

An Integrated Qualitative and Quantitative Biochemical Model Learning Framework Using Evolutionary Strategy and Simulated Annealing

Zujian Wu¹ · Wei Pang² · George M. Coghill²

Received: 16 August 2014 / Accepted: 20 April 2015 / Published online: 3 May 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract Both qualitative and quantitative model learning frameworks for biochemical systems have been studied in computational systems biology. In this research, after introducing two forms of pre-defined component patterns to represent biochemical models, we propose an integrative qualitative and quantitative modelling framework for inferring biochemical systems. In the proposed framework, interactions between reactants in the candidate models for a target biochemical system are evolved and eventually identified by the application of a qualitative model learning approach with an evolution strategy. Kinetic rates of the models generated from qualitative model learning are then further optimised by employing a quantitative approach with simulated annealing. Experimental results indicate that our proposed integrative framework is feasible to learn the relationships between biochemical reactants qualitatively and to make the model replicate the behaviours of the target system by optimising the kinetic rates quantitatively. Moreover, potential reactants of a target biochemical system can be discovered by hypothesising complex reactants in the synthetic models. Based on the

biochemical models learned from the proposed framework, biologists can further perform experimental study in wet laboratory. In this way, natural biochemical systems can be better understood.

Keywords Evolutionary algorithms · Heuristic algorithms · Qualitative model learning · Quantitative model learning · Systems biology

Introduction

Understanding inherent mechanisms and principles in the biochemical systems is one of the main tasks when modelling such systems. To effectively investigate a biochemical system of interest, *in silico* analysis can be performed to reveal and formalise the underlying cellular functions and biochemical processes. Two different but complementary methods, quantitative and qualitative model learning approaches [3], can be applied to model biochemical systems: a given cellular system can be described and analysed mathematically in a quantitative manner until desired biochemical behaviour is replicated in a virtual cellular environment, for instance, a web-based environment for kinetic modelling and dynamic simulation of cellular networks—WebCell [36]; meanwhile, a biochemical system can be qualitatively modelled and identified through qualitative model learning (QML) [49, 52, 53] when only incomplete knowledge and imperfect data are available. The above facts motivate us to develop an integrative qualitative and quantitative model learning framework, and we expect that by making use of the advantages of both learning approaches better learning performance will be achieved to assist wet-laboratory research.

Zujian Wu and Wei Pang have contributed equally to this work.

✉ Wei Pang
pang.wei@abdn.ac.uk
Zujian Wu
zujian.wu@jnu.edu.cn
George M. Coghill
g.coghill@abdn.ac.uk

¹ College of Information Science and Technology, Jinan University, Guangzhou 510632, Guangdong, People's Republic of China

² School of Natural and Computing Sciences, University of Aberdeen, Aberdeen AB24 3UE, Scotland, UK

In quantitative modelling approaches, a dynamic biochemical system is mathematically represented to model molecular mechanisms at a quantitative level, and interactions of molecules may be discovered through such modelling process. In further biochemical analysis and wet-laboratory experiments, more biochemical assumptions may be suggested and verified with the help of such precise quantitative analysis. In addition, cell–cell interactions can also be studied through quantitative simulation or system identification [37].

With the recent development of experimental techniques, large omics data have been made available, and this makes it possible to employ quantitative modelling approaches to analyse the dynamics of biochemical systems at the genomic level and address various biochemical issues. For instance, metabolome, fluxome, transcriptome and/or proteome can all be improved and completed by the use of quantitative approaches. Moreover, topologies of biochemical systems can be identified [29, 65] and parameter values can be numerically determined or estimated [32, 43].

In qualitative modelling approaches, qualitative information extracted from imprecise and incomplete data is used to model real-world problems, which becomes the task of qualitative reasoning (QR) [20, 35]. Continuous aspects of a given dynamic system, for instance, space, time and quantity, can be represented or inferred automatically in QR. In QR-based research, qualitative values, such as, *high*, *medium*, *low*, *zero*, *positive* and *negative*, can be used to describe complicated dynamic systems, instead of using precise numerical values. Therefore, behaviours of target biochemical systems can be predicted and reasoned qualitatively *in silico* with the support of QR [30].

One of the subfields of QR, qualitative differential equation model learning (QML) [48], involves the construction of qualitative differential equation models of dynamic systems from observed data and existing knowledge. QML has been well developed in the last two decades, and examples of QML systems include MISQ [60], GENMODEL [26], QOPH [15], QSI [10], ILP-QSI [17], and the most recent QML-Morven [47, 50] as well as its scalable version, QML-AiNet [51]. QML is a complementary approach to quantitative system identification [37] and it works well in reasoning dynamic behaviours of biochemical systems, especially when only noisy and sparse experimental data are available. QML can infer plausible qualitative models for a given target biochemical system (for instance, in [70] an integrative modelling approach is studied for stepwise qualitative exploration of biochemical interactions), and these plausible qualitative models could be directly examined by biologists or further

refined by quantitative approaches depending on specific research tasks.

In a biochemical system, behaviours of reactants, including the interactions between these reactants, are determined by kinetic laws and concentrations of species. In the presence of abundant quantitative data and sufficient knowledge, it is straightforward to employ sophisticated quantitative modelling approaches and tractable computational tools to first build quantitative models by presuming the model structures and then fit the numeric parameters of these models. However, when available data and knowledge are not enough to assume model structures and perform quantitative analysis, it is essential to use qualitative model learning approaches to first qualitatively infer the model structure and then analyse biochemical systems at a qualitative level [47, 63].

In this research, we propose an integrative framework to explore the biochemical model space at both qualitative and quantitative levels. More specifically, an evolution strategy (ES) [5, 62] is employed in the qualitative approach to perform effective selection and composition of functional modules and heuristically evolve model structures towards the target biochemical systems. Then simulated annealing (SA) [31] is used to quantitatively optimise model kinetic rate constants obtained from the qualitative approach. The motivation of employing ES and SA in our framework is that these two metaheuristics algorithms are suitable for searching qualitative and quantitative model space, respectively, and they have been proven to be effective in similar problems in our previous work [66, 67]. For a comprehensive review of employing ES and SA in biochemical systems identification, the reader is referred to [68], in which the optimisation of model structure and kinetic rates was studied by the hybrid use of ES and SA, respectively. In addition, for a general review of evolutionary algorithms and related issues on their applications, the reader is referred to [46]. It is also worth mentioning that in [22], a co-evolutionary algorithm is used to infer differential equation models of the target system from time series data in a reverse engineering manner.

The rest of this paper is organised as follows. In Sect. 2, biochemical components and models are briefly described and how these components work together are also illustrated. We introduce the background knowledge about the quantitative and qualitative modelling approaches in Sects. 3.2 and 3.1, respectively. In Sect. 4, we present the integrative quantitative and qualitative modelling framework. Some case studies and simulation results with analysis are reported in Sect. 5. In Sect. 6, we compare our system with relevant ones. Finally, Sect. 7 concludes our research with discussions on future work.

Representation of Biochemical Models

In this research, we reuse the definitions of biochemical building blocks (components) and composition of these components (models), as reported in our previous work [66, 69], to represent the target biochemical system in the form of Petri nets [44]. Components are basic building blocks, and a synthetic model is composed from these components by following a set of composition rules [66]. Definitions of biochemical components and models as well as the applications of composition rules are detailed in our previous work [66, 68]. In this section, a brief introduction about the components and model composition is given as follows.

In general, biochemical components can be defined by Petri nets, which can also be applied to define the two patterns (binding and unbinding) for instantiating components. As shown in Eqs. (1) and (2), the binding and unbinding patterns are defined for further model learning tasks.



In Eq. (1), P_1 represents a reactant acting as a substrate; P_2 denotes a reactant acting as an enzyme; P_3 ($P_3 = P_1|P_2$) is a complex synthesised from P_1 and P_2 at a reaction kinetic rate constant k_1 , which is for a synthetic process. It should be noted that in this research we use the symbol ‘|’ joining the labels of the two reactants to represent a complex.



In Eq. (2), k_2 is a reaction kinetic rate constant for a disassociation process; Complex P_3 is either disassociated to two reactants P_1^* and P_2^* , which form the complex P_3 itself (an inverse process of the binding pattern presented in Eq. (1), thus $P_1^* = P_1$ and $P_2^* = P_2$), or converted into a new product ($P_1^* \neq P_1$) and an enzyme ($P_2^* = P_2$) which is one of the reactants forming the complex P_3 described in Eq. (1).

