Pergamon

Comput. Biol. Med. Vol. 25, No. 3, pp. 321–334, 1995

0010–4825/95 $9.50 + 0.00

0010-4825(95)00019-4

INTERACTIVE MODELLING AND SIMULATION OF BIOCHEMICAL NETWORKS

RALF HOFESTÄDT* and FRANK MEINEKE

Universität Koblenz, FB Informatik, Rheinau 1, 56075 Koblenz, Germany; and University of Leipzig, Department of Medical Computer Science and Statistics, Liebigstr. 27, 04103 Leipzig, Germany

(*Received* 8 *September* 1994; *revised* 26 *January* 1995)

Abstract—The analysis of biochemical processes can be supported using methods of modelling and simulation. New methods of computer science are discussed in this field of research. This paper presents a new method which allows the modelling and analysis of complex metabolic networks. Moreover, our simulation shell is based on this formalization and represents the first tool for the interactive simulation of metabolic processes.

Simulation shell modelling Metabolic pathways Metabolic networks
Metabolic bottleneck

1. INTRODUCTION

Methods of molecular biology make possible the isolation, sequence analysis and synthesis of genes and enzymes [1]. This method produces an exponential growth of biological data. With regard to gene sequences (enzymes) the following database systems are popular—GENBANK, EMBL and JJDJ (SWISSPROT, PIR)—and can be ordered by CD ROM or used via electronic mail [2]. Moreover, biotechnological methods facilitate the analysis of biochemical processes. Enzymes are biosynthetic products of specific genes, which catalyse biochemical processes. Metabolic pathways are cascades of biochemical reactions, which can interact and create complex metabolic networks [3]. Analysis and synthesis of metabolic networks are the main aims in the field of metabolic engineering [4]. In the case of metabolic pathways Boehringer analyzed and prepared all data [5]. The so-called Boehringer wall picture, which represents these data by a graphical notation, can be requested from Boehringer Mannheim. Based on this dataset different database systems are implemented. A fundamental component for the realization of metabolic engineering is the implementation of integrated information systems, which represent genes, enzymes and metabolic pathways. In this domain of research, Karp has developed the first metabolic information system for *E. coli* [6]. Moreover, modelling of metabolic processes in combination with such information systems will be the basic tool for metabolic engineering. Therefore, dynamic models which allow the implementation of useful interactive simulation programmes are important. With regard to modelling metabolic processes, different models are discussed. Abstract models are based on binary automata or logical approaches which facilitate the qualitative discussion on an abstract level [7, 8]. Analytical models are based on the use of differential equations which make possible the exact simulation of concentration rates. However, the simulation of kinetic effects is possible. The disadvantage of this method is the effort of the computational complexity and the fact that concentration rates are not available nowadays. Discrete models facilitate the qualitative modelling of metabolic networks and are based on the theory of formal languages, automata, graph theoretical approaches and methods of artificial intelligence.

* Present address: Universität Leipzig, Inst. für Mediz. Informatik, Statistik und Epidemiologie, Liebigstr. 27, 04103 Leipzig, Germany.

The aim of our work is to develop a new concept of modelling and simulation of metabolic processes. Therefore, we defined a grammatical formalization which allows the simultaneous modelling of genetic processes, biosynthetic processes and cell communication processes. This model is an extension of the semi-Thue system [9] and represents special rules for specific approaches. Our formalization, genetic grammar [10], allows the definition of different languages, which discuss different biological aspects [11]. Moreover, we implemented a simulation shell based on genetic grammar. While using this simulation shell, different disadvantages became visible. Based on this project we developed a probabilistic rule based system defining a universal rule, which is able to describe different metabolic processes in a biochemically orientated language. Besides, we developed a user friendly shell which allows the use of this simulator without any previous knowledge of programming and computing.

In this paper we present a new concept which represents the first method of interactive modelling and simulation of complex metabolic processes. The approach is important in the research field of metabolic engineering, because analysis of metabolic pathways becomes more and more important [4] in biomedicine and biotechnology. The reason is that genetic defects cause diseases and ought to be identified. An important application is the detection of metabolic bottlenecks [12, 13] because such configurations signal specific concentration rates, based on genetic defects.

2. BIOCHEMICAL BASIC KNOWLEDGE

The metabolism is based on biochemical reactions. Modelling and simulation are important in understanding the behaviour of biochemical reactions. Therefore, the empirical data must be interpreted. In the case of genes, enzymes and biochemical reactions multiple data are available. Several models have been developed. The main gap in the field of modelling and simulation is the development of an integrative and interactive simulation shell [14]. Integrative means that this model enables us to discuss biosynthetic processes, gene regulation processes and cell communication processes. Therefore, integrative models allow the discussion of complex metabolic networks.

The genetic information (DNA) controls metabolism indirectly [15]. The protein synthesis of structural genes produces specific enzymes which catalyse biochemical reactions. The transcription of these genes has to be regulated by enzymatic mechanisms. The fundamental model of gene regulation is based on the model of Jacob and Monod [16]. Regarding this model the operon, which consists of the promotor, structural gene(s), and the terminator sequence, is the primary unit of the gene regulation complex. The RNA polymerase makes contact with the promotor sequence and starts the transcription process. The affinity of the promoter/RNA polymerase complex is defined by specific sequences (Pribnow box). Homeotic genes, transposons, enhancer and silencer genes demonstrate that gene regulation is a complex process regarding eucariotic cells [1]. The biochemical control of a cell is defeated by biochemical reactions, which change substrates into products ($S \rightarrow P$). This can be done spontaneously or can be catalysed by specific enzymes. Mostly biochemical reactions are two-way processes catalysed by specific enzymes. Therefore, concentration rates are important ($S \leftrightharpoons P$). However, the flux of biosynthetic processes is controlled by the enzyme affinity, the enzyme concentration and the reaction rate. Specific regulatory proteins and enzymes—influence proteins—are able to modify these parameters.

In the case of two-way biochemical reactions, the enzyme will catalyse biochemical reactions from the higher to the lower level of concentration rates. Moreover, kinetic effects are important [17]. Most biochemical reactions follow the Michaelis–Menten kinetic scheme characterized by the following equation:

$$V = -\mathrm{d}S/\mathrm{d}t = V^*_{\max} S/S + K_\mathrm{m},$$

where V is the reaction rate, S is the substrate concentration at the given rate of reaction, $V_{\max}$ is the maximum reaction rate regarding substrate saturation and K_m is the Michaelis constant. V and K_m are two constants that characterize the interactions of the enzyme

with its substrate. Enzymes can be controlled by modifying the affinity, efficiency and specification of the enzyme. However, genes and their regulation mechanisms, biosynthesis and their catalysis, and cell communication processes are called elementary metabolic processes, which define the behaviour of metabolism. All these processes build metabolic networks, which are connected with elementary metabolic processes, which influence each other in a well defined way.

Moreover, to develop a useful integrative model the main features of metabolic processing must be considered and implemented. Therefore, we will discuss the characteristics of the metabolic processes. Operons, known as genetic instructions, are activated depending on the specific promotor affinity which can be influenced by specific DNA units as silencers or enhancers. Therefore, the genetic instruction must be interpreted as a probabilistic instruction. Operons can be activated simultaneously, that means metabolic processing is based on the concept of parallel processing. A main feature of the genetic information is the modular organisation of the genome which is demonstrated by the function of homeotic genes [15]. Mutator genes, transposons and virus genomes demonstrate the dynamic of the genetic instruction sets. Moreover, the metabolic processing is based on the concept of dataflow processing. Based on these features we will develop a rule based system, which models and simulates integrative metabolic processes.

3. THEORETICAL BASIC KNOWLEDGE

In this paper the standard mathematical symbols are used. $\cup$ denotes the union of sets, $\subseteq$ denotes the sub set, and * denotes the Kleene star operator. [0, 1] denotes all rational numbers between 0 and 1. An alphabet is a finite set of symbols. A (formal) language is a set of strings of symbols from any one alphabet. A semi-Thue system is a Chomsky type 0 grammar [9]. This is a 4-tuple $G=(V, T, P, S)$. V is a finite set of variables (also called non-terminals or categories), each of which represents a language. The languages represented by the variables are described recursively in terms of each other and primitive symbols are called terminals. T represents the finite set of terminals. We assume that V and T are disjoint. P is a finite set of productions; each production is of the form $a \rightarrow b$, where $a \in (V \cup T)^*$. Finally, S is a special variable called the start symbol.

The language generated by a grammar can be defined by two relations; ($\Rightarrow$) and ($\overset{*}{\Rightarrow}$) between strings in $(V \cup T)^*$. If $a \rightarrow b$ is production of P and a and b are any strings in $(V \cup T)^*$, then $xay \Rightarrow xby$. We say that the production $a \rightarrow b$ is applied to the string xay to obtain xby or that xay directly derives xby. Suppose that $a_1, \ldots, a_n$ are strings in $(V \cup T)^*$, $n \geq 1$, and $a_1 \Rightarrow a_2 \Rightarrow \ldots \Rightarrow a_n$. Then we say $a_1 \overset{*}{\Rightarrow} a_n$ or a_1 derives a_n in grammar G.

A multi-set m upon a set X is a function m: $X \rightarrow \mathbb{N}$. Elements and their quantity are combined in a set of 2-tuples. Example: Let $X=\{a, b, c\}$. A possible multi-set upon X would be: $\{(a, 2), (b, 0), (c, 1)\}$. For two multi-sets m and k upon X the usual set-operations are defined:

$$x \in m := m(x) > 0$$

$$k \subseteq m := \forall a \in X: k(a) \leq m(a)$$

$$k + m := k(d) + m(d)$$

$$k - m := k(d) - m(d).$$

4. METHOD

There are many models and simulation shells available in the research field of modelling and simulation of metabolic processes. Most of them are based on differential equations which try to simulate biosynthetic processes [18, 19]. However, no model or simulation shell exists which is able to discuss integrative metabolic processes. Our idea is to transform the analyzed metabolic knowledge into an integrative model. Metabolism is based on metabolites, metabolic structures and biochemical reactions. Metabolites and

metabolic structures can be interpreted as a specific alphabet and the biochemical reactions are rules for this alphabet. Therefore, we choose the theory of formal languages to develop a suited model. However, Chomsky grammar must be expanded into the analyzed features of metabolic processes. Therefore, a probabilistic, parallel rule-based system will be presented. The implementation of this formalization presents the first interactive simulation shell for the modelling of metabolic processes [20].

4.1. *Probabilistic parallel rule-based system*

Our model is an extension of the genetic grammar [10]. Moreover, using a universal rule, this formalization allows the representation of genetic, biosynthetic and cell communication processes. Furthermore, it is necessary to enlarge this discrete model by adding concentration rates for each metabolite. Metabolites are substances or substance concentrations, which can be modified by biochemical reactions. Enzymes are specific proteins which catalyse biochemical reactions. Inducers and repressors are metabolites, which are able to speed up or slow down (prevent) biochemical reactions. The biochemical space (cell state) of a cell is a mixture of these components. The set of all cell states will be denoted by Z. By these definitions the abstract metabolism is given by the actual cell state and the biochemical reaction rules. The metabolic rule is the basic unit of the metabolic system. This is a universal rule, able to describe all biochemical reactions. Moreover, this rule set defines the set Z indirectly.

Definition 4.1: *metabolic rule.* Let Z be a finite set of cell states. A 5-tuple (B, A, E, I, p) with $p \in [0, 1]$ and $B, A, I, R \subseteq Z$ is called a metabolic rule, p is called rule probability, B (before) a set of preconditions, A (after) a set of postconditions, E (enzyme) a set of catalysed conditions and I (inhibitor) a set of inhibitor conditions.

Based on the metabolic rule we are able to define the basic model.

Definition 4.2: *metabolic system.* $G = (Z, R)$ is called a metabolic system, Z is a finite set of cell states, $S \in Z$ is called the start state and R is a set of metabolic rules, called a metabolic rule set.

In the following we define the meaning of the metabolic system. The integration of the analyzed metabolic features is the basic idea of this formalization. This is the reason we develop a stochastic parallel derivation mechanism, which will describe the change of actual cell states depending on the specified rule set. Therefore, the set of all activated rules must be fixed. This will be the first step of the derivation process. A rule is activated if all preconditions of this rule are also elements of the actual state $z \in Z$. Moreover, effects of inducer and inhibitor elements must be considered. If such influential metabolites are elements of the actual state z, then the probability of this rule will be modified by inhibitor and inducer effects (the rule probability will be modified by these elements). The function CALCULATE(z, r) will determine the absolute probability value of rule r depending on state z. A random generator (RANDOM) using the absolute probability value of the input works as a Boolean function and will produce either positive or negative results (true or false). Regarding the Boolean value true (false), $r \in R$ is activated (deactivated) and goes into action.

Definition 4.3: *activation.* Let $G = (Z, R)$ be a metabolic system, $r = (B, A, E, I, p) \in R$ a rule and $z \in Z$ a cellular state, r is activated by z (in symbols r_z), iff $\forall x \in z.\ \forall x \in B\ x \in Z$ $A(z) = \{r \in R : r \text{ is activated by } z\}$ is called the set of activated rules by z.

Any activated rule $r \in R$ can go into action. The action of r will modify the actual cellular state of the metabolic system. Elements of the actual cellular state, which are elements of the before set of rule r will be eliminated in z and all elements of the after component will be added to z. Therefore, the action of rule r can produce a new state $z' \in Z$.

Definition 4.4: *action.* Let $G = (Z, R)$ be a metabolic system, $z \in Z$ the actual cellular state and $r_z = (B, A, E, I, p) \in R$. The action of r_z is defined by:

If RANDOM(CALCULATE(z, r)) = true then $z' = (z\text{-}B) \cup A$.
The action of r_z will be described in symbols by $z \rightarrow z'$.

The one step derivation of a metabolic system is defined by the (quasi) simultaneous action of all activated rules. Therefore, we consider the set of all activated rules and determine two new sets: the before set and the after set. The before set includes all B elements of the activated rules. The definition of the after set is analagous. Using these sets the one-step derivation could be interpreted as an addition and subtraction of elements.

Definition 4.5: *one-step derivation*. Let $G = (Z, R)$ be the metabolic system, $z \in Z$ the actual cellular state, $A(z)$ the set of all activated rules under z and $B_z = \{B: \exists r \in A(z)\ B \in r_z\}$ and $A_z = \{A: \exists r \in A(z)\ A \in r_z\}$. The simultaneous action of $A(z)$ is called one-step derivation, iff $z' = (z - B_z) \cup A_z$. The one-step derivation is described in symbols by $z \Rightarrow z'$.

Each action could be interpreted as an independent event. Therefore, the probability of each one-step derivation can be calculated with respect to the absolute probability values of all activated and deactivated rules. In our simulation system this will be done by multiplying these values.

Definition 4.5 defines the parallel derivation procedure. However, based on the one-step derivation we can define the derivation inductively. However, based on the one-step derivation, probability can be calculated for any derivation.

Definition 4.6: *derivation*. Let $G = (Z, R)$ be a metabolic system. $x \in Z^+$ is called derivation in G, iff $|x| = 1$ or $|x| > 1$ and $\exists y' \in Z^* \exists z'$, $z'' \in Z$: $x = z'z''y$ and $z''y$ is a derivation and $z' \Rightarrow z''$.

In the case of analytical modelling it is necessary to expand our model using abstract concentration rates. To realize this requirement of each component (metabolite) specific integer values must be assigned. These values can be interpreted as abstract concentration rates. Regarding the metabolic system, these effects can be included using the formalization of multi-sets. Therefore, the definition of the metabolic system must be modified. Regarding the activation of a rule the concentration rate of any before component must be satisfied in connection with corresponding metabolites of the actual state. The concentration rate of this metabolite must be higher or equal in comparison with the corresponding before component of this metabolic rule. Moreover, the function CALCULATE must be modified. In this case the influence of all concentrations of inductor and repressor metabolites will determine the absolute rule probability. All activated rules can go into action simultaneously. With regards to corresponding metabolites of the actual state (integer values) the addition and subtraction of the concentration rates of all before and after components is needed. In this chapter we present only the fundamental part of this formalization.

Definition 4.7: *metabolic concentration*. Let $z \in Z$ be a state. The multi-set k: $z \rightarrow \mathbb{N}$ is called the metabolic concentration. K denotes the set of all metabolic concentrations.
Notation: The actual state $z = \{\text{Lactose, Glucose}\}$ represents two metabolites. The multi-set k of z, $k(z) = \{(\text{Lactose, 12}), (\text{Glucose, 22})\}$, represents 12 units of lactose and 22 units of glucose. In the following we use a specific notation for metabolic concentrations: ⟦12 Lactose, 22 Glucose⟧.

Based on the formalization of multi-sets, the analytical metabolic systems, which enables the discussion of kinetic effects, can be defined.

Definition 4.8: *analytical metabolic system*. The 3-tuple $G = (Z, R, k)$ with $A \in Z$ the start state, $k \in K$ a multi-set (K: $A \rightarrow \mathbb{N}$) and R is a finite set of metabolic rules, where $r = (B, A, E, I, p) \in R$ with $p \in [0, 1]$ and B, A, E, I are metabolic concentrations, is called the metabolic system.

The definition of activation is fundamental. Regarding multi-sets the activation of any rule $r \in R$ depends on the specified concentration rate of each rule component, the concentration rate of the actual state and the absolute rule probability.

Definition 4.9: *activation*. Let $G=(Z, R, k)$ be an analytical metabolic system, $r \in R$ a rule and $z \in Z$ a cellular state. r is activated by z (in symbols r_z), iff $\forall x \in B\ x \in z$ and $k(x) \leqslant k(z)$.
$A(z)=\{r \in R : r$ is activated by $z\}$ is called the set of activated rules by z.

Based on this definition the one-step derivation can be modified. All activated metabolic rules can go into action simultaneously. Regarding the set of activated rules, all after concentrations must be added to the actual state and all before concentrations must be subtracted from the actual state.

Definition 4.10: *one-step derivation*. Let $G=(Z, R, k)$ be an analytical metabolic system, z the actual cellular state, $A(z)$ the set of all activated rules under z and

$$B_z=\{B:\exists r \in A(z) \quad B \in r_z\} \text{ and } |B_z| \coloneqq \sum_{b \in B} k(b),$$

$$A_z=\{A:\exists r \in A(z) \quad A \in r_z\} \text{ and } |A_z| \coloneqq \sum_{a \in A} k(a).$$

The simultaneous action of $A(z)$ is called one-step derivation, iff

$$z'=\{x : x \in A_z \text{ or } \forall x \in z, y \in B_z\, x=y \text{ and } k(x)-k(y)>0\}.$$

The one-step derivation is described in symbols by $z \Rightarrow z'$.

Using the one-step derivation operator we can define the derivation of an analytical metabolic system inductively (see definition 4.6).

4.2. *Metabolica*

We developed a simulation shell (Metabolica) based on the theory of the analytical metabolic system [20]. This simulation shell can be used by applying biochemical terms which are implemented in C and run on a SUN Sparc workstation. When Metabolica is started the rule set and the start configuration is either located automatically or is manually defined. Its main parts are the rule editor and the configuration editor/browser.

The rule editor handles the construction and definition of metabolic rules and rule sets. This is done by describing the elements of the specific rule. The integer value, which can be placed in front of each specific rule component, represents the specific abstract concentration rate. Rules are identified by an obligatory comment. Sets of rules may be saved, loaded and merged.

The main window has a number of different functions depending on the selected mode, e.g. the start/set mode allows the loading or entering of the start configuration of the metabolic system. The pathways mode allows the viewing of every reached configuration of any derivated generation. The trade off between time and space vs accessibility is parameterized by the number of pathways which are calculated for every generation. To limit memory usage all intermediate configurations can be disposed of.

To begin with biochemical environment must be defined by metabolic rules. In the next stage the actual state (substrate) must be defined using the start set mode. Moreover, a specifically set goal (product) can be defined, which will be represented by a specific concentration rate of metabolites. The system can produce different derivations simultaneously and the number of pathways can be specified by the user. Moreover, the user can define a derivation number. The pathway browser of our system will produce derivations until this number or until the set goal is reached. The system includes statistical tools which show the historical path of all derivations and the concentration values in flux.

5. APPLICATIONS

The implemented universal metabolic rule allows the formalization of biosynthesis, gene expression, gene regulation and cell communication process. Regarding biosynthetic processes the before, after, inducer and repressor components are used. For example:

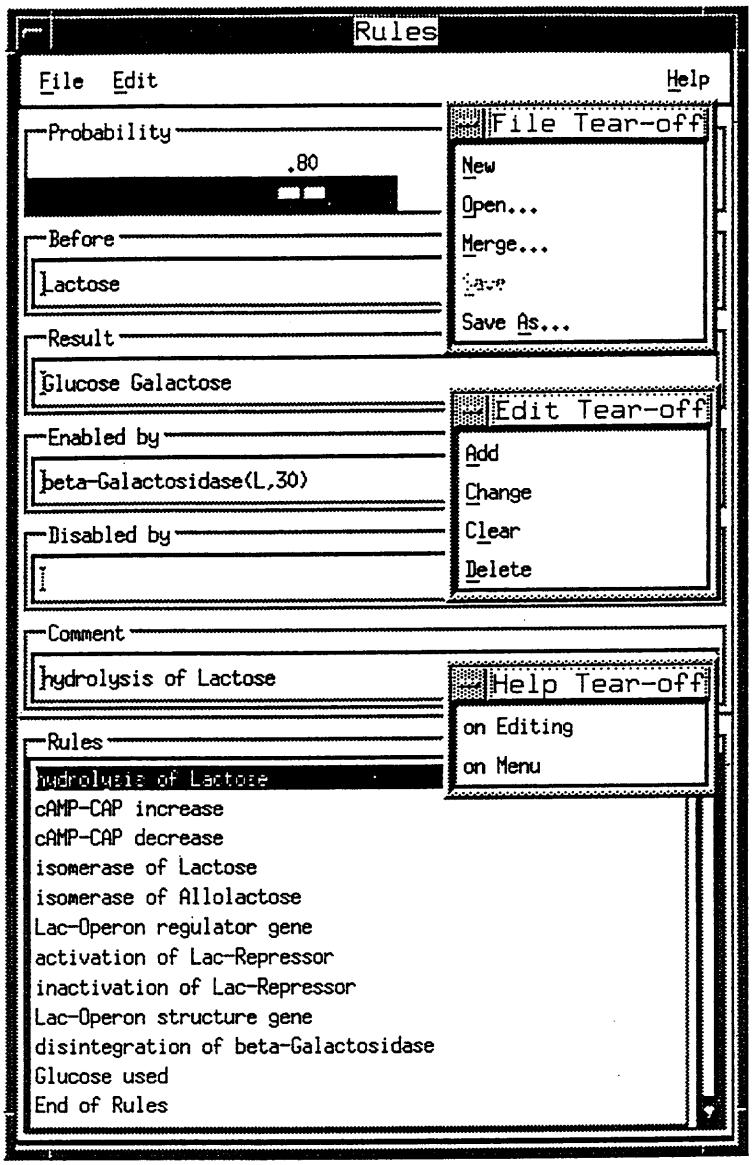

Fig. 1. Rule editor of Metabolica [20].

enzyme E_1 will catalyse the biochemical process S_1 into S_2. This can be expressed by: $B = [\![S_1]\!]$, $A = [\![S_2]\!]$ and $E = [\![E_1]\!]$. Moreover, we can add any concentration. For example: $B = [\![15\ S_1]\!]$, $A = [\![12\ S_2]\!]$, $E = [\![2\,E_1]\!]$. The probability value models the flux of this biochemical rule, which can be influenced by specific inducer and repressor metabolites depending on their concentration.

In the case of simple cell communication processes only the before or after component will be used. By doing this, we obtain the following interpretation: substance A enters the cell by endocytotic processes. Therefore, we have to define a rule where only the after component will be assigned by specific substances. Moreover, such processes can be influenced by specific events, which can be formulated regarding inducer and repressor components.

Normally, metabolites will disintegrate after a specific time interval. This can be expressed by a rule which only represents the specific before component. Moreover, concentration rates and specific influence components can be defined.

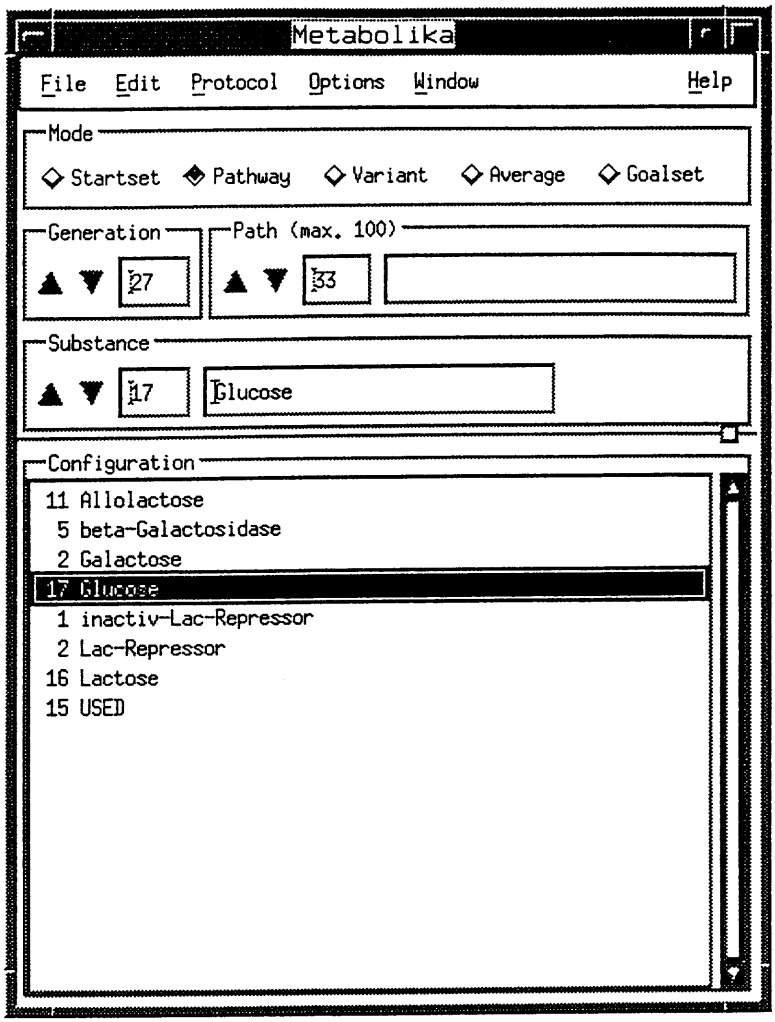

Fig. 2. Main window-browser.

In the case of gene regulation the activity of operons can be modelled easily. If we choose an operon which represents two genes (S_1, S_2), two operator genes (O_1, O_2) and one enhancer sequence, then this can be expressed by:

$$B = [\![\text{RNA-polymerase, ribosome, amino acid, tRNA}]\!]$$

$$A = [\![S_1, S_2]\!],\ E = [\![IO_1, IO_2]\!] \text{ and } I = [\![O_1, O_2]\!].$$

Therefore, it is easy to develop complex metabolic pathways and complex metabolic networks using metabolic rule sets. This will be demonstrated by a simple example, which combines different levels of metabolic processes: Isoleucin biosynthesis (*E. coli*).

The Isoleucin biosynthesis, a shown in Fig. 3, by *E. coli* serves as an example of a simple feedback control system, where the biosynthetic product, Isoleucin, disables the activity of the enzyme, which catalyzes the most important stage of the biochemical reaction [1]. It is assumed in this model, that genes produce enzymes directly and continue with a low probability, denoting the promotor affinity. The same value is used to model the probability of spontaneous enzyme decay. The complete set of rules can be given by:

$$r_1 = (\emptyset, [\![\text{Threonin-Desaminase}]\!], \emptyset, \emptyset, 0.1),$$
$$r_2 = (\emptyset, [\![P_1]\!], \emptyset, \emptyset, 0.1),$$
$$r_3 = (\emptyset, [\![\text{Reducto-Isomerase}]\!], \emptyset, 0, 0.2),$$

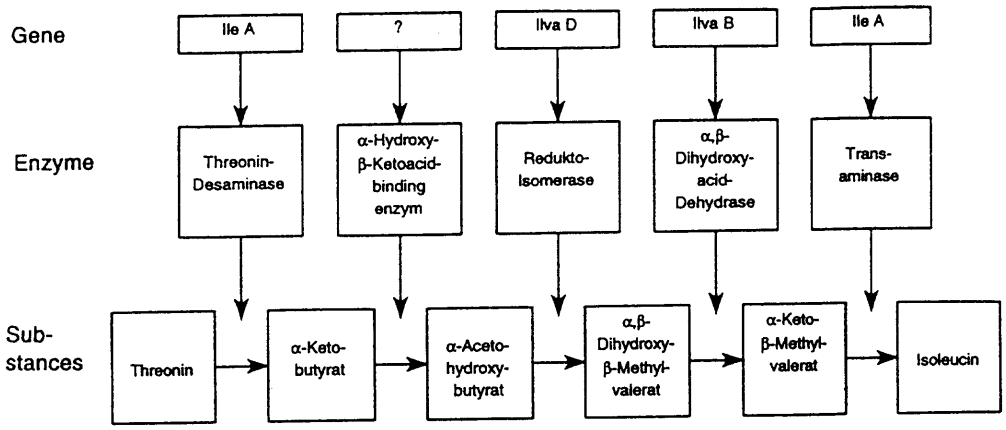

Fig. 3. Isoleucin biosyntheis (*E. coli*).

$r_4 = (\emptyset, [\![\alpha\beta\text{-Dihroxy-Dehydrase}]\!], \emptyset, \emptyset, 0.1)$,
$r_5 = (\emptyset, [\![\text{Transaminase}]\!], \emptyset, \emptyset, 0.1)$,
$r_6 = ([\![\text{Threonin-Desaminase}]\!], \emptyset, \emptyset, \emptyset, 0.4)$,
$r_7 = ([\![P_1]\!], \emptyset, \emptyset, \emptyset, 0.4)$,
$r_8 = ([\![\text{Redukto-Isomerase}]\!], \emptyset, \emptyset, \emptyset, 0.4)$,
$r_9 = ([\![\alpha\beta\text{-Dihroxy-Dehydrase}]\!], \emptyset, \emptyset, \emptyset, 0.4)$,
$r_{10} = ([\![\text{Transaminase}]\!], \emptyset, \emptyset, \emptyset, 0.4)$,
$r_{11} = ([\![\text{Threonin-Desaminase}]\!], [\![\text{Threonin}]\!], [\![\alpha\text{-Ketobutyrat}]\!], [\![\text{Isoleucin}]\!], 0.9)$,
$r_{12} = ([\![\alpha\text{-Ketobutyrat}]\!], [\![\alpha\text{-Acetohydroxy-butyrat}]\!], [\![P_1]\!], \emptyset, 0.9)$,
$r_{13} = ([\![\alpha\text{-Acetohydroxy-butyrat}]\!], [\![\alpha\beta\text{-Dihroxy-Dehydrase}]\!], [\![\text{Redukto-Isomerase}]\!], \emptyset, 0.9)$,
$r_{14} = ([\![\alpha\beta\text{-Dihroxy-Dehydrase}]\!], [\![P_2]\!], [\![\alpha\beta\text{-Dihroxy-Dehydrase}]\!], \emptyset, 0.9)$,
$r_{15} = ([\![P_2]\!], [\![\text{Isoleucin}]\!], [\![\text{Transaminase}]\!], \emptyset, 0.9)$.

The meaning of P_1 and P_2 is:

$$P_1 = \alpha\text{-Hydroxy-}\beta\text{-Keto and } P_2 = \alpha\text{-Keto-}\beta\text{-Methylvalerat.}$$

Our simulation shell allows, for example, the discussion of the delayed effect of sudden gene-defect. Starting with a start set of [[30 Threonin]] Fig. 4 shows how Isoleucin concentration increases and how Threonin developed velocity changes through the first 150 generations. By such graphical analysis, bottlenecks can be detected directly. Moreover, the produced statistical data in the case of a specific derivation can be analyzed using common software tools (mathematica).

However, this model allows the simulation of complex metabolic networks and the grammatical formalization allows the definition and implementation of different languages. These languages represent specific biological aspects [11]. For example, it is possible to produce the set of all possible pathways, to produce metabolic pathways depending on specific conditions (as for example the probability value), to search for the appearance of specific substances (as for example toxic substances), etc. These languages are specific tools. Moreover, the user can modify specific cell states or rule states in any configuration of the simulation and can also control the simulation forward and backward through the derivation space of this grammar. This is the main assumption for the interactive feature of our system. For specific metabolic questions specific analysis tools must be implemented. Therefore, the bottleneck detection is one of the actual questions and will be discussed in detail.

A metabolic bottleneck is a configuration which signals a high or low concentration rate of metabolites [3]. The detection of such configurations and metabolites is of interest because they cause diseases based on genetic defects. In the case of biomedicine, the detection of such configurations is of increasing interest [4]. However, in the case of biotechnology DNA recombination needs theoretical tools which allow the simulation of

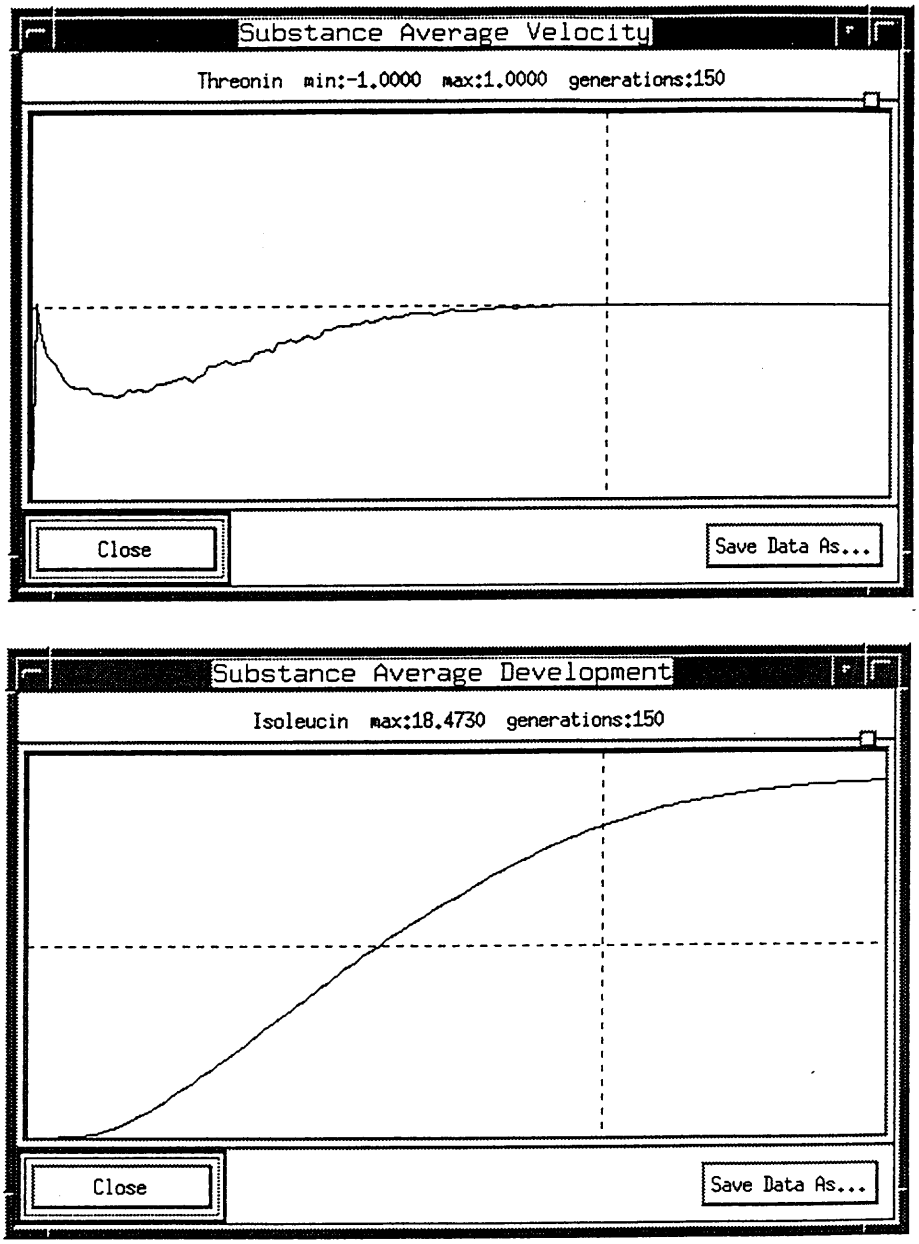

Fig. 4. Concentration development within the Isoleucin system.

well defined gene transformations. Therefore, interactive simulation tools are necessary. Definition 4.10 defines a parallel derivation operator, where any metabolite of the actual state is able to activate different rules in the same one-step derivation. If such rules go into action simultaneously high and low concentration rates can be produced by a one-step derivation.

Example 5.1:
$[\![15\,A]\!]$ is the actual state $z \in Z$ and
$R = \{(5\,A, 7\,B, \emptyset, \emptyset, 1.0), (8\,A, 4\,D, \emptyset, \emptyset, 1.0), (4\,A, 3\,C, \emptyset, \emptyset, 1.0)\}$.
All rules are activated by z and the one-step derivation can produce the state $[\![-2\,A, 7\,B, 4\,D, 3\,C]\!]$.

Important configurations are metabolic states which represent extremely high or low concentration rates. We have to explain the meaning of this derivation operator.

Definition 5.1: *metabolic bottleneck.* Let $G=(Z, R, k)$ be an analytical metabolic system, $A=a_1, \ldots, a_n$ a derivation, M a metabolite and h, l specific integer values. Each derivation is called the (h, l)-bottleneck in configuration $i \in 1 \ldots n$, iff

$$\exists a_i M \leqslant l \vee M \geqslant h \text{ with } i=1 \ldots n$$

If $M \leqslant l (M \geqslant h)$ A is called a negative (positive) (h, l)-bottleneck.

A metabolic bottleneck represents a critical concentration rate of one or more considered metabolites. In order to discuss the cause of these effects we have to distinguish between two cases.

Case 1: *positive bottleneck*

(a) A specific rule will produce the metabolite: the simplest example can be shown by a system which consists of two rules: $r_1=(\emptyset, 10A, \emptyset, \emptyset, 1.0)$ and $r_2=(10A, \emptyset, \emptyset, \emptyset, 0.2)$. r_1 will be activated in any derivation and r_2 only after four derivation steps. Therefore, the metabolite A increases permanently.

(b) More than one rule will produce the same metabolite and only a few rules will consume this metabolite: a simple example of such configurations is: $r_1=(B, 4A, \emptyset, D, 0.8)$, $r_2=(V, 8A, E, \emptyset, 1.0)$, $r_3=(N, 3A, \emptyset, \emptyset, 0.9)$ and $r_4=(6A, H, \emptyset, \emptyset, 0.5)$. The producer rules are superior in numbers and in a higher equent of activation. Therefore, the metabolite rate will increase permanently.

Case 2: *negative bottlenecks (similar to case 1)*

The simultaneous action of all activated rules is the basic element of our detection mechanism. The derivation operator matches the before components of all rules independently with the metabolites of the actual state and calculates all potential activations. In nature only a few biochemical reactions can be activated—but our model allows the visibility of all potential one-step derivations. The simultaneous activation of some rule combinations will produce the bottleneck and signal the potential critical configuration of a metabolic pathway.

Therefore, in example 5.1 any possible combination of this rule set is allowed. But the action of all rules will produce a negative concentration value (negative bottleneck). Our simulation shell does not continue the simulation using negative components because these configurations are without any biological meaning. However, each actual state will be stored (on tape) and all negative components will be changed to 0 (deleted) before starting the next one-step derivation. The following proposition shows that this formalization of the metabolic processes realizes the complete detection of metabolic bottlenecks.

Proposition

The metabolic system is complete in the case of bottleneck detection.

Proof idea:

Let $G=(S, R)$ be a metabolic system, where R represents all biochemical reactions of the metabolic network and S represents the actual start configuration. For each derivation $S, a_1, \ldots, a_n$ with $n \in \mathbb{N}$ represents a_i a state which represents a finite number of metabolites. For each metabolite the concentration rate can be calculated by considering the rate in state a_{i-1} and the concentration rates of the activated rules. This will produce the actual concentration value $x=\mathbb{N}$ If $x \leqslant l$ and $x \geqslant H$, where $h, l \in \mathbb{N}$ denotes a specific threshold, then one of the following cases is true:

Case $x \leqslant 1$:

(1) This component was modified by one rule in the previous derivation step.

In this case the bottleneck is located: this component and the historical pathway of this component, can then be considered by the metabolic system.

(2) This component was modified by more than one rule in the previous derivation step.

In this case the system shows the global bottleneck constellation, which consists of the component with his historical pathway and all activated rules.

Case $x > h$: is similar to case $x \leqslant 1$.

6. RELATED WORKS

The simulation of metabolic processes is based on specific models, which can be classified into the class of abstract, discrete and analytical models. The abstract models based on automata and logical models which allow the global discussion of fundamental aspects [7, 8]. The goal of analytical models is the exact quantitative simulation where the analysis of kinetic features of enzymes is important. The paper of Waser *et al.* [18] presents a computer simulation of phosphor-fructocinase. This enzyme is part of the glycolyse metabolism and catalyses a chemical reaction. Waser *et al.* model all kinetic features of the metabolic reaction by computer simulation. This computer program is based on chemical reaction rules which are described by differential equations. Franco and Canelas simulate the purine metabolism by differential equations where each reaction is described by the relevant substance and the catalytic enzyme using the Michaelis constant of each enzyme [19]. Discrete models are based on state transition diagrams. Simple models of this class are based on simple production units which can be combined. Overbeek presented an amino acid production system where a black-box with an input set and an output set describes a specific production unit [21]. The graphical model of Kohn and Letzkus [22, 23], which allows the discussion of metabolic regulation processes, is representative for the class of graph theoretical approaches. They expand the graph theory by specific functions which allow to modelling of dynamic processes. In this case the approach of Petri nets is a new method. Reley *et al.* [24] presented the first application of Petri nets in the work field of molecular biology. This formalism is able to model metabolic pathways. The highest abstraction level of this model class is represented by expert systems [25] and object oriented systems [26]. Expert systems and object oriented systems are developed by higher programming languages (Lisp, C++) and allow the modelling of metabolic processes by facts/classes (proteins and enzymes) and rules/classes (chemical reactions). The grammatical formalization is able to model complex metabolic networks [27, 28]. Based on this formalization the metabolic system was developed as a specific rule-based system.

7. SUMMARY

In the research field of biotechnology and biomedicine modelling and simulation of metabolic pathways, gene expression, and complex metabolic networks is of interest. Numerous models are available which can be classified into abstract, analytical and discrete models. Nowadays discrete models are based on new methods of computer science. The disadvantage of these models and simulation shells is that no approach allows the modelling of complex metabolic networks.

Database systems which represent all sequenced gene structures and/or amino acids of enzymes are complex. Moreover, many metabolic pathways are known and collected by Boehringer [5]. In the research field of molecular biology the necessity of integrative systems is realised [14]. Therefore, Karp developed the first information system which represents genes (sequences, function), enzymes (amino acids, function and structure), and metabolic pathways of *E. coli* [6]. The fixed data representation is the disadvantage of such systems.

New methods in this research field are probabilistic networks and grammatical formalizations [10, 27]. Both concepts are suitable because they represent the natural behaviour of these systems. Biochemical reactions are defined by rules and the metabolites of the biosystem are represented by the actual configuration. Based on this formalization we developed an integrative rule based model which allows the modelling and simulation of complex metabolic processes. Our formalization is based on the theory of semi-Thue systems with respect to the analyzed features of metabolic processing. These features are parallel, probabilistic, dynamic and data flow processing. Moreover, we defined a universal rule which is able to model all different biochemical reactions using the biochemical terms of notation. Our simulation shell is programmed in C and works on the SUN workstation [20]. However, this is the first interactive simulation shell

to be used in modelling complex metabolic networks. This model and simulation shell can be used to support the discussion of specific biochemical questions. Therefore, this system can be used to analyse and synthesise metabolic processes. The detection of genetic or metabolic defects, caused by metabolic bottlenecks, can be performed using the metabolic system. In the field of biomedicine specific metabolic configurations must be detected. Such configurations are a sign of genetic defects, which can be repaired by DNA recombination. In this paper we have presented a derivation operator which is able to detect all metabolic bottlenecks [19]. However, the completeness of our strategy is given.

REFERENCES

1. E. Gardner, E. Simmons and D. Snustad, *Principles of Genetics*. New York (1991).
2. G. v. Heijne, *Sequence Analysis in Molecular Biology*. Academic Press, San Diego (1987).
3. M. Mavrovouniotis, G. Stephanopoulos and G. Stephanopoulos, Computer-aided synthesis of biochemical pathways, *Biotechnol. Bioengng* **36**, 1119–1131 (1990).
4. J. Bailey, Toward a science of metabolic engineering, *Science* **252**, 1668–1674 (1991).
5. G. Michal, *Biochemical Pathways*. Boehringer Mannheim, Penzberg (1993).
6. P. Karp, A knowledge base of the chemical compounds of intermediary metabolism, *CABIOS* **8**, 347–357 (1992).
7. S. Kauffman, Metabolic stability and epigenesis in randomly constructed genetic nets, *J. theor. Bio.* **22**, 437–467 (1969).
8. R. Thomas, Regulatory networks seen as asynchronous automata: a logical description, *J. theor. Biol.* **153**, 1–23 (1991).
9. J. Hopcroft and J. Ullman, *Introduction to Automata Theory, Languages and Computatioon*, Addison-Wesley (1990).
10. R. Hofestädt, A simulation shell to model metabolic pathways, *J. Syst. Analysis Modelling Simulation* **11**, 253–262 (1993).
11. R. Hofestädt, *Grammatical formalization of metabolic processes*. in *Proceedings Intelligent Systems for Molecular Biology*, L. Hunter and J. Shavlik, Eds, pp. 181–189 MIT/AAAI Press, (1993).
12. M. Mavrovouniotis, Group contributions for estimating standard Gibbs energies of formation of biochemical compounds in aqueous solution, *Biotechnol. Bioengng* **36**, 1070–1082 (1990).
13. R. Hofestädt, Detection of metabolic bottlenecks, *Proceedings 12th International Congress of the European for Medical Informatics*, Lisbon 1994, P. Barahona, M. Velos and J. Bryant, Eds, 121–127 (1994).
14. D. Fell, Metabolic control analysis: a survey of its theoretical and experimental development, *Biochem. J.* **286**, 313–330 (1992).
15. B. Lewin, *Genes*. Wiley, New York (1990).
16. F. Jacob and J. Monod, Genetic regulatory mechanisms in the synthesis of proteins, *J. Mol. Biol.* **3**, 318–356 (1961).
17. L. Michaelis and M. Menten, Die Kinetik der Invertinwirkung, *Biochem. Z.* **49**, 333–369 (1913).
18. M. Waser *et al.* Computer modeling of muscle phosphofructokinase, *J. theor. Biol.* **103**, 295–312 (1983).
19. R. Frnaco and E. Canela, Computer simulation of purine metabolism, *European J. Biochem.* **144**, 305–315 (1984).
20. R. Hofestädt and F. Meineke, METABOLICA–a rule based system to model metabolic processes, *Proceedings International Symposium and Workshop: MacroMolecules, Genes, and Computers*, T. Smith (Ed.). Waterville Valley, Boston (1993).
21. R. Overbeek, Logic programming and genetic sequence analysis: a tutorial, *Proceedings of Logic Programming*, MIT Press (1992).
22. M. Kohn and W. Letzkus, A graph-theoretical analysis of metabolic regulation, *J. theor. Biol.* **100**, 293–304 (1982).
23. M. Kohn and R. Lemieux, Identification of regulatory properties of metabolic networks by graph theoretical modeling, *J. theor. Biol.* **150**, 3–25 (1991).
24. R. Reley, M. Mavrovouniotis and M. Liebman, Petri net representations in metabolic pathways, in *Proceedings Intelligent Systems for Molecular Biology*, L. Hunter and J. Shavlik, eds, pp. 341–349. MIT/AAAI Press (1993).
25. D. Brutlag, D. Galper and D. Millis, Knowledge-based simulation of DNA metabolism: prediction of enzyme action, *CABIOS* **7**, 9–19 (1991).
26. H. Stoffers *et al.*, METASIM: object-oriented modelling of cell regulation, *CABIOS* **8**, 443–449 (1992).
27. J. Collado-Vides, A transformational-grammar approach to the study of the regulation of gene expression, *J. theor. Biol.* **136**, 403–425 (1989).
28. R. Hofestädt, Modelling molecular counters with stochastic cellular grammars, *Informatik, Biometrie und Epidemiologie in Medizin und Biologie* **4**, 169–177 (1993).

About the Author—RALF HOFESTÄDT studied computer science and biology at the University of Bonn (1979–1985). Afterwards he worked at the Department of Theoretical Computer Science at the University of Bonn (1985–1990) where his research field was modelling metabolic processes using formal languages. In 1990 he finished his Ph.D. thesis at the University of Bonn. From 1990 to 1994 he was assistant professor in the research field of parallel systems at the

University of Koblenz-Landau. In 1994 he presented his habilitation thesis at the University of Koblenz-Landau. This thesis included the theory of metabolic rule based systems. He is presently working at the Department of Medical Computer Science at the University of Leipzig.

About the Author—FRANK MEINEKE studied computer science at the University of Koblenz (1987–1994). He is presently working at the Department of Medical Computer Science at the University of Leipzig.