On the Implementation of Quantitative Model Refinement

Bogdan Iancu[1,2,3], Diana-Elena Gratie[1,2,3], Sepinoud Azimi[1,2,3], and Ion Petre[1,2,3]
{biancu,dgratie,sazimi,ipetre}@abo.fi

[1] Computational Biomodeling Laboratory
[2] Turku Centre for Computer Science
[3] Department of IT, Åbo Akademi University
Joukahainengatan 3-5, FIN-20520 Åbo

Abstract. The iterative process of adding details to a model while preserving its numerical behavior is called *quantitative model refinement*, and it has been previously discussed for ODE-based models and for *kappa*-based models. In this paper, we investigate and compare this approach in three different modeling frameworks: rule-based modeling, Petri nets and guarded command languages. As case study we use a model for the eukaryotic heat shock response that we refine to include the acetylation of the heat shock factor. We discuss how to perform the refinement in each of these frameworks in order to avoid the combinatorial state explosion of the refined model. We conclude that Bionetgen (and rule-based modeling in general) is well-suited for a compact representation of the refined model, Petri nets offer a good solution through the use of colors, while the PRISM refined model may be much larger than the basic model.

Keywords: Quantitative model refinement, heat shock response, acetylation, rule-based modeling, Petri nets, model checking.

1 Introduction

Systems biology aims to holistically characterize highly complex biological systems. A hierarchical system-level representation is very adequate in this context. Formal frameworks turn out to be fundamental in the effort of understanding the behavior of such complex systems, see [12, 21]. The abstractions that lie at the core of these formalisms need to be refined to incorporate more details.

We focus in this paper on the implementation of model refinement. Within the model development process, we examine *data refinement* through three different frameworks – *rule-based modeling, Petri nets* and *guarded command languages* – and discuss their capabilities for the efficient construction of a refined model. For rule-based modeling we used the Bionetgen framework and RuleBender, for Petri nets we chose Snoopy and Charlie as modeling tools, while for modeling with guarded command languages we used PRISM. Data refinement, as described in [3] and [10], assumes the replacement of one species in the model with several of its variants, called subspecies. This type of refinement is adequate for representing post-translational modifications of proteins, e.g., acetylation, phosphorylation, etc. Given a protein P, one can indicate its state regarding

post-translational modifications by replacing it with its variants. This substitution also implies a refinement of all complexes involving protein P and of all reactions involving either P or any such complex, see [10]. This might induce a combinatorial state explosion of the refined model, as in the case of ODE-based models, see [10]. The main question we are answering is whether one can avoid this problem in the other three frameworks we investigate in this paper and build a compact representation of the refined model.

We consider as a case study for our analysis the heat shock response mechanism, as described in [20] and [10]. Throughout the paper, the model in [20] will be referred to as the *basic* heat shock response model, while the model in [10] will be referred to as the *refined model*.

All models developed in this paper are available for download at [11]. Due to space restrictions, some of the details of this work were omitted. For full details, we refer the reader to [6].

2 The Heat Shock Response (HSR)

The eukaryotic *heat shock response* is a highly conserved bio-regulatory network that controls cellular function impairment produced by protein misfolding as a result of high temperatures. Elevated temperatures have proteotoxic effects on proteins, inducing protein misfolding and leading to the formation of large aggregates that thereafter trigger apoptosis (controlled cell death). Cell survival is promoted by a defense mechanism, which consists in restoring protein homeostasis by augmenting the level of molecular chaperones, see [22].

We consider the basic molecular model for the eukaryotic heat shock response proposed in [20]. *Heat shock proteins* (hsp's) play a key role in the heat shock response mechanism by chaperoning the *misfolded proteins* (mfp's). Due to their affinity to mfp's, hsp's form hsp:mfp complexes and help the misfolded proteins refold. The heat shock response is regulated by the transactivation of the hsp-encoding genes. In eukaryotes, some specific proteins, called *heat shock factors* (hsf's), promote gene transcription. In the absence of environmental stressors, heat shock factors are predominantly found in a monomeric state, extensively bound to heat shock proteins. Raising the temperature causes the correctly folded proteins (prot) to misfold and hsp:hsf complexes to break down. This switches on the heat shock response by releasing hsf's, which quickly reach a DNA binding competent state, see [20, 23].

Heat stress induces dimerization (hsf_2) and, subsequently, trimerization (hsf_3) of hsf's, enabling the binding of the hsf trimers to the promoter site of the hsp-encoding gene, called *heat shock element* (hse). Subsequently, DNA binding triggers the transcription and translation of the hsp-encoding gene, inducing hsp synthesis, see [20, 22]. Once the level of heat shock proteins is sufficiently elevated for the cell to withstand thermal stress, hsp synthesis is turned off. Heat shock proteins sequestrate heat shock factors and break hsf dimers and trimers, constituting hsp:hsf complexes. The explicit molecular reactions constituting the model can be found in [20].

The numerical setup of the basic model (in terms of initial concentrations and kinetic constants) can be found in [20]. Acetylation has been shown to have an extensive

influence in regulating the heat shock response, we refer the reader to [24]. To this end, we consider the acetylation of heat shock factors implemented through data refinement.

3 Quantitative Model Refinement

Quantitative model refinement was investigated in [4, 19] regarding rule based modeling and applied to two ampler ODE-based models in [10, 18].

3.1 Quantitative Model Refinement

A reaction-based model can be refined to incorporate more information regarding its reactants and/or reactions. There are two types of refinement, either of the *data* (data refinement) or of the *reactions* (process refinement). In this study, we focus on the first refinement type. Considering that one's interest lies especially on data, a species in a model could be refined by replacing it with several of its subspecies, a routine called *data refinement*. When the interest is focused on reactions, the model can be refined by replacing a collective reaction, accounting for a specific process, by a set of reactions depicting the transitional steps of the process. The last type of refinement is called *process refinement*, see [10].

The notion of quantitative model refinement has been previously addressed in systems biology in the context of rule based modelling, see [4, 5, 7, 19]. The rule based modelling framework embodies the concept of *data refinement*, as previously introduced, implementing agent resolution as a fundamental constituent, [7]. The key refinement method in this context is rule refinement, an approach that requires the refinement of the set of rules ensuring the preservation of the dynamic behavior of the system, see [19].

We present here the *quantitative model refinement* of reaction models following the discussion in [10]. Consider a model M, comprising a number m of species $\Sigma = \{A_1, A_2, \ldots, A_m\}$ and n of reactions r_i, $1 \leq i \leq n$, as follows:

$$r_i \ : \ S_{i,1}A_1 + S_{i,2}A_2 + \ldots + S_{i,m}A_m \xrightarrow{k_i} S'_{i,1}A_1 + S'_{i,2}A_2 + \ldots + S'_{i,m}A_m,$$

where $S_{i,1}, \ldots, S_{i,m}, S'_{i,1}, \ldots, S'_{i,m} \geq 0$ are the stoichiometric coefficients of r_i and $k_i \geq 0$ is the kinetic rate constant of r_i. We discuss here a continuous, mass-action formulation of the model based on ODEs. For some details on this approach we refer to [14].

Model M can be refined to distinguish between various subspecies of any species in the model, for example, A_1. The distinction between the subspecies is very often drawn by post-translational modifications such as acetylation, phosphorylation, sumoylation, etc. All previously mentioned subspecies of A_1 take part in all reactions A_1 engaged in, conceivably obeying a different kinetic setup. Given model M and species A_1, substituting subspecies $B_1, \ldots, B_l$ for species A_1 in M leads to attaining a new model M_R, comprising species $\{A'_2, A'_3, \ldots, A'_m\} \cup \{B_1, \ldots, B_l\}$, for some $l \geq 2$, where variables A'_i, $2 \leq i \leq m$ from M_R, coincide with A_i from model M and $B_1, \ldots, B_l$ substitute for species A_1 in M_R. Furthermore, each reaction r_i of M is replaced in the new model M_R by all possible reactions $r_{i,j}$ of the following form:

$$r_{i,j} \;:\; (T^j_{i,1}B_1 + \ldots + T^j_{i,l}B_l) + S_{i,2}A'_2 + \ldots + S_{i,m}A'_m \xrightarrow{k_{i,j}} (T'^j_{i,1}B_1 + \ldots + T'^j_{i,l}B_l) + S'_{i,2}A'_2 + \ldots + S'_{i,m}A'_m,$$

where $k_{i,j}$ is the kinetic rate constant of $r_{i,j}$ and $(T^j_{i,1}, \ldots, T^j_{i,l}, T'^j_{i,1}, \ldots, T'^j_{i,l})$ are all possible nonnegative integers so that $T^j_{i,1} + \ldots + T^j_{i,l} = S_{i,1}$ and $T'^j_{i,1} + \ldots + T'^j_{i,l} = S'_{i,1}$. Model M_R is said to be a *data refinement of model M on variable* A_1 if and only if the following conditions are fulfilled:

$$[A_i](t) = [A'_i](t), \tag{1}$$

$$[A_1](t) = [B_1](t) + \ldots + [B_l](t), \tag{2}$$

for all $2 \leq i \leq m$, $t \geq 0$. Fulfilling these conditions depends on the numerical setup of model M_R, i.e., on the kinetic constants of its reactions (both those adopted from the basic model, as well as those newly introduced in the construction) and on the initial concentrations of its species.

3.2 Adding the Acetylation Details to the HSR Model through Data Refinement

We start from the basic model of the heat shock response, introduced in [20], where no post-translational modification of hsf is taken into account, and we refine all species and complexes that involve hsf taking into consideration one acetylation site for every hsf molecule. We follow here the discussion in [10]. The aim is to refine the basic model and preserve its numerical properties. For hsf_2, hsf_3, $\mathsf{hsf}_3\colon\mathsf{hse}$ and $\mathsf{hsp}\colon\mathsf{hsf}$, the refinement is performed conforming to the number of hsf constituents respectively. This leads to the following data refinements: $\mathsf{hsf} \to \{\mathsf{rhsf}, \mathsf{rhsf}^{(1)}\}$; $\mathsf{hsf}_2 \to \{\mathsf{rhsf}_2, {\mathsf{rhsf}_2}^{(1)}, {\mathsf{rhsf}_2}^{(2)}\}$; $\mathsf{hsp}\colon\mathsf{hsf} \to \{\mathsf{rhsp}\colon\mathsf{rhsf}, \mathsf{rhsp}\colon\mathsf{rhsf}^{(1)}\}$; $\mathsf{hsf}_3 \to \{\mathsf{rhsf}_3, {\mathsf{rhsf}_3}^{(1)}, {\mathsf{rhsf}_3}^{(2)}, {\mathsf{rhsf}_3}^{(3)}\}$; $\mathsf{hsf}_3\colon\mathsf{hse} \to \{\mathsf{rhsf}_3\colon\mathsf{rhse}, {\mathsf{rhsf}_3}^{(1)}\colon\mathsf{rhse}, {\mathsf{rhsf}_3}^{(2)}\colon\mathsf{rhse}, {\mathsf{rhsf}_3}^{(3)}\colon\mathsf{rhse}\}$. The refinement based on the above data refinements involves substantial changes in the list of reactions. For example, the reversible reaction of dimerization $2\,\mathsf{hsf} \rightleftarrows \mathsf{hsf}_2$ in the basic model is replaced by three reactions as follows: $2\,\mathsf{rhsf} \rightleftarrows \mathsf{rhsf}_2$; $\mathsf{rhsf} + \mathsf{rhsf}^{(1)} \rightleftarrows {\mathsf{rhsf}_2}^{(1)}$; $2\mathsf{rhsf}^{(1)} \rightleftarrows {\mathsf{rhsf}_2}^{(2)}$.

The refined model of [10] consists of 20 species and 55 irreversible reactions, compared to 10 species and 17 irreversible reactions in the basic model of [20].

4 Quantitative Refinement in Rule-Based Models

4.1 A RuleBender Implementation of the Basic HSR Model

This section focuses on the RuleBender implementation of the basic heat shock response model, as introduced in Section 2. We model all reactions to follow the principle of mass action. Conforming to the implementation presented here, Bionetgen source code comprises a set of twelve rules, which generate a total number of seventeen irreversible reactions. Due to the symmetry that some of the species exhibit, the collision frequency (e.g. in our case dimerization, trimerization, etc) and the existence of multiple paths from substrates to products in some reactions (e.g. for the heat shock response model, the

unbinding of trimers), kinetic rate constants for those specific reactions are multiplied in Bionetgen by diverse symmetry and/or statistical factors, see [2]. For example, the collision frequency of two different types of reactants A and B, $A+B$, is twice that of identical types of reactants $A+A$. Another example concerns the multiple reaction paths from reactants to products, which may generate statistical factors. Preserving the fit of the heat shock response model attained in [20] required a multiplication of some rate constants by the inverse of the aforementioned factors respectively.

RuleBender generates during the process of model development a contact map which depicts the connectivity between the molecules. The contact map for the basic model of the heat shock response is shown in Figure 1.

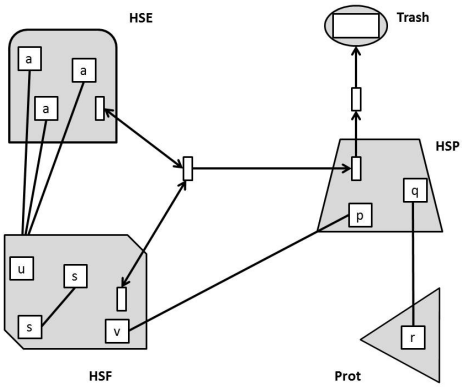

Fig. 1. The RuleBender generated contact map for the basic model of the heat shock response. It depicts the possible interconnections among the model's species.

One can notice in Figure 1 that hsf's have been represented as having 4 sites (s, s, u, v). The two s sites are involved in the generation of dimers and trimers. The other two sites, u and v, are used to illustrate the process of DNA binding/unbinding and hsf sequestratation/dimer (trimer) dissipation. Trimers are considered to be circular structures, each of the 's' site of one hsf being bound to the 's' sites of the consequent hsf's, no two hsf's having both sites 's' bound to the same partner. The promoter, hse, has been represented as having three identical sites (a, a, a), so as to be connected to the trimer in such a way that the symmetry is not affected. Heat shock proteins are modeled to have two sites 'p' and 'q', used for the modelling of unbinding of dimers and trimers and for the sequestration of misfolded proteins. The model takes into account a species called Prot, which has a site with two possible states, one of which accounts for misfolded proteins 'm' and another one 'f', that accounts for folded proteins. A "dummy" component, called Trash, has been introduced to help encode the degradation of heat shock proteins.

The contact map in Figure 1 illustrates the connectivity between the species in the model. The link between the 's' sites of the hsf molecule denotes the formation of dimers and trimers through the agency of these sites. Once trimers are formed, they can bind to

the heat shock element (hse), the connection being illustrated by three links connecting hsf trimers to the heat shock element (one can notice three 'a' sites the heat shock element component exhibits). The middle connector encodes for a number of reactions, such as: DNA unbinding, HSP synthesis and breaking of dimers and trimers. The link between the site 'v' of the hsf component and the site 'p' of the hsp component illustrates hsf sequestration. The link between the hsp component and the prot component encodes the following reactions: protein misfolding, protein refolding and mfp sequestration. By linking the component *Trash* to the hsp component, we encoded for the degradation of hsp's.

We chose a deterministic simulation for the basic model. The simulation results for DNA binding for a temperature of 42°C showed that RuleBender prediction are in accordance with the results reported in [20].

4.2 A RuleBender Implementation of the Acetylation-Refined HSR Model

We focus in this section on the acetylation-refinement of the heat shock response, as described in [10]. There are several changes to do in Rulebender to refine the basic model so as to include the acetylation of hsf's. The syntax of the rules remains, in this case, unchanged, since all reactions, in this model, take place regardless of the acetylation status of the molecules. We brought changes in the definition of hsf's, by introducing one acetylation site, 'w', which can be either acetylated or not, and in the initial concentrations of the molecules. The initial concentrations were set conforming to [10].

As expected from the refinement conditions, the simulation of the refined model for a temperature of 42°C showed that the Rulebender prediction for the refined model and the one for the basic model are the same.

5 Quantitative Refinement in Petri Net Models

5.1 A Petri Net for the Basic HSR Model

A standard Petri net model for the heat shock response was previously reported in [1]. We focus here on a Snoopy continuous Petri net implementation of the basic heat shock response model, shown in Figure 2. The network has 10 places and 17 transitions, encoding the 10 species and 17 irreversible reactions in the basic model definition of [20]. Verifying the model required the analysis of several properties. For instance, the model is covered by T-invariants; also, the P-invariants reported by Charlie encode all mass conservation relations reported in the ODE-based model of [20]. Moreover, all places except HSP are covered by P-invariants, which means that they are bounded. The three mass conservation relations yield three constants (accounting for the total amount of HSF, HSE and protein molecules in the system, respectively), that have been used in the PRISM implementation of the model.

5.2 Petri Nets for the Acetylation-Refined HSR Model

For the refined heat shock response that includes two types of hsf's (acetylated and non-acetylated [10]), we chose an implementation based on colored continuous Petri nets.

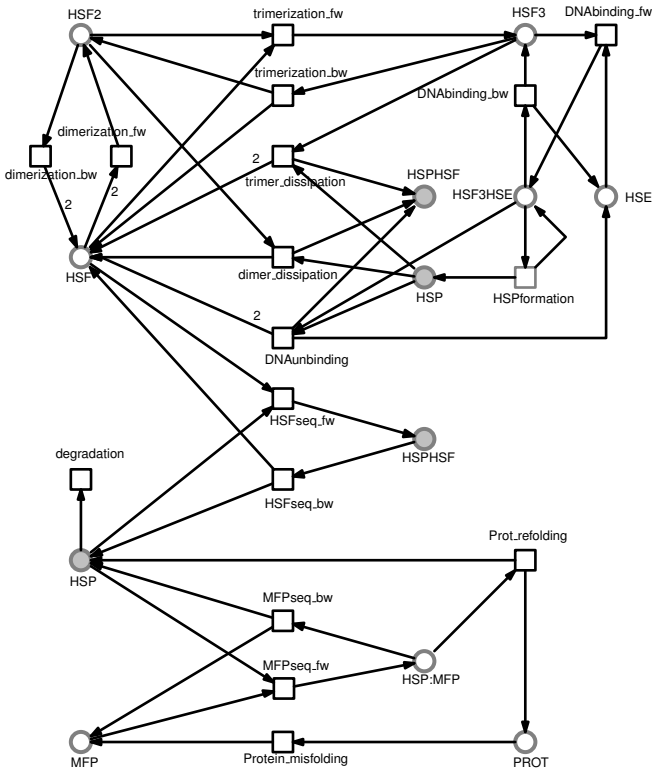

Fig. 2. Snoopy implementation of the basic heat shock response model. The text next to a place (transition) denotes the identifier of that particular place (transition). Arc multiplicities greater than 1 are included in the picture. The dashed gray circles are logical places (they may appear several times, but they represent the same species).

There are several ways of reasoning about refined species within this framework. For example, the dimer of a protein with a site that can be acetylated (1) or non-acetylated (0) can be either seen as an entity with 0, 1, or 2 acetylated sites, or as a compound where the order of the acetylated sites counts (i.e. (0,0), (0,1), (1,0), (1,1)). Depending on the approach one takes, the colored representation will have different color sets, different number of transitions and different kinetic constants.

We modeled the refined heat shock response using two approaches: one focused on keeping the structure of the basic model intact, with the same transitions and kinetic constants (we call this model *transition-focused*). This is the most compact representation. The other approach aimed to minimize the number of colors used in the model (we call this model *color-focused*). This approach uses as few colors as possible, at the cost of a complicated representation, with many conditions in a transition, and also introducing new transitions in the colored representation. Due to space limitations, we present here only the color-focused model. A more detailed description of both approaches can be found in [6].

Several choices had to be made during the modeling process. We detail the modeling options for the dimerization and trimerization of acetylated and non-acetylated hsf's. There are three types of dimers that can be formed: non-acetylated ($\mathsf{hsf_2}^{(0)}$), single-acetylated ($\mathsf{hsf_2}^{(1)}$) and double-acetylated dimer ($\mathsf{hsf_2}^{(2)}$). One way of modeling the dimers is using a color set with three colors of type int (0, 1, and 2 denoting the number of acetylated sites). Another approach is using a cartesian product $\{0,1\} \times \{0,1\}$. When modeling hsf trimers, one could consider, for example, one of the following three color sets: a color set $\mathsf{Tri} = \{0,1,2,3\}$, a compound color set $\mathsf{Compound} = \{0,1\} \times \{0,1,2\}$, or a compound set $\mathsf{Trimer} = \{0,1\} \times \{0,1\} \times \{0,1\}$. For the color-focused refinement, we chose the simple integer color sets.

All reactions involving the decomposition of complexes containing hsf's required additional transitions. For example, the trimer dissipation reaction $\mathsf{hsf_3} + \mathsf{hsp} \rightarrow \mathsf{hsp}:\mathsf{hsf} + 2\,\mathsf{hsf}$ is split into three transitions. One covers the case when all hsf's in the trimer have the same acetylation value (i.e. $\mathsf{hsf_3}$ has color 0 or 3). In this case, there is no distinction between which hsf binds to hsp and which two hsf's become unbound, and the kinetic constant for this transition is the same as the corresponding one in the basic model. When $\mathsf{hsf_3}$ has color 1 or 2, there are two binding possibilities: hsp binds to either a non-acetylated hsf, or to an acetylated hsf. For the two transitions representing these possibilities, the kinetic constant is half of the corresponding one in the basic model (following the reasoning explained in [10]).

When simulating a colored Petri net, Snoopy first unfolds it, in other words it creates an equivalent Petri net. Each place instance (each color) will correspond to a place in the unfolded net, and each transition instance (each binding) will correspond to a transition in the unfolded net; for details on colored Petri nets unfolding, see [17]. The color-focused refined model has 10 places and 25 transitions, and its corresponding flattened Petri net has 20 places and 56 transitions. This representation, although more complex than the transition-focused one, encodes a smaller flattened network. Both the transition- and color-based refinements have been compared with the basic model predictions, and they are all equivalent (data not shown).

6 Quantitative Refinement in PRISM Models

6.1 A PRISM Implementation of the Basic HSR Model

We implemented the basic heat shock response as a CTMC model that defines all possible guards (in this case reactions) within a single module. The PRISM model consists of 10 variables, each of them corresponding to one of the reactants in the model, and 17 guards representing the 17 irreversible reactions of the system. The values for upper bounds of the variables are taken from our Petri net model's P-invariants and mass-conservation relations. Upper bounds are used both for allocating memory and in the guarded commands. For example the guard corresponding to *dna binding* is expressed as follows: $\mathsf{hsf_3} >= 1 \wedge \mathsf{hse} >= 1 \wedge \mathsf{hsf_3}:\mathsf{hse} <= N-1 \rightarrow \mathsf{hsf_3} * \mathsf{hse} * k_5 : (\mathsf{hsf_3}' = \mathsf{hsf_3} - 1) \wedge (\mathsf{hse}' = \mathsf{hse} - 1) \wedge (\mathsf{hsf_3}:\mathsf{hse}' = \mathsf{hsf_3}:\mathsf{hse} + 1)$, where N represents the upper bound for hse in the system.

It is noteworthy to mention that the PRISM model could be obtained from the Petri net model via some format manipulations in Snoopy. However, we decided to write the

model from the very beginning in order to be able to compare the modeling effort in each chosen framework.

6.2 A PRISM Implementation of the Acetylation-Refined HSR Model

The approach we took in Sections 4 and 5 to implement the acetylation-refined heat shock response model was through a compact representation of the acetylated species. Whereas colors of the places and arc expressions were employed to represent the refinement in the Petri net model, in modeling with RuleBender the solution was to introduce a new acetylation site for every hsf molecule. Both methods used structured data types for the species, thus concealing the complexity of the model in a compact representation. In PRISM this requires a method to represent the acetylation details in the definition of hsf, i.e. a composite data type. Since PRISM currently supports only simple data type (e.g. integer, boolean) variables in the model, such a definition is not possible. Alternatively, we implemented the acetylation-refined model through introducing new variables describing all possible acetylation configurations of hsf and hsf complexes. This was similar to the ODE-based approach to quantitative model refinement discussed in [10].

The refined heat shock response model is built based on the refinements given in Section 3.2 by refining all reactants and complexes involving hsf. In this approach, the strategy is to replace each guard involving any refined reactant by the guards considering all possible refined reactions.

One could also use parallel modules to implement the refinement but this approach would not help reducing the complexity of the model.

The complete PRISM implementation of the refinement is not listed here due to space limitations. The numerical setup of this model is based on [10].

6.3 Model Checking of the HSR Models

According to [15], the maximum number of states that PRISM can handle for CTMCs is 10^{10}. In both our models (basic and refined version of the heat shock response), the number of all possible states in the system exceeds this limit. This is a known problem for biological systems in PRISM, see [8]. Several studies have addressed this issue, see e.g., [8, 9, 16]. One of the investigated approaches is *approximate verification* of probabilistic systems, where a Monte-Carlo algorithm is used to approximate the probability of a temporal formula to be true, see [9]. We used this method to verify the desired properties of the heat shock response model. In this approach a large number of stochastic paths is sampled for the model and based on the defined properties, the result for each run is obtained. The information produced in this way gives an approximate result for the probability that the desired property holds for the model.

We are interested in verifying two properties discussed in [20]. The properties are: (i) the validity of three mass-conservation relations and (ii) the level of DNA binding eventually returns to the basal values, both at $37°C$ and at $42°C$.

In order to check the mass conservation properties, we used the G operator which checks if the property remains true at all states along the path. The three properties we were interested in are listed as follows:

- $p =?$ [G $\mathsf{hsf} + 2\,\mathsf{hsf}_2 + 3\,\mathsf{hsf}_3 + 3\,\mathsf{hsf}_3{:}\,\mathsf{hse} + \mathsf{hsp}{:}\,\mathsf{hsf} = \mathsf{hsf}_{const}$],
- $p =?$ [G $\mathsf{hse} + \mathsf{hsf}_3{:}\,\mathsf{hse} = \mathsf{hse}_{const}$],
- $p =?$ [G $\mathsf{prot} + \mathsf{mfp} + \mathsf{hsp}{:}\,\mathsf{mfp} = \mathsf{prot}_{const}$],

where $\mathsf{hsf}_{const}, \mathsf{hse}_{const}$ and prot_{const} represent the total amounts of $\mathsf{hsf}, \mathsf{hse}$ and prot respectively. These properties check if the mass-conservation relations, corresponding to the level of $\mathsf{hsf}, \mathsf{hse}$ and prot, are valid in all the states. In each case, the value of p was confirmed to be one, which was to be expected, with confidence level 95%, i.e. the mass conservation laws are respected in the model.

For the second property, we verified in PRISM that for time points larger than 14400, the value of $\mathsf{hsf}_3{:}\,\mathsf{hse}$ reactants returns to their initial value. We formulated the following property: $p =? \quad [F >= 14400 \quad \mathsf{hsf}_3{:}\,\mathsf{hse} = 3]$. The probability value calculated by PRISM was one for this property as well, with confidence level 95%.

We also checked if the model confirms the experimental data of [13] on DNA binding. One approach could be to run the simulation for many times and plot the average run. Due to the memory issues of the PRISM, we were not able to follow this approach. Since we are using a stochastic model, our second approach was to check the probability of having a data point within the interval $[0.9 \cdot d, 1.1 \cdot d]$ in the time period $[0.9 \cdot t, 1.1 \cdot t]$, where d is the experimental data point at time t. The confidence interval for all the properties and the number of simulations were 95% and 150 respectively. We interpret the high values we obtained as a result as a confirmation that the two PRISM models are in accordance with the experimental data of [13].

7 Discussion

We focused in this paper on analyzing the capability of three different frameworks to implement the concept of quantitative model refinement: rule based modelling (with Bionetgen), Petri nets (with Snoopy) and guarded command languages (with PRISM). Handling the combinatorial explosion due to accounting for a post-translational modification throughout our refinement proved to be fundamentally different in the approaches we considered. These modeling methods are not restricted to the analysis of our case study solely, but their applicability extends to other reaction-based models. Rule-based modelling tackles the complexity of refinement through a compact model representation based on a partial presentation of the details of the model species, leading to more effective model construction and analysis techniques. Colored Petri nets integrate programmability by including data types (color sets) as an intrinsic property of places. The color set assignation reflects on the structure of the network, affecting the dimensions of the corresponding flattened network. PRISM model checker promotes a low level implementation of data structures and it does not allow the modeler to introduce more complex data structures.

Our study shows that some modeling frameworks are more suitable for model refinement than others, with respect to the compactness of the representation of the refined model. A key ingredient for this is the spectrum of internal data structures supported by the modeling framework. Data structures may encapsulate a large amount of information, and their effective manipulation can substantially reduce the complexity of a model's

representation. RuleBender provides data structures suitable for modeling biological systems: species, sites, links, partial description of species, rendering a straightforward refinement procedure with a very compact representation. In contrast, Petri nets are not primarily a biology-focused framework. Colored Petri nets introduce programmability in this modeling formalism, incorporating data types into the places of the network. New data types can be implemented based on primitive built-in types and composition rules. In refining a Petri net model, one has to define the appropriate data structures, and associate a biological meaning to each of them. The modeling choices affect both the compactness of the representation and the complexity of the corresponding flattened Petri net model. PRISM on the other hand only supports primitive data types. This translates into an explicit detailing of all elements of the refined model.

Our study shows that quantitative model refinement is a potentially viable approach to building a large biomodel. The approach can be used together with a multitude of modeling paradigms, allowing the modeler to increase the level of details of the model, while preserving its numerical behavior. Moreover, on any level of detail one can switch from a modeling paradigm to another, taking full advantage of the various analysis tools made possible in different model formulations, in terms of fast simulations, model checking or compact model representation. While our case-study shows the potential of the quantitative model refinement approach to model building, its scalability remains to be tested on a larger case study.

Acknowledgement

This research was partially supported by Academy of Finland under project 267915.

The authors thank Monika Heiner for her help with issues related to Snoopy and Charlie, James Faeder and Leonard Harris for advice on the Bionetgen implementation of the heat shock response, and Adam Smith for technical support regarding RuleBender.

References

1. Back, R., Ishdorj, T., Petre, I.: A petri net formalization of heat shock response model. In: Petre, I., Rozenberg, G. (eds.) Workshop on Natural Computing and Graph Transformations. pp. 19–28 (2008)
2. Blinov, M., Yang, J., Faeder, J., Hlavacek, W.: Graph theory for of biochemical networks. Transactions on Computational Systems Biology VII pp. 89–106 (2006)
3. Czeizler, E., Rogojin, V., Petre, I.: The phosphorylation of the heat shock factor as a modulator for the heat shock response. In: Proceedings of the 9th International Conference on Computational Methods in Systems Biology. pp. 9–23. ACM (2011)
4. Danos, V., Feret, J., Fontana, W., Harmer, R., Krivine, J.: Rule-based modelling and model perturbation. Transactions on Computational Systems Biology XI pp. 116–137 (2009)
5. Faeder, J., Blinov, M., Goldstein, B., Hlavacek, W.: Rule-based modeling of biochemical networks. Complexity 10(4), 22–41 (2005)
6. Gratie, D., Iancu, B., Azimi, S., Petre, I.: Quantitative model refinement in four different frameworks, with applications to the heat shock response. Tech. Rep. 1067, TUCS (2013)
7. Harmer, R.: Rule-based modelling and tunable resolution. EPTCS 9, 65–72 (2009)

8. Heath, J., Kwiatkