP} \quad (20d)$$
Again, we have combined the phosphorylation of $E_{i+1}$ and the subsequent dephosphorylation of $E_i$. In the minimum case $n = 1$, we get:

$$E_iPP + E_{i+1} + NTP = E_i + 2P + E_{i+1}P + NDP \quad (21a)$$

$$E_iPP + E_{i+1} + 2NTP = E_i + 2P + E_{i+1}PP + 2NDP \quad (21b)$$
\[ E_{i+1}P + F_i + 2 \text{NTP} = E_{i+1} + P + F_iPP + 2 \text{NDP} \quad (21c) \]
\[ E_{i+1}PP + G_i + 2 \text{NTP} = E_{i+1} + 2P + G_iPP + 2 \text{NDP} \quad (21d) \]

For the other reaction equations of the cascade, again equations analogous to eq. (18) can be used.

The diverging double-phosphorylation cascade and its simplified form with merged phosphorylation steps are shown in Figures 5 and 6, respectively. The latter scheme gives rise to two EFMIs, consisting of the reactions R1, R2, R4, R6, and R7, and of R1, R3, R5, R8, and R9. The corresponding overall reactions are:

### Table 3. Metatool Input for the Simplified Double-Phosphorylation Cascade Shown in Figure 4

| -ENZREV | -ENZREV | R1 R2 R3 |
| -METINT | E2 E2PP E3 E3PP |
| -METEXT | NDP NTP E1 E1PP E4 E4PP P |
| -CAT | R1 : 2 NTP + 1 E2 + 1 E1PP = 1 E2PP + 2 NDP + 1 E1 + 2 P |
| | R2 : 2 NTP + 1 E3 + 1 E2PP = 1 E3PP + 2 NDP + 1 E2 + 2 P |
| | R3 : 2 NTP + 1 E4 + 1 E3PP = 1 E4PP + 2 NDP + 1 E3 + 2 P |

**FIG. 5.** Diverging double-phosphorylation cascade, comprising the kinases E₁–E₃, F₁–F₃, and G₁–G₃. Rectangles, external metabolites; ellipses, internal metabolites.
E₁PP + F₃ + 9 NTP = E₁ + F₃PP + 9 NDP + 9 P \hspace{1cm} (22a)
E₁PP + G₃ + 10 NTP = E₁ + G₃PP + 10 NDP + 10 P \hspace{1cm} (22b)

The complete Metatool input is given in Table 4.

4. EXAMPLE FROM INSULIN SIGNALING

As a real example for a diverging cascade, we modeled part of the insulin signaling pathway in humans based on data from the Transpath database (Krull et al., 2006). The model (Metatool input given in Table 5) starts with a binding reaction between insulin and its receptor:

$$\text{Insulin} + \text{InsR} = \text{IIR}$$

where IIR denotes the insulin-insulin-receptor complex. This complex phosphorylates itself and becomes active. After a binding reaction with Shc, a subsequent phosphorylation and a further complexation with Grb-2 and Sos, the resulting enzyme complex is able to catalyze the GDP-GTP-exchange reaction that activates Ras. Formally, the exchange of GDP and GTP can be modeled in the same way as a

Table 4. Metatool Input for the Simplified Diverging Double-Phosphorylation Cascade Shown in Figure 6

-ENZREV
-ENZIRREV
R1 R2 R3 R4 R5 R6 R7 R8 R9
-METINT
E2 E2PP E3 E3P E3PP F1 F1PP F2 F2PP G1 G1PP G2 G2PP
-METEXT
NDP NTP E1 E1PP F3 F3PP G3 G3PP P
-CAT
R1 : 2 NTP + 1 E2 + 1 E1PP = 1 E2PP + 2 NDP + 1 E1 + 2 P
R2 : 1 NTP + 1 E3 + 1 E2PP = 1 E3P + 1 NDP + 1 E2 + 2 P
R3 : 2 NTP + 1 E3 + 1 E2PP = 1 E3PP + 2 NDP + 1 E2 + 2 P
R4 : 2 NTP + 1 F1 + 1 E3P = 1 F1PP + 2 NDP + 1 E3 + 1 P
R5 : 2 NTP + 1 G1 + 1 E3PP = 1 G1PP + 2 NDP + 1 E3 + 2 P
R6 : 2 NTP + 1 F2 + 1 F1PP = 1 F2PP + 2 NDP + 1 F1 + 2 P
R7 : 2 NTP + 1 F3 + 1 F2PP = 1 F3PP + 2 NDP + 1 F2 + 2 P
R8 : 2 NTP + 1 G2 + 1 G1PP = 1 G2PP + 2 NDP + 1 G1 + 2 P
R9 : 2 NTP + 1 G3 + 1 G2PP = 1 G3PP + 2 NDP + 1 G2 + 2 P

Table 5. Metatool Input for the Modeled Part of the Insulin Pathway Shown in Figure 7

-ENZREV
R01 R03 R05 R06 R09
-ENZIRREV
R02 R04 R07 R08 R10 R11 R12 R13 R14 R15
-METINT
ERK ERK(P) Grb–2_Sos IIR IIR(PY) IIR(PY)_Shc IIR(PY)_Shc(PY) IIR(PY)_Shc(PY)_Grb–2_Sos
InsR Insulin MEK MEK(P) RSK RSK(P) Raf Raf(P) Ras_GDP Ras_GTP Ras_GTP_Raf(P) Shc Sos
-METEXT
ADP ATP CREB CREB(PS) GDP GTP NDP NTP P PFKFB–2 PFKFB–2(PS) S6 S6(P)
-CAT
R01 : 1 Insulin + 1 InsR = 1 IIR
R02 : 1 IIR + 1 ATP = 1 IIR(PY) + 1 ADP
R03 : 1 IIR(PY) + 1 Shc = 1 IIR(PY)_Shc
R04 : 1 IIR(PY)_Shc + 1 NTP = 1 IIR(PY)_Shc(PY) + 1 NDP
R05 : 1 Grb–2 + 1 Sos = Grb–2_Sos
R06 : 1 IIR(PY)_Shc(PY) + 1 Grb–2_Sos = IIR(PY)_Shc(PY)_Grb–2_Sos
R07 : 1 IIR(PY)_Shc(PY)_Grb–2_Sos + 1 Ras_GDP + 1 GTP = 1 Ras_GTP + 1 GDP + 1 Insulin + 1 InsR + 1 Shc + 1 Grb–2 + 1 Sos + 2 P
R08 : 1 Raf + 1 ATP = 1 Raf(P) + 1 ADP
R09 : 1 Ras_GTP + 1 Raf(P) = Ras_GTP_Raf(P)
R10 : 1 Ras_GTP_Raf(P) + 1 MEK + 1 NTP = 1 MEK(P) + 1 NDP + 1 Ras_GDP + 1 Raf + 2 P
R11 : 1 ERK + 1 NTP + 1 MEK(P) = 1 ERK(P) + 1 NDP + 1 MEK + 1 P
R12 : 1 RSK + 1 NTP + 1 ERK(P) = 1 RSK(P) + 1 NDP + 1 ERK + 1 P
R13 : 1 PFKFB–2 + 1 ATP + 1 RSK(P) = 1 PFKFB–2(PS) + 1 ADP + 1 RSK + 1 P
R14 : 1 CREB + 1 ATP + 1 RSK(P) = 1 CREB(PS) + 1 ADP + 1 RSK + 1 P
R15 : 1 S6 + 1 ATP + 1 RSK(P) = 1 S6(P) + 1 ADP + 1 RSK + 1 P

P, phosphorylation in general; PS, phosphorylation at a serine; PY, phosphorylation at a tyrosine.
The kinase cascade continues via MEK, ERK, and RSK. The branching point is given by the kinase RSK, which catalyzes (in its phosphorylated form) the phosphorylation (and hence activation) of the enzyme PFKFB-2, the transcription factor CREB and the ribosomal protein S6. PFKFB-2 is a bifunctional enzyme, known as 6-phosphofructo-2-kinase (EC 2.7.1.105) and fructose-2,6-bisphosphate 2-phosphatase (EC 3.1.3.46), both being important in hexose metabolism. CREB and S6 affect glycolysis and protein synthesis, respectively. The corresponding reactions are labeled R_{13}–R_{15} in Table 5. The modeled cascade is shown in Figure 7.

This system gives rise to three EFMs. They have reactions R_{01}–R_{12} in common and differ only in the last reaction (R_{13}, R_{14}, R_{15}). The corresponding overall reactions read as follows:

\[
\begin{align*}
3 \text{ ATP} + \text{GTP} + 4 \text{ NTP} + \text{PFKFB} - 2 &= 3 \text{ ADP} + \text{GDP} + 4 \text{ NDP} + 7 \text{ P} + \text{PFKFB} - 2(\text{PS}) \\
3 \text{ ATP} + \text{CREB} + \text{GTP} + 4 \text{ NTP} &= 3 \text{ ADP} + \text{CREB(PS)} + \text{GDP} + 4 \text{ NDP} + 7 \text{ P} \\
3 \text{ ATP} + \text{GTP} + 4 \text{ NTP} + \text{S6} &= 3 \text{ ADP} + \text{GDP} + 4 \text{ NDP} + 7 \text{ P} + \text{S6(P)}
\end{align*}
\]
These three EFMs can be interpreted in biological terms as follows. Insulin activates fructose/mannose metabolism, glycolysis as well as protein synthesis. Although these three effects usually occur simultaneously, each EFM could operate alone when the other two are non-functional, for example, when some component proteins are knocked out. For this system, the elementary signaling routes are quite obvious. We have taken this example as a proof of concept. In larger networks, the signaling routes can no longer be detected by inspection.

5. DISCUSSION

Here, we have presented an approach for detecting elementary signaling routes in enzyme cascades (e.g., phosphorylation cascades). To this end, we adapted elementary flux modes (EFM) analysis, which had been established earlier for detecting pathways in metabolism (Schuster et al., 1999, 2000). A schematic application of the concept of EFMs to enzyme cascades would lead to the trivial result that the EFMs reflect particular enzyme cycles (e.g., phosphorylation-dephosphorylation cycles) rather than the routes of information transfer. In order to describe the latter in a suitable way, we have started from the reasoning that signaling usually implies signal amplification or at least a constant signal strength, but never a diminution. In the limit case of constant signal strength, each active enzyme molecule should activate exactly one enzyme molecule at the next level of the cascade. By this coupling, we can write reaction equations that lead to EFMs representing routes of information transfer. The same routes are relevant when signal amplification occurs. By this reasoning, the application of elementary flux modes analysis and related methods such as extreme pathway analysis (Schilling et al., 2000) and minimal T-invariants (Starke, 1990) to intracellular signaling systems has been put on a firm theoretical basis. Thus, EFMs can be calculated, for example, by the program Metatool (von Kamp and Schuster, 2006) also for enzyme cascades. Earlier, these analyses have been applied in a formal way to such systems without a theoretical justification (Xiong et al., 2004; Heiner et al., 2004; Sackmann et al., 2006).

As a proof of concept, we have applied the presented method to part of the insulin signaling network. The three resulting EFMs can be interpreted in biochemical terms. The first EFM leads to a serine phosphorylated and thus active 6-phosphofructo-2-kinase/fructose 2,6-bisphosphatase (EC 3.1.3.46/EC 2.7.1.105). This bifunctional enzyme is part of the fructose and mannose metabolism, where it catalyzes the reaction from $\beta$-D-fructose 6-phosphate to $\beta$-D-fructose 2,6-bisphosphate, and back. The output of the second EFM is the activated transcription factor CREB(PS), which regulates (together with the transcription factor forkhead box O1A) the expression of the insulin-like growth factor binding protein 1. That protein is necessary for binding the insulin-like growth factor 1, which in turn regulates cell growth, glucose metabolism, and several other processes. The third EFM results in the phosphorylated ribosomal protein S6. Hence, the further outcome of this signaling route is activation of protein synthesis. For all of these functions, see the Transpath database (Krull et al., 2006).

Of course, the entire insulin signaling network has many more functions, such as activation of glyco- genesis and enhancing glucose uptake in muscle cells and adipocytes by increasing the number of GLUT4 transporters in their membranes.

The enzyme cascades analyzed here can be considered as relay races with changing batons because it might appear as if the phosphates were transferred along the cascades, but they are replaced at each level. This situation differs from group transfer pathways such as the phophotransferase system (PTS) for which the analogy to a relay race is more appropriate because the “baton” is not replaced. Papin and Palsson (2004) have studied the latter type of systems.

The enzyme Ras, which we have included in our model, belongs to the G-proteins and in particular to the family of small GTPases. Such enzymes are activated by replacing the GDP they are complexed with in their inactive form, by GTP. The substitution is catalyzed by so-called guanine nucleotide exchange factors (GEFs) (Krauss, 2003). In our case, this is IIR(PY)_Shc(PY)_Grb-2_Sos, the complex of insulin, insulin receptor, Src homologous and collagen protein, growth-factor receptor-binding protein 2, and the enzyme Son of Sevenless. The inactivation of these activated G-proteins is triggered by their intrinsic ability to hydrolyze GTP to GDP and phosphate (Gomperts et al., 2002). From the point of view of the biochemical structure of the active form of the G-protein, the replacement of GDP by GTP is equivalent to a phosphorylation. Therefore, our formalism is also applicable to this class of signaling proteins.

Our analysis differs from the approach proposed by Alon (2007) in that we determine routes going through the entire signaling network, starting from an initial signal and leading to some cellular response. In contrast,
Alon (2007) extracts modular (non-overlapping) network motifs, which are smaller than the signaling routes determined here, which can have subroutes in common.

It is promising to apply robustness analysis (Wilhelm et al., 2004; Behre et al., 2008) and the concept of minimal cut sets (Klamt and Gilles, 2004; Klamt, 2006) proposed earlier for metabolic networks to enzyme cascades. Since these methods use EFMs, their application is straightforward. Also the concept of enzyme subsets (Pfeiffer et al., 1999) and the refined concepts in flux coupling analysis (Burgard et al., 2004) can be applied. On Boolean networks, information transfer has been analyzed, and minimal intervention sets have been defined without explicitly considering mass balance constraints (Klamt et al., 2006). It would be interesting to investigate the interrelations between these approaches.

6. APPENDIX

Components involved in the modeled part of the human insulin pathway are described in the following table. P denotes phosphorylation in general; PS, phosphorylation at a serine; and PY, phosphorylation at a tyrosine.

<table>
<thead>
<tr>
<th>Abbreviated component name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADP</td>
<td>Adenosine 5'-diphosphate</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine 5'-triphosphate</td>
</tr>
<tr>
<td>CREB</td>
<td>cAMP-responsive element-binding protein</td>
</tr>
<tr>
<td>CREB(PS)</td>
<td>cAMP-responsive element-binding protein (serine phosphorylated)</td>
</tr>
<tr>
<td>ERK</td>
<td>Extracellular signal-regulated kinase 1</td>
</tr>
<tr>
<td>ERK(P)</td>
<td>Extracellular signal-regulated kinase 1 (phosphorylated)</td>
</tr>
<tr>
<td>GDP</td>
<td>Guanosine 5'-diphosphate</td>
</tr>
<tr>
<td>Grb-2</td>
<td>Growth-factor receptor-binding protein 2</td>
</tr>
<tr>
<td>Grb-2_Sos</td>
<td>Complex of Grb-2 and Sos</td>
</tr>
<tr>
<td>GTP</td>
<td>Guanosine 5'-triphosphate</td>
</tr>
<tr>
<td>IIR</td>
<td>Insulin insulin receptor complex</td>
</tr>
<tr>
<td>IIR(PY)</td>
<td>Insulin insulin receptor complex (tyrosine phosphorylated)</td>
</tr>
<tr>
<td>IIR(PY)_Shc</td>
<td>Complex of IIR(PY) and Shc</td>
</tr>
<tr>
<td>IIR(PY)_Shc(PY)</td>
<td>Complex of IIR(PY) and Shc (tyrosine phosphorylated)</td>
</tr>
<tr>
<td>IIR(PY)_Shc(PY)_Grb-2_Sos</td>
<td>Complex of IIR(PY)_Shc(PY) and Grb-2_Sos</td>
</tr>
<tr>
<td>InsR</td>
<td>Insulin receptor</td>
</tr>
<tr>
<td>Insulin</td>
<td>Insulin</td>
</tr>
<tr>
<td>MEK</td>
<td>ERK-kinase (mitogen-activated protein kinase)</td>
</tr>
<tr>
<td>MEK(P)</td>
<td>ERK-kinase (phosphorylated)</td>
</tr>
<tr>
<td>NDP</td>
<td>Nucleoside diphosphate</td>
</tr>
<tr>
<td>NTP</td>
<td>Nucleoside triphosphate</td>
</tr>
<tr>
<td>P</td>
<td>Orthophosphate</td>
</tr>
<tr>
<td>PFKFB-2</td>
<td>6-Phosphofructo-2-kinase/fructose 2,6-bisphosphatase (EC 3.1.3.46 / EC 2.7.1.105)</td>
</tr>
<tr>
<td>PFKFB-2(PS)</td>
<td>PFKFB-2 (serine phosphorylated)</td>
</tr>
<tr>
<td>Raf</td>
<td>Raf-1, serine/threonine-specific kinase (EC 2.7.11.1)</td>
</tr>
<tr>
<td>Raf(P)</td>
<td>Raf (phosphorylated)</td>
</tr>
<tr>
<td>Ras_GDP</td>
<td>GTPase from the Ras family (inactive form)</td>
</tr>
<tr>
<td>Ras_GTP</td>
<td>GTPase from the Ras family (active form)</td>
</tr>
<tr>
<td>Ras_GTP_Raf(P)</td>
<td>Complex of Ras_GTP and Raf(P)</td>
</tr>
<tr>
<td>RSK</td>
<td>Ribosomal S6 kinase</td>
</tr>
<tr>
<td>RSK(P)</td>
<td>Ribosomal S6 kinase (phosphorylated)</td>
</tr>
<tr>
<td>S6</td>
<td>Ribosomal protein S6</td>
</tr>
<tr>
<td>S6(P)</td>
<td>Ribosomal protein S6 (phosphorylated)</td>
</tr>
<tr>
<td>Shc</td>
<td>Src homologous and collagen protein (adaptor protein)</td>
</tr>
<tr>
<td>Sos</td>
<td>Son of Sevenless (GTPase-controlling signal molecule)</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

We thank Steffen Klamt, Ina Koch, and Regina Samaga for stimulating discussions. Financial support by the BMBF (German Ministry for Education and Research) to J. Behre (HepatoSys Program and Jena Center for Bioinformatics) is gratefully acknowledged.

DISCLOSURE STATEMENT

No competing financial interests exist.

REFERENCES


Address reprint requests to:
Jörn Behre
Faculty of Biology and Pharmaceutics
Section of Bioinformatics
Friedrich Schiller University Jena
Ernst-Abbe-Platz 2
D-07743 Jena, Germany

E-mail: joern.behre@uni-jena.de