We can take the binding pattern for the instantiation of a component, and in the binding process, two reactants are combined to form a complex reactant. We can also take the unbinding pattern for instantiating a component, and in the unbinding process, a complex reactant is divided into two reactants. In this way, complicated components consisting of more than three reactants can be represented by the composition of components instantiated from these two patterns. As mentioned before, in this research qualitative model, learning with ES is used to evolve model structures generated from these components, and kinetic rate constants associated with the interactions between the reactants in these components are optimised by quantitative model learning with SA.

We here briefly introduce the formation of components: given two sets of reactant labels provided by the user: a set of reactants as species $S_{species}$ and another set of reactants as enzymes $S_{enzymes}$. Each element in $S_{species}$ is selected in turn to be combined with each element in $S_{enzymes}$ to produce a complex and a new reactant, based on the mass-action 1 (MA1) kinetic law [6]. For instance, species A from $S_{species}$ is combined with enzyme E from $S_{enzymes}$ to form a complex $A|E$ and a new reactant AP . The choice of kinetic laws used to describe biochemical models depends on the pathway that we will study. In practice, two kinetic laws are often used to describe a biological pathway: the law of mass-action [24] and Michaelis–Menten kinetics [41]. Mass-action kinetics indicate that the rate of a biochemical interaction is proportional to the concentrations of biochemical reactants. A system of nonlinear ODEs defines the rate of change of reactants within experimental time. Thus, an enzyme acts as a catalyst facilitating the reaction in an enzymatic reaction, which is possible to be investigated by biological experiments. Based on literature reviews [64], it is reasonable to use the law of mass-action to model the target biochemical system, which is the RKIP pathway (introduced later in Sect. 5.1). While the law of Michaelis–Menten assumes that the substrate is in instantaneous chemical equilibrium with the complex, and this assumption is not suitable for describing the RKIP pathway. Moreover, the RKIP pathway has been well studied in [13, 21] by using the mass-action law to describe the quantitative relationships between substrates and complex. In order to follow previous research results and develop new modelling methods based on the same biochemical pathway, in this research we also use the mass-action law to describe a pathway and do not consider Michaelis–Menten kinetics at this stage. Although the Michaelis–Menten kinetics are not used in this research, it is possible to use a hybrid method of employing both Michaelis–Menten and mass-action laws to study different parts/scales of biochemical systems as in our previous work [49, 52]. However, this is beyond the scope of this research.

A synthetic enzymatic reaction is shown in Eq. (3). In this equation, the symbol ‘ \rightleftharpoons ’ indicates that the reaction is reversible; the symbol ‘ \rightarrow ’ represents a non-reversible reaction; the symbol ‘|’ indicates that a complex reactant is generated from the two reactants (as described before); and the letter P after the species label A means a new product generated from A .



Therefore, three atomic components can be obtained from the enzymatic reaction shown in Eq. (3): ‘ $A + E \rightarrow A|E$ ’, ‘ $A|E \rightarrow A + E$ ’ and ‘ $A|E \rightarrow AP + E$ ’.

Given a set of reactants, components can be generated from these reactants and used as modules to construct a biochemical model. Components of a model represented by Petri nets are connected by merging the same ‘nodes’ (Places) among these components [66]. Figure 1 shows a Petri net model for an enzymatic reaction consisting of three components connected with each other by merging the same reactants. ‘T1, T2, and T3’ stand for ‘reactions’ (Transitions), whose kinetic rate constants are marked as ‘k1, k2, and k3’ on the places, respectively. The numbers (for example, 4 and 5) associated with the places represent the initial concentrations of reactants [68].

Qualitative and Quantitative Model Learning Approaches

Qualitative Model Learning

Qualitative States

A dynamic system can be described at a qualitative level, and its important behavioural properties are captured by a set of qualitative states and possible transitions between these states [19, 35]. A qualitative state is a complete assignment of qualitative values to all variables in the system and considered as a ‘snapshot’ of the system. The dynamic system under investigation could potentially demonstrate such possible qualitative states under specific conditions and a correct model built for the system should produce these qualitative states (and only these states if all variables are known).

Table 1 shows a set of qualitative states derived from a qualitative model [71]. Each row in this table represents an individual qualitative state. For each variable, its magnitude and rate of change of the current state are illustrated by the qualitative signs: *pos* (positive), *zer* (zero), and *neg* (negative). For example, if the qualitative value of a variable *A* is $\langle zer, pos \rangle$, this means the magnitude of *A* is zero and the rate of change is positive, which indicates that the value of *A* is increasing.

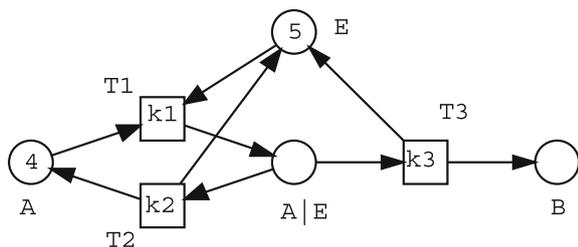


Fig. 1 A graphical representation of a Petri net model for an enzymatic reaction consisting of three components

For the legal transitions between qualitative states, transition rules (e.g., rules presented in QSIM [35]) are employed to calculate them. A sequence of qualitative states forms a *qualitative behaviour*, and the terminal states of the qualitative behaviour are often equilibrium ones, in which all variables remain constant [49].

Qualitative Differential Equations

One of the well-studied formalisms of the qualitative model used by QR is qualitative differential equations (QDEs), which have been used by QSIM [33, 35] and Morven [14, 16].

Ordinary differential equations (ODEs) quantitatively describe the behaviour of a dynamic system. A QDE model is the abstraction of a set of ODE models sharing the same model structure but with varying parameter values. Figure 2 shows the relationships between the real-world dynamic systems, ODE models, and QDE models in terms of quantitative and qualitative behaviour.

Formally, a qualitative differential equation (QDE) is a tuple, $\langle V, Q, C, T \rangle$, each of which is defined in [35] and briefly described as follows: *V* is a set of qualitative variables, each of which is a ‘reasonable’ function of time; *Q* is a set of quantity spaces, one for each variable in *V*; *C* is a set of qualitative constraints applying to the variables in *V*, and each variable in *V* must appear in some constraints; *T* is a set of transitions between qualitative states. In summary, a QDE is the conjunction of all its qualitative constraints, and each constraint links some qualitative variables.

According to above definitions, qualitative constraints constitute a QDE, which restricts the generation of possible qualitative states. In addition, a quantity space is composed of several qualitative values that could be taken by a variable. In this research, for the application of ES to evolve the topologies of biochemical models, we use QDEs to represent qualitative models in terms of structure, and for all variables we used the signs quantity space [30], which is composed of three qualitative values: *positive*, *zero* and *negative*.

Quantitative Model Learning

Quantitative modelling of biochemical processes has been intensively used in biochemical research [1, 7, 12, 55, 61]. Preliminary quantitative analysis of biochemical systems has been very difficult to perform, due to the inherent complexity of biochemical processes [25, 57]. Biochemical systems have been modelled by employing Petri nets theory [44], including enzymatic cascades and synergistic binding of ligands to enzymes [3, 38, 45]. In primary research, the issues of quantitative analysis of metabolic

Table 1 A set of qualitative states

State ID	A	AP	B	BP
1	⟨zer , pos⟩	⟨pos , neg⟩	⟨pos , neg⟩	⟨pos , neg⟩
2	⟨pos , pos⟩	⟨pos , pos⟩	⟨zer , pos⟩	⟨pos , neg⟩
3	⟨pos , zer⟩	⟨pos , zer⟩	⟨pos , neg⟩	⟨pos , neg⟩
4	⟨pos , pos⟩	⟨pos , pos⟩	⟨pos , neg⟩	⟨pos , neg⟩
5	⟨pos , neg⟩	⟨zer , pos⟩	⟨pos , neg⟩	⟨pos , neg⟩
6	⟨zer , pos⟩	⟨pos , neg⟩	⟨pos , zer⟩	⟨pos , neg⟩
7	⟨pos , zer⟩	⟨pos , neg⟩	⟨pos , zer⟩	⟨pos , neg⟩
8	⟨pos , neg⟩	⟨pos , zer⟩	⟨pos , zer⟩	⟨pos , neg⟩
9	⟨pos , zer⟩	⟨pos , zer⟩	⟨pos , zer⟩	⟨pos , neg⟩
10	⟨pos , neg⟩	⟨pos , pos⟩	⟨pos , zer⟩	⟨pos , neg⟩
11	⟨zer , zer⟩	⟨zer , zer⟩	⟨zer , zer⟩	⟨zer , zer⟩
12	⟨pos , pos⟩	⟨pos , zer⟩	⟨zer , zer⟩	⟨zer , zer⟩
13	⟨pos , pos⟩	⟨pos , zer⟩	⟨pos , zer⟩	⟨pos , zer⟩
14	⟨pos , pos⟩	⟨pos , neg⟩	⟨pos , pos⟩	⟨pos , zer⟩

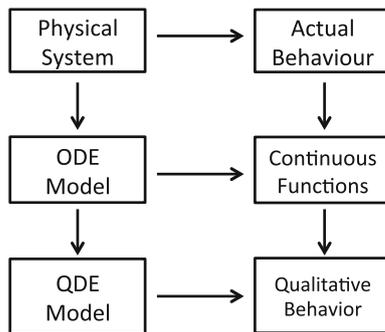


Fig. 2 This diagram is a slightly modified version of the one presented in [34]. The diagram shows that all models are abstractions of the real-world systems. Qualitative models are related to ordinary differential equations, but are more expressive of incomplete knowledge

pathways have been studied by Reddy et al. [58, 59]. Research efforts on applying Petri nets to model biochemical processes and the description of current challenges of constructing biochemical pathways by Petri nets are detailed in [4, 11, 39, 56].

In biochemistry, a chemical reaction is a process of converting molecules of reactants known as substrates into products within a specific time period. Biochemical systems are composed of interacting molecular species, whose dynamics are governed by the corresponding chemical reactions. A biochemical model is fully characterised by the initial concentration of each molecular species and the specifications of the reactions with their kinetic rate laws. The dynamics of the molecular species can be described by an ordinary differential equation (ODE) as shown below [67]:

$$\frac{dX_i}{dt} = \sum_j \left(\mu_{ij} \cdot \gamma_j \prod_{k=1}^m X_k^{f_{jk}} \right), \quad (4)$$

where X_i represents one entity of the model, for instance, the metabolite concentration, protein concentrations, or the level of gene expression; j represents the biochemical reaction that affects the dynamics of the species; μ_{ij} is the stoichiometric coefficient; γ_j is the rate constant; f_{jk} stands for kinetic order; and m denotes the number of reactants involved in the reactions.

The use of Petri nets in biochemical systems comes as a natural and intuitive solution, as biochemical reactions are inherently bipartite, and Petri nets can describe relations between biochemical entities [44]: the biochemical reactions and entities can be mapped onto transitions and places, respectively. A continuous Petri net can be represented by a system of ODEs which describes biochemical reactions in models. In this research, we map ODEs from a set of biochemical reactions and optimise the associated kinetic rate constants quantitatively.

An Integrated and Complementary Biochemical Model Learning Framework

The Modelling Framework

In this section, we propose an integrated and complementary qualitative and quantitative biochemical modelling framework (2QBMF) for modelling biochemical systems, and this framework can identify the structure and kinetic rate constants simultaneously.

Figure 3 illustrates the details of the modelling framework. First, initial biochemical model seeds are synthesised by employing a pair of operators ‘Addition and Subtraction’ to compose biochemical components from a pre-specified component library, before we perform model topological exploration and rate constants optimisation.

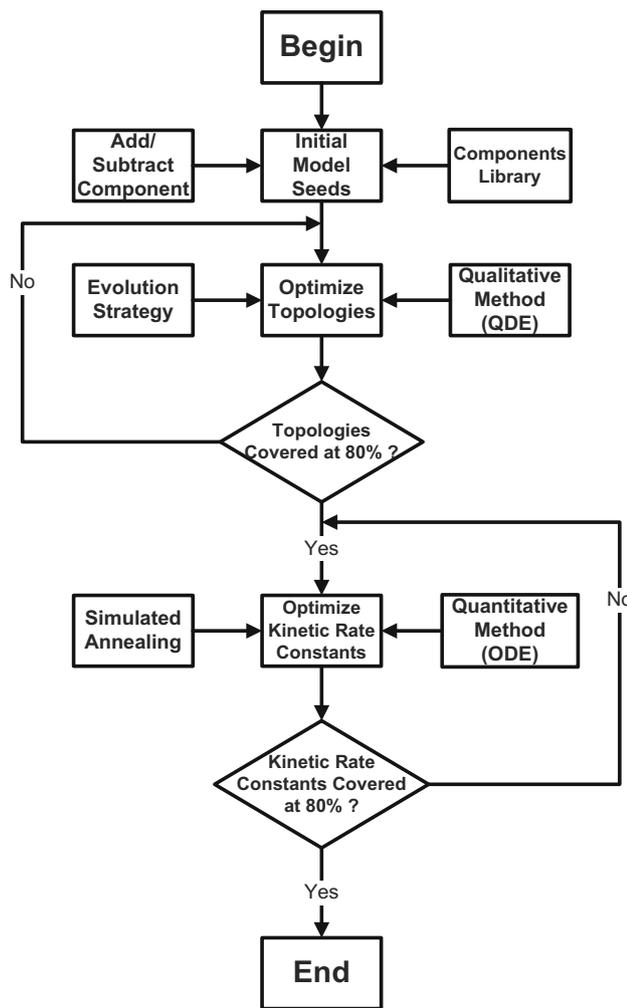


Fig. 3 A complementary qualitative and quantitative-based biochemical modelling framework (2QBMF)

Qualitative model learning (QML) presented in Sect. 3.1 is then applied to explore the qualitative model space, that is, to explore all possible model structures. As the implicit search space for model structure may be huge, to effectively perform such exploration we use evolution strategy (ES); once we determine the model structure, quantitative model learning (QuatML) described in Sect. 3.2 is employed to optimise the kinetic rate constants, and we use simulated annealing (SA) to perform such optimisation.

While the model structure is finely tuned by the component addition and subtraction operations of ES in an iterative manner, qualitative differential equations (QDEs) converted from the Petri net model are used to describe and analyse the qualitative states generated by simulating each modified model. Modification of model structure is accepted, if the interactions between biochemical reactants in the model can achieve a high rate of coverage (e.g. 80%), which is defined as the percentage of the observed qualitative states covered by the QDE model.

After the model structure has been explored, kinetic rate constants associated with the biochemical reactions are globally optimised by employing ordinary differential equations (ODEs) in the SA process. The ODEs are performed to mathematically analyse influence of kinetic rate constants on the reactant concentration. If the target rate constants are achieved at a high rate coverage (e.g. 80%), the optimisation of the reaction rates is accepted.

In this way, the metaheuristic algorithms ES and SA are used to carry out qualitative analysis and quantitative estimation in a complementary manner. The heuristic evolutionary searching mechanism of ES supports the model structure exploration, and the global search strategy of SA helps the optimisation of kinetic rate constants to drive the model behaviours approaching the target ones.

Qualitative Model Structure Evolution

In the qualitative modelling approach, each individual evolved by ES is a Petri net model. At each generation of the evolutionary process, the topology of the model is evolved through the application of genetic addition and subtraction operators. An evolutionary algorithm ($\mu+\lambda$)-ES [5] is employed to iteratively evolve model structure at a qualitative level. In order to test the evolutionary qualitative modelling process in a simplest scenario, we choose a simple (1 + 1)-ES to generate offspring models. Further, advanced ($\mu+\lambda$)-ES will be performed and investigated more thoroughly in the future.

Algorithm 1 ES based qualitative modelling approach to evolve biochemical systems

Require: A set of model seeds M_i ($i=1,2,\dots,N$) for a target biochemical system; given qualitative states QS_T .
Ensure: Developed models of the target system based on qualitative information of evolved M_i .

- 1: **while** Maximum number of ES generations is not achieved **do**
- 2: $M_i' \leftarrow M_i$, by adding a component C to M_i ; or by subtracting a component C from M_i ;
- 3: Apply QDEs to qualitatively simulate M_i' and produce $QS_{M_i'}$;
- 4: Perform calculation of the fitness function $F(QS_T, QS_{M_i'})$;
- 5: **if** $F(QS_T, QS_{M_i'}) = 1$ or 80% of the interactions are obtained **then**
- 6: Call SA based quantitative modelling approach in Algorithm 2 to quantitatively optimize $QS_{M_i'}$;
- 7: **else**
- 8: Reject M_i' and reuse M_i in the next ES generation;
- 9: **end if**
- 10: **end while**
- 11: **Return** a set of qualitatively evolved models for the target biochemical system.

The pseudo-code of employing ES to perform qualitative model learning of biochemical systems is shown in Algorithm 1. A set of model seeds M_i is prepared for structurally approaching the target biochemical system.

These model seeds can be obtained from literature or knowledge of wet-laboratory experiments. In general, a composed model M_i' can be evolved by applying addition or subtraction operators [66] to M_i . In this research, the addition and subtraction operators are applied by a fixed number of generations, which is adopted based on our previous experience in quantitative simulations [67]. Model M_i' is simulated with JMorven [8], a qualitative simulation engine. JMorven will generate a set of qualitative states QS for comparison with QS_T in a fitness function F. If the fitness value is equal to 1 (the range of fitness value is from 0 to 1, and the bigger the better) or 80 % of target interactions are obtained (the interactions between reactants in a model indicate the structure of the model), an SA-based quantitative modelling approach in Algorithm 2 will be used to optimise the kinetic rate constants. Otherwise, the mutated M_i' is rejected and the mutation is performed in the next generation. The qualitative model learning process is terminated when the ES stopping conditions are satisfied. At the end of the modelling process, a set of final best qualitative models will be obtained with explored topologies by the qualitative analysis.

Note that, one of the predefined conditions, obtaining 80 % of target interactions, is not triggered during the qualitative optimisation process in all simulations performed in this research. The number 80 % was inspired by the Pareto Principle [54] (or called ‘80–20 rule’) which states that roughly 80 % of the effects come from 20 % of the causes for many events. Thus, in our research, it is possible to allow potential good qualitative models to be preserved during the evolutionary process by considering a small group of mutated models with 80 % interactions generated. But after checking our simulation results, we found that the condition of generating 80 % interactions was not hold and switching from Algorithm 1 to Algorithm 2 was always not triggered by this condition. The reason was that fitness values of models under constructions always first reach the value ‘1’ before 80 % interactions are generated. However, with an aim to obtain good qualitative models from simulations in further research, we still keep the condition of 80 % interactions as one of criteria for switching from qualitative modelling in Algorithm 1 to quantitative modelling in Algorithm 2.

Quantitative Model Parameter Optimisation

In the quantitative modelling approach, ODEs are used to mathematically describe biochemical models and parameters of the models are investigated. In this research, kinetic rate values are parameters to be optimised by SA. SA is a heuristic algorithm suitable for globally searching optimal solutions in a very large solution space, and it can avoid getting trapped into local optima. As mentioned in

Sect. 3.1.2, a set of ODEs sharing the same structure can be represented by a QDE model; thus, a QDE model obtained from qualitative model learning can be quantitatively optimised to obtain corresponding ODEs. This is achieved by using the QDE model as the skeleton of an ODE model and estimating the parameter values of this ODE model.

Algorithm 2 Quantitative modelling approach for optimization of biochemical systems based on SA

Require: A set of models M_i ($i=1,2,\dots,N$) with corresponding initial kinetic rate values associated with biochemical interactions, iteration number $IterNum$, cooling rate α , system temperature T and minimum stop temperature T_{Min} .
Ensure: Optimized kinetic rate constants within the M_i .

- 1: **while** $T > T_{Min}$ **do**
- 2: **while** $IterNum! = 0$ **do**
- 3: Mutate kinetic rate values of M_i by *Gaussian* distribution $N(\mu, \sigma)$;
- 4: Apply ODEs to quantitatively calculate and evaluate mutated rate values in M_i ;
- 5: Accept M_i based on the Metropolis algorithm;
- 6: **end while**
- 7: Reset $IterNum$;
- 8: Lower T by α ;
- 9: **end while**
- 10: **Return** M_i with optimized kinetic rate constants.

Algorithm 2 shows the pseudo-code of quantitative optimisation of kinetic rates in models M_i . Models M_i constructed from previous qualitative model learning stage have been evolved qualitatively to obtain their proper model structures, and these models are to be quantitatively investigated, because the values of the associated rates are essential to obtain expected quantitative behaviours. The parameter t is the current SA system temperature ($t = T$), and $IterNum$ is the number of iterations at each system temperature.

Models M_i with optimised rate values are accepted or rejected according to a classical Metropolis mechanism [40]. Accepted M_i are preserved as new start points for the next run of model optimisation. Models M_i with different sets of rate values and structures are optimised heuristically at different SA system temperatures by a cooling rate α , and the whole process will stop when system temperature reaches the minimum temperature T_{Min} .

Note that, due to the probabilistic and random nature of SA [2], a model with a poor estimated fitness value could be generated and accepted. Thus, corresponding biochemical reaction rates associated with different topologies of the models could result in the acceptance of non-optimum models during the model learning process.

Qualitative and Quantitative Models Evaluation

Models are simulated and evaluated after qualitative structural mutation and quantitative optimisation of rate

values. There are two different evaluation methods, one for qualitative and another for quantitative model learning.

Qualitative Fitness Function

As mentioned before, qualitative models explored by ES are simulated with JMorven [8]. JMorven can produce a set of qualitative states given a qualitative model. The given target qualitative states of a biochemical system are compared with the states generated from the qualitative model. The number of matched qualitative states between the target system and a synthetic model is recorded and considered as part of the fitness evaluation.

A qualitative state for the model evaluation purpose is an assignment of N variables which appear in both the target biochemical system and explored model. There could be M qualitative states, and a vector is used to record each of these M qualitative states. In this vector, each element is the assignment of one variable. To evaluate a composed model, the following two sets will be compared element by element: one is the set of qualitative states generated by the composed model, and another is the given set of states demonstrated by the target system. In this way, a fitness value of a qualitative model is calculated by considering the overlapping part of the above two sets.

Figure 4 shows the comparison between a synthetic model and a target biochemical system in terms of qualitative states. A synthetic model produces a set of qualitative states QS_G , and a set of given target qualitative states QS_T is compared with QS_G to calculate the coverage of qualitative states in f_1 . Another comparison is performed by comparing QS_G with the set of all possible states QS_C , in which f_2 is the rate of matched qualitative states produced by the model under estimation. Therefore, qualitative evaluation of a composed model can be described by jointly considering f_1 and f_2 in a fitness function F . Details of the calculation of f_1 , f_2 and F are shown in Eqs. (5)–(7) [71].

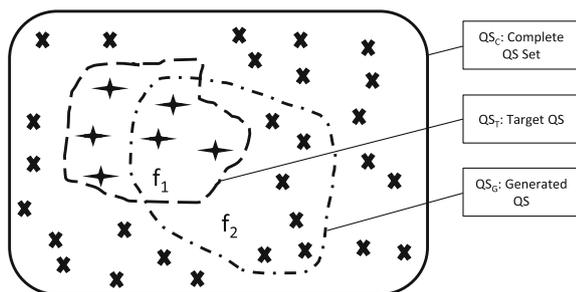


Fig. 4 Comparison between generated qualitative states from a synthetic model and target states in a given biochemical system

$$f_1 = \frac{|QS_G \cap QS_T|}{|QS_T|} \tag{5}$$

$$f_2 = \frac{|QS_G \cap QS_C|}{|QS_C|} \tag{6}$$

$$F = 1 - \frac{1}{1 + f_1 + \frac{1}{1 + f_2}} \tag{7}$$

In the above Eqs. (5)–(7), ‘|·|’ denotes the number of states in the set, $|QS_G \cap QS_C|$ indicates the set of overlapping states in both QS_G and QS_C . Two qualitative states are the same if their assignments of all variables are the same.

The value of f_1 ranges from 0 (worst) to 1 (best), because the more matched qualitative states between QS_G and QS_T , the bigger the value of f_1 . The value of f_2 ranges from 0 (best) to 1 (worse), as the less matched spurious qualitative states between QS_G and QS_C , the better the quality of the generated model. A fitness function F is summarised by standardising f_1 and f_2 , and the value of F ranges from 0 (worst) to 1 (best).

Note that there could be different synthetic reactants in a composed model during the evolutionary model learning process, because of the use of genetic addition and subtraction operators. In this research, we specify the reactants to be compared during the model evaluation process. Thus, we discard a composed model of which all reactants are not in the vector of variables for comparison with the target system.

Quantitative Fitness Function

Quantitative behaviours of a model are compared with the target ones, which are represented in the form of time series data for the concentrations of species, e.g., enzymes, proteins, and complexes. The behaviours of the species in the target system can be obtained from observations of a biochemical system from the wet laboratory and corresponding computational model in dry laboratory. As mentioned before, the optimisation of kinetic rates will result in the model behaviours approaching the target ones.

A set of reference data M_T is used for the target system, and there are N generated time series $X_T = (X_1, X_2, \dots, X_N)$ which represent the behaviours of N species, and $N \geq 1$. There are P data points in each time series $X_i = (x_i^1, x_i^2, \dots, x_i^P)^T$, $i = 1, \dots, N$. There are M time series $X_G = (\hat{X}_1, \hat{X}_2, \dots, \hat{X}_M)$ which describes the behaviours of M species in a constructed model M_G , with P data points in each time series $\hat{X}_j = (\hat{x}_j^1, \hat{x}_j^2, \dots, \hat{x}_j^P)^T$, $j = 1, \dots, M$. Intersection between M_T and M_G is defined by $X_C = X_T \cap X_G = (X_1, X_2, \dots, X_n)$, $1 \leq n \leq N$. The difference between the behaviours of M_T and M_G is calculated by averaging the difference of behaviours of each species in X_C by a paired comparison of the P data points.

As shown in Eq. (8), the difference of behaviours for one species X_k in X_C is measured by the *Euclidean* distance, where η is the total number of compared substrates in X_C [66].

$$d_{M_T, M_G}(X_k) = \frac{1}{\eta} \sum_{k=1}^{\eta} \sqrt{\sum_{t=1}^P (x_k^t - \hat{x}_k^t)^2} \quad (8)$$

Note that, $\eta = n$ if the compared substrates are from the intersection X_C , ($X_C = X_T \cap X_G$, $|X_C| = n$, $1 \leq n \leq N$), and $\eta = n'$ if the compared substrates are from a specific X'_C , ($X'_C \subseteq X_C$, $|X'_C| = n'$, $1 \leq n' \leq n$). In this research, quantitative modelling is a minimisation problem; therefore, the smaller the fitness value, the better the generated model.

Simulation and Analysis

In order to test the effectiveness of our proposed integrative modelling framework, we use a real-world biochemical system, the ‘Ras/Raf-1/MEK/ERK’ signalling pathway [72], which is called ‘The RKIP pathway’ in this research for ease of description. In further research, there will be more key biochemical systems to be used as proof-of-concept examples. Details of the RKIP pathway structure are retrieved from literature as well as experiments, and corresponding qualitative states are abstracted from quantitative values, which are obtained from a simulator Snoopy [27].

We firstly test the feasibility of our learning framework on a small scale: we specify a small number of generations and populations and run experiments with a computer equipped with Intel Core 2 Duo CPU (2.4 GHz) and 4 GB memory; the total number of generations in ES is 100; the number of individual models as model seeds is 20; and in current research, each individual will be mutated definitely without any application of a probability. Thus, individual model is under construction by involving two operators ‘addition’ and ‘subtraction’: addition of one component is performed at each generation, and subtraction of one component is carried out at every two generations which allows model structures to be mutated quickly, according to our previous experience in quantitative simulation [67]; each model seed is an atomic component randomly selected from the component library [66] and applied to the initial population. Various parameter settings of ES and SA (for instance, numbers of generations applying subtraction, addition and crossover in ES; initial temperatures, cooling rates and iteration numbers in SA) have been tested individually throughout the experiments. It appeared that some modelling issues arose as follows on a small scale: if applying the subtraction operator to mutate models more than every five generations, the sizes of model structures would

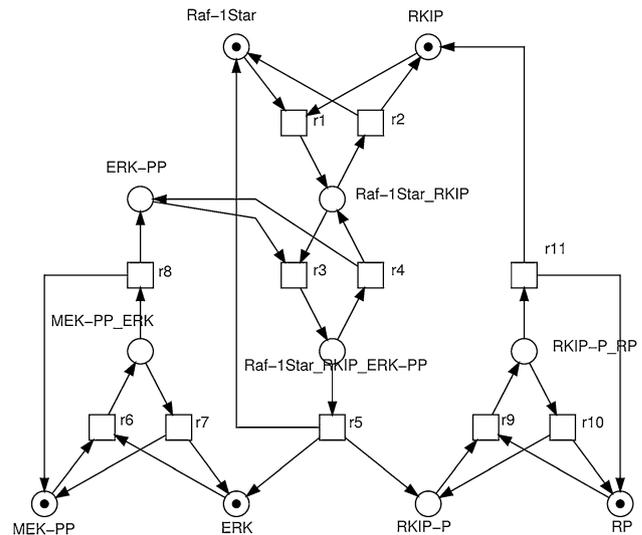


Fig. 5 A graphical representation of the ERK signalling pathway regulated by RKIP. Figure is from [21]

be out of control. If the number of evolutionary generations is more than 100, the simulation process would take a longer time to complete due to the high cost of qualitative and quantitative simulations. If the initial temperature is set too high (e.g. > 100) and the iteration number is > 50 in SA, the convergence of model fitness function would be very slow. Therefore, parameter settings of ES and SA in this research are designed according to performance considerations and empirical selection. A comprehensive study of ES and SA parameter settings will be investigated for their effects on a large scale by employing the high-performance computing (HPC) environment in further research. In this way, simulation and analysis on a large scale would present an overall influence of different algorithm parameter settings on the model learning process.

The RKIP Pathway

Signalling pathways play a pivotal role in many key cellular processes [18]. The abnormality of cell signalling can cause the uncontrollable division of cells, which may lead to cancer. The RKIP pathway is one of the most important and intensively studied signalling pathways, transfers the mitogenic signals from the cell membrane to the nucleus [72]. It is de-regulated in various diseases, ranging from cancer to immunological, inflammatory and degenerative syndromes, and thus represents an important drug target. Ras is activated by an external stimulus, via one of many growth factor receptors; it then binds to and activates Raf-1 to become Raf-1*, or activated Raf, which in turn activates MAPK/ERK Kinase (MEK) which in turn activates Extracellular signal Regulated Kinase (ERK). This cascade (Raf-1 \rightarrow Raf-1* \rightarrow MEK \rightarrow ERK) of protein interaction

controls cell differentiation with the effect being dependent upon the activity of ERK. RKIP inhibits the activation of Raf-1 by binding to it, disrupting the interaction between Raf-1 and MEK, thus playing a part in regulating the activity of the ERK pathway.

With the aim of understanding the role of reactant RKIP in the biochemical pathway, many computational models have been developed and ultimately suggest new therapies [9, 13]. In this research, the RKIP pathway is used to test the proposed modelling framework. Detailed biological description of the RKIP pathway can be found in [13], and a graphical representation of the RKIP pathway in Petri nets from [21] is shown in Fig. 5.

Analysis of Explored Components

Qualitative model learning focuses on the exploration of components in candidate biochemical models. During the model structure learning process, instantiated components based on given patterns are added into model seeds and components are subtracted from the models. Thus, statistical analysis of the frequency of explored components is important for obtaining information about essential biochemical interactions, the problem of which is very difficult to be addressed in wet laboratory. Moreover, structures of generated models may vary from the target biochemical systems because of the component composition. Regarding the aims of generating interest reactants and hidden complexes associated with the biochemical reactions, it is also helpful to generate synthetic reactants involved in the biochemical models, which can be presented to biologists for further experimental examination.

Table 2 shows the top 11 frequently explored components in the qualitatively constructed models. Compared to the structure of the target RKIP pathway, most of these explored components do not exist in the target RKIP pathway:

- Two explored components are in the target RKIP pathway. NO. 1 ' $RKIP + Raf1 \rightarrow RKIP|Raf1$ ' and NO. 4 ' $ERK + MEKPP \rightarrow ERK|MEKPP$ ' components exist in the target RKIP pathway, which means our ES-based qualitative model learning can explore biochemical reactions correctly.
- Other explored components are potential/alternative biochemical reactions in the target RKIP pathway. Most of learned components with high frequency (appearance) are obtained from the analysis of simulation results. After comparing synthetic reactants involved in these components with the target pathway, we can find that alternative biochemical interactions may have influence on the consumption and accumulation of the species concentrations. Thus, this may interest biologists, who can further examine these alternative interactions in wet laboratory.

It is pointed out that functions of these synthetic reactants can be analysed by synthetic biology techniques, which aim to design new biochemical functionalities for specific tasks in real-world applications.

Approximating Species Behaviours

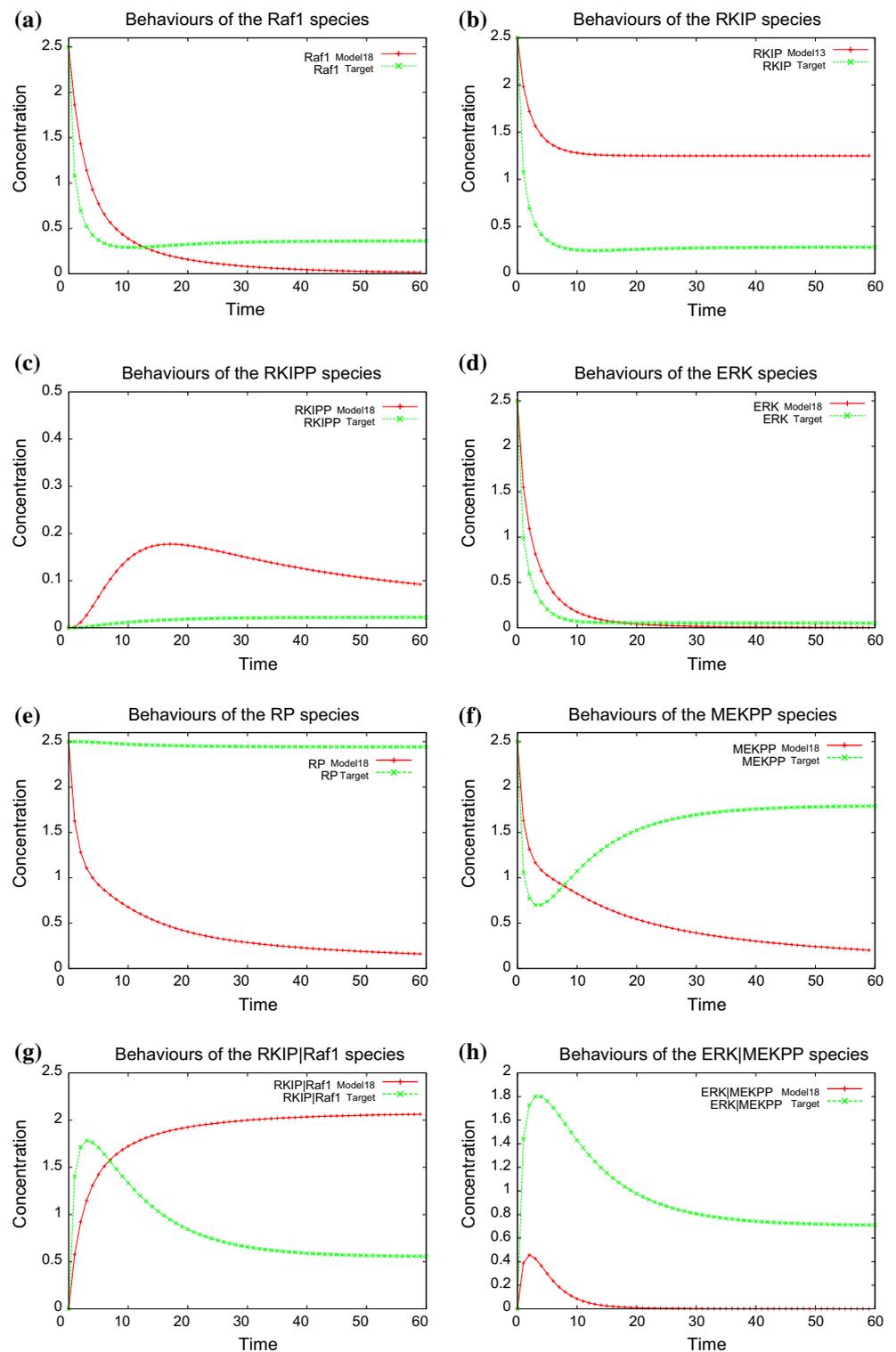
Biochemical characteristics of constructed model are described by the changes in species concentrations, which are determined by the biochemical reactions and associated kinetic rates. Quantitative behaviours of the model are represented by time series data, which is from the process of kinetic rates optimisation by SA. Thus, results of kinetic rate optimisation can be examined by comparing the species behaviours between the learned and target models quantitatively.

Figure 6 presents the comparison between the behaviours of eight target species in one of the best learned models and those in the target RKIP pathway. From a biologist's point of view, the 'shapes' of species behaviours meaning the qualitative change of species concentrations in

Table 2 Frequency of generated components in models

NO.	Reactants and reactions	Frequency	Target reactions
1	$RKIP+Raf1 \rightarrow RKIP Raf1$	405	Yes
2	$MEKPP+RKIP \rightarrow MEKPP RKIP$	407	No
3	$RKIPP Raf1 \rightarrow RKIPPP+Raf1$	415	No
4	$ERK+MEKPP \rightarrow ERK MEKPP$	429	Yes
5	$ERK MEKPP \rightarrow ERK+MEKPP$	429	No
6	$MEKPP RKIPP \rightarrow RKIPPP+MEKPP$	433	No
7	$RKIP+RP \rightarrow RKIP RP$	469	No
8	$MEKPP RKIP \rightarrow RKIPP+MEKPP$	471	No
9	$ERKPP MEKPP \rightarrow ERKPP+MEKPP$	483	No
10	$MEKPP+RKIPP \rightarrow MEKPP RKIPP$	639	No
11	$ERKPP MEKPP \rightarrow ERKPP+MEKPP$	761	No

Fig. 6 Comparison of species behaviours between one best learned model and the target RKIP pathway **a** Raf1, **b** RKIP, **c** RKIPP, **d** ERK, **e** RP, **f** MEKPP, **g** RKIP|Raf1, **h** ERK|MEKPP



experiments are more important than the exact numerical values of species concentrations while performing experiments. Thus, Fig. 6 shows that the behaviours of the learned model qualitatively agrees those demonstrated by the target pathway. This indicates that our proposed model

learning framework can effectively optimise the kinetic rates associated with biochemical reactions, which in turn drive the species behaviours of the synthetic models to approach the target species behaviours in an approximate manner.

In future study, when performing quantitative optimisations of associated kinetic rate values on a large scale, dissimilarities of these molecular species behaviours could be further reduced by increasing the number of obtained reactants and interactions between these reactants. Another way of driving the shapes of molecular species behaviours is to generate appropriate interactions between target molecular species and hidden complex in the evolved models at optimised kinetic rate values.

Comparison with Relevant Systems

There are a number of biochemical modelling frameworks which consider both qualitative and quantitative perspectives.

For instance, Guerriero has used the gp130/JAK/STAT signalling pathway as a case study for modelling and performs analysis using the Bio-PEPA process algebra in [23]. In that work, the PRISM model-checker [28] was used to verify a number of qualitative properties of a model generated by the Bio-PEPA Workbench. The qualitative properties of a PRISM model include deadlock states, species invariants, reachability, reversibility and liveness. These properties are intended to be consistency checks on the model, which allow the presence of possible human errors to be checked in the modelling process and the model can be guaranteed to behave as expected. Then, the PRISM model-checker was used to perform quantitative analysis: state reward-based properties were considered to observe the time series for some of the quantitative species of the system; additional (semi-)quantitative measures were also defined for further quantitative computation.

Milazzo [42] developed a formalism for the description of biological systems, called Calculus of Looping Sequences (CLS), based on term rewriting. Some typical features of process calculi for concurrency were included in CLS. Formalisms were introduced for the description of biological systems and formal tools were provided for the verification of properties of biological systems. Qualitative aspects of biological systems (structure and presence/absence of certain molecules) were also considered, concluding that it is only possible to verify qualitative properties such as the reachability of particular states or causality relationships between biological events. Moreover, a stochastic extension of CLS (Stochastic CLS) was developed, in which quantitative aspects of biological systems such as time and probabilities were taken into account.

However, the above methods did not suggest how to apply both quantitative and qualitative methods and their integration to optimise the candidate models in terms of their structure and kinetic rates. Our learning framework focuses on the optimisation of structure and

kinetic rates within the same framework, and we employ Petri nets to represent models and QML to infer and verify qualitative models. These features make our approach distinguish from the above-mentioned approaches.

Finally, it is worthwhile to investigate the application of other evolutionary algorithms and soft computing techniques in our integrative learning framework. Therefore, it will be interesting to compare the performance of ES and SA with that of other evolutionary algorithms and soft computing techniques when they are applied to our learning framework. In future research, we plan to employ classical bio-inspired algorithms, for instance, the genetic algorithm (GA) and ant colony optimisation (ACO) algorithm, in our learning framework to investigate whether these algorithms are suitable for our learning tasks or could improve the overall learning performance.

Conclusions

The lack of biochemical knowledge and the limitation of experimental techniques are reasons that studying biochemical systems in wet laboratory is a time-consuming and expensive task. Modelling biochemical systems *in silico* helps the investigation of biochemical systems in nature. Thus, it is important and interesting for life scientists to use alternative routes to study biochemical systems. Existing modelling approaches are either qualitative when there are only sparse, noisy data and incomplete knowledge available, or quantitative when there are sufficient reliable quantitative data. However, in many circumstances, the model structure for the underlying system and associated reactants may not be well identified due to the problem nature, the lack of data and knowledge, and technical limitations. Therefore, there is a need to perform model identification at both qualitative and quantitative levels, that is, qualitatively explore the model structure space (the topology of a model) and quantitatively optimise the parameters of the target biochemical systems in an integrative manner.

In this research, we show how the identification of biochemical systems can be performed and evolved in an integrative manner by reusing, composing, and evolving biochemical modules qualitatively and by mutating kinetic rates quantitatively. The main issues of the integrative qualitative and quantitative model learning are addressed in this research: firstly, interactions between reactants (including potential reactants to be discovered) are learned by the qualitative model learning approach with an evolutionary algorithm; secondly, kinetic rates in a generated biochemical model are quantitatively optimised so that behaviours of the target biochemical systems are

reproduced in synthetic models. Experimental results have shown that our proposed integrative qualitative and quantitative model learning framework is feasible and effective in the presence of incomplete knowledge and qualitative data. We point out that our model learning framework can be applied in both the context of computational systems biology for biochemical system identification and synthetic biology for the modular design of desired biochemical systems.

In future research, a high-performance computing environment will be used to improve the learning performance of our framework. By performing parallelised construction of biochemical models, for instance, models for metabolic or signalling pathways participated by some key species, much more biochemical meaningful models can be obtained and further investigated by biologists in wet laboratory. Furthermore, after generating a large number of biochemical models, some biochemical interaction patterns can be statistically analysed and summarised for later reuse as modules for piece-wise model composing [68]. Another interesting research direction is to investigate the effectiveness of more established and state-of-the-art heuristic algorithms, especially evolutionary algorithms, when they are applied to our integrative qualitative and quantitative system identification tasks. Finally, we expect that our learning framework can be implemented as a user interactive program for biologists so that they can make use of the program to facilitate experiments and generate feedbacks.

Acknowledgments The authors would like to thank the support on this research by the CRISP Project (Combinatorial Responses In Stress Pathways) funded by the BBSRC (BB/F00513X/1) under the Systems Approaches to Biological Research (SABR) Initiative.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Akman OE, Watterson S, Parton A, Binns N, Millar AJ, Ghazal P. Digital clocks: simple boolean models can quantitatively describe circadian systems. *J R Soc Interface*. 2012;9(74):2365–82.
2. Anily S, Federgruen A. Simulated annealing methods with general acceptance probabilities. *J Appl Probab*. 1987;24(3):657–67.
3. Baldan P, Cocco N, Marin A, Simeoni M. Petri nets for modelling metabolic pathways: a survey. *Nat Comput*. 2010;9(4):955–89.
4. Baldan P, Cocco N, Marin A, Simeoni M. Petri nets for modelling metabolic pathways: a survey. *Nat Comput*. 2010;9:955–89.
5. Beyer H, Schwefel H. Evolution strategies: a comprehensive introduction. *Nat Comput*. 2002;1(1):3–52.
6. Breitling R, Gilbert D, Heiner M, Orton R. A structured approach for the engineering of biochemical network models, illustrated for signalling pathways. *Brief Bioinform*. 2008;9(5):404–21.
7. Brim L, Češka M, Dražan S, Šafránek D. Exploring parameter space of stochastic biochemical systems using quantitative model checking. In: Sharygina N, Veith H, editors. *Computer aided verification, of lecture notes in computer science*, vol. 8044. Berlin: Springer; 2013. p. 107–23.
8. Bruce AM, Coghil GM. Parallel fuzzy qualitative reasoning. In: Rinner B, Hofbaur M, Wotowa F, editors. *19th International Workshop on Qualitative Reasoning*. 2005. p. 110–116.
9. Calder M, Gilmore S, Hillston J. Modelling the influence of RKIP on the ERK signalling pathway using the stochastic process algebra PEPA. In: *Transactions on computational systems biology*, vol. 4230. New York: Springer; 2004. p. 1–23.
10. Cem AC. Say and S. Kuru. Qualitative system identification: deriving structure from behavior. *Artificial Intell*. 1996;83(1): 75–141.
11. Chaouiya C. Petri net modelling of biological networks. *Brief Bioinform*. 2007;8(4):210–9.
12. Chelliah V, Juty N, Ajmera I, Ali R, Dumousseau M, Glont M, Hucka M, Jalowicki G, Keating S, Knight-Schrijver V, Lloret-Villas A, Natarajan KN, Pettit JB, Rodriguez N, Schubert M, Wimalaratne SM, Zhao Y, Hermjakob H, Le Novre N, Laibe C. *Biomodels: ten-year anniversary*. *Nucleic Acids Res*. 2014; 43(D1):D542–8.
13. Cho KH, Shin SY, Kim HW, Wolkenhauer O, McFerran B, Kolch W. Mathematical modeling of the influence of RKIP on the ERK signaling pathway. In: Priami C, editor. *Computational methods in systems biology (CSMB'03) of LNCS*, vol. 2602. New York: Springer; 2003. p. 127–41.
14. Coghil GM, Chantler MJ. Mycroft: a framework for qualitative reasoning. In: *Second international conference on intelligent systems engineering*. 1994. p. 43–48.
15. Coghil GM, Garrett SM, King RD. Learning qualitative models in the presence of noise. In: *Proceedings of 16th International Workshop on Qualitative Reasoning, QR'02*. 2002. p. 27–36.
16. Coghil GM. Mycroft: a framework for constraint based fuzzy qualitative reasoning. Phd thesis, Heriot-Watt University, Edinburgh, 1996.
17. Coghil GM, Srinivasan A, King RD. Qualitative system identification from imperfect data. *J Artificial Intell Res*. 2008; 32(1):825–77.
18. Elliot WH, Elliot DC. *Biochemistry and molecular biology*. 2nd ed. Oxford: Oxford University Press; 2002.
19. Feng H, Shao C, Wang Z. Using qualitative description for the dynamic analysis of virus infection. In: Wang T, Li X, Chen S, Wang X, Abbas H, Iba H, Chen G, Yao X, editors. *Simulated evolution and learning, of lecture notes in computer science*, vol. 4247. Berlin: Springer; 2006. p. 640–7.
20. Forbus KD. Qualitative reasoning. *CRC Handbook of Computer Science and Engineering*. 1996. p. 715–733.
21. Gilbert D, Heiner M. From petri nets to differential equations—an integrative approach for biochemical network analysis. In: Donatelli S, Thiagarajan PS, editors. *Proceedings of the 27th international conference on applications and theory of petri nets and other models of concurrency—ICATPN 2006, of lecture notes in computer science*, vol. 4024. Berlin: Springer; 2006. p. 181–200.
22. Gormley P, Li K, Wolkenhauer O, Irwin GW, Du D. Reverse engineering of biochemical reaction networks using co-evolution with eng-genes. *Cognit Comput*. 2013;5(1):106–18.
23. Guerriero ML. Qualitative and quantitative analysis of a bio-pepa model of the gp130/jak/stat signalling pathway. *Transactions on Computational Systems Biology XI*. 2009. p. 90–115.
24. Guldberg CM, Waage P. Studies concerning affinity. C. M. Forhandlinger: Videnskabs-Selskabet i Christiania, 35 1864.
25. Habibi I, Emamian ES, Abdi A. Quantitative analysis of intracellular communication and signaling errors in signaling networks. *BMC Syst Biol*. 2014;8:89.

26. Hau DT, Coiera EW. Learning qualitative models of dynamic systems. *Mach Learn*. 1993;26:177–211.
27. Heiner M, Herajy M, Liu F, Rohr C, Schwarick M. Snoopy—a unifying petri net tool. In: Haddad S, Pomello L, editors. *Application and theory of petri nets*, of lecture notes in computer science, vol. 7347. Berlin: Springer; 2012. p. 398–407.
28. Hinton A, Kwiatkowska M, Norman G, Parker D. Prism: a tool for automatic verification of probabilistic systems. In: Hermanns H, Palsberg J, editors. *Tools and algorithms for the construction and analysis of systems*, of lecture notes in computer science, vol. 3920. Berlin: Springer; 2006. p. 441–4.
29. Kholodenko BN, Kiyatkin A, Bruggeman FJ, Sontag E, Westerhoff HV, Hoek JB. Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. *Proc Natl Acad Sci*. 2002;99(20):12841–6.
30. King RD, Garrett SM, Coghill GM. On the use of qualitative reasoning to simulate and identify metabolic pathways. *Bioinformatics*. 2005;21(9):2017–26.
31. Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. *Science*. 1983;220(4598):671–80.
32. Kremling A, Fischer S, Gadkar K, Doyle FJ, Sauter T, Bullinger E, Allgöwer F, Gilles ED. A benchmark for methods in reverse engineering and model discrimination: problem formulation and solutions. *Genome Res*. 2004;14(9):1773–85.
33. Kuipers B. Qualitative simulation. *Artificial Intell*. 1986;29:289–338.
34. Kuipers B. Qualitative simulation: then and now. *Artificial Intell*. 1993;59(1–2):133–40.
35. Kuipers B. *Qualitative reasoning: modeling and simulation with incomplete knowledge*. Cambridge: The MIT Press; 1994.
36. Lee D, Yun C, Cho A, Hou BK, Park S, Lee SY. Webcell: a web-based environment for kinetic modeling and dynamic simulation of cellular networks. *Bioinformatics*. 2006;22(9):1150–1.
37. Ljung L. *System identification: theory for the user*. New York: Pearson Education; 1998.
38. Marwan W, Wagler A, Weismantel R. Petri nets as a framework for the reconstruction and analysis of signal transduction pathways and regulatory networks. *Nat Comput*. 2011;10(2):639–54.
39. Matsuno H, Li C, Miyano S. Petri net based descriptions for systematic understanding of biological pathways. *IEICE Trans Fundam Electron Commun Comput Sci website*. 2006;E89–A:3166–74.
40. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *J Chem Phys*. 1953;21(6):1087–92.
41. Michaelis L, Menten ML. Die kinetik der invertinwirkung. *Biochem Z*. 1913;49:333–69.
42. Milazzo P. Qualitative and quantitative formal modeling of biological systems. PhD thesis. April 2007.
43. Moles CG, Mendes P, Banga JR. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res*. 2003;11:2467–74.
44. Murata T. Petri nets: properties, analysis and applications. *Proc IEEE*. 1989;77(4):541–80.
45. Napione L, Manini D, Cordero F, Horvth A, Picco A, De Pierro M, Pavan S, Sereno M, Veglio A, Bussolino F, Balbo G. On the use of stochastic petri nets in the analysis of signal transduction pathways for angiogenesis process. In: Degano P, Gorrieri R, editors. *Computational methods in systems biology*, of lecture notes in computer science, vol. 5688. Berlin: Springer; 2009. p. 281–95.
46. Nobakhti A. On natural based optimization. *Cognit Comput*. 2010;2(2):97–119.
47. Pang W. QML-morven: a framework for learning qualitative models. Aberdeen: Aberdeen University; 2009.
48. Pang W, Coghill GM. Learning qualitative differential equation models: a survey of algorithms and applications. *Knowl Eng Rev*. 2010;25:69–107.
49. Pang W, Coghill GM. An immune-inspired approach to qualitative system identification of biological pathways. *Nat Comput*. 2011;10(1):189–207.
50. Pang W, Coghill GM. Qml-morven: a novel framework for learning qualitative differential equation models using both symbolic and evolutionary approaches. *J Comput Sci*. 2014;5(5):795–808.
51. Pang W, Coghill GM. Qml-ainet: an immune network approach to learning qualitative differential equation models. *Appl Soft Comput*. 2015;27:148–57.
52. Pang W, Coghill GM. An immune network approach to learning qualitative models of biological pathways. In: *IEEE congress on evolutionary computation (CEC)*, 2014. July 2014. p. 1030–1037.
53. Pang W, Coghill GM. Learning qualitative metabolic models using evolutionary methods. In: *Fifth international conference on frontier of computer science and technology (FCST)*, 2010. Aug 2010. p. 436–441.
54. Pareto V. *Manuale deconomia politica*. 1906. English translation, A.M. Kelly 1971.
55. Peil J. Quantitative description and mathematical modelling of biological systems by means of nonlinear approximations demonstrated at the example of the body length growth of man. *Biom J*. 1979;21(1):57–70.
56. Peleg M, Rubin D, Altman RB. Using petri net tools to study properties and dynamics of biological systems. *J Am Med Inform Assoc*. 2005;12(2):181–99.
57. Qin T, Tsoi L, Sims K, Lu X, Zheng W. Signaling network prediction by the ontology fingerprint enhanced bayesian network. *BMC Syst Biol*. 2012;6(Suppl 3):S3.
58. Reddy VN, Mavrovouniotis ML, Liebman MN. Petri net representations in metabolic pathways. In: *Proceedings of the 1st international conference on intelligent systems for molecular biology*. 1993. AAAI Press. p. 328–336.
59. Reddy VN, Liebman MN, Mavrovouniotis ML. Qualitative analysis of biochemical reaction systems. *Comput Biol Med*. 1996;26(1):9–24.
60. Richards BL, Kraan I, Kuipers BJ. Automatic abduction of qualitative models. In: *Proceedings of the tenth national conference on Artificial intelligence, AAAI'92*. 1992. AAAI Press. p. 723–728.
61. Samaga R, Klamt S. Modeling approaches for qualitative and semi-quantitative analysis of cellular signaling networks. *Cell Commun Signal*. 2013;11(1):43.
62. Schwefel H. *Kybernetische Evolution als Strategie der experimentellen Forschung in der Strömungstechnik*. : Diplomarbeit, Technische Universität Berlin, Hermann Föttinger-Institut für Strömungstechnik, März; 1965.
63. Steggle LJ, Banks R, Shaw O, Wipat A. Qualitatively modelling and analysing genetic regulatory networks: a petri net approach. *Bioinformatics*. 2007;23:2006.
64. Villaverde AF, Banga JR. Reverse engineering and identification in systems biology: strategies, perspectives and challenges. *J R Soc Interface*. 2013;11(91):20130505.
65. Vlad MO, Arkin A, Ross J. Response experiments for nonlinear systems with application to reaction kinetics and genetics. *Proc Natl Acad Sci*. 2004;101(19):7223–8.
66. Wu Z, Yang S, Gilbert D. A hybrid approach to piecewise modelling of biochemical systems. In: Coello CAC, Cutello V, Deb K, Forrest S, Nicosia G, Pavone M, editors. *Parallel problem solving from nature—PPSN XII*, of lecture notes in computer science, vol. 7491. Berlin: Springer; 2012. p. 519–28.
67. Wu Z, Grosan C, Gilbert D. Empirical study of computational intelligence strategies for biochemical systems modelling. In: Terrazas G, Otero FEB, Masegosa AD, editors. *Nature inspired cooperative strategies for optimization (NICSO 2013)*, of studies in computational intelligence, vol. 512. New York: Springer International Publishing; 2014. p. 245–60.

68. Wu Z. A generic approach to behaviour-driven biochemical model construction. PhD thesis, London/Uxbridge: Brunel University, October 2012.
69. Wu Z, Gao Q, Gilbert D. Target driven biochemical network reconstruction based on petri nets and simulated annealing. In: Proceedings of the 8th international conference on computational methods in systems biology, CMSB '10. 2010. New York: ACM. p. 33–42.
70. Wu Z, Pang W, Coghill GM. An integrative top-down and bottom-up qualitative model construction framework for exploration of biochemical systems. *Soft Comput.* 2014. p. 1–16.
71. Wu Z, Pang W, Coghill GM. Stepwise modelling of biochemical pathways based on qualitative model learning. In: UKCI. 2013. p. 31–37.
72. Yeung K, Janosch P, McFerran B, Rose DW, Mischak H, Sedivy JM, Kolch W. Mechanism of suppression of the Raf/MEK/Extracellular signal-regulated kinase pathway by the Raf kinase inhibitor protein. *Mol Cell Biol.* 2000;20(9):3079–85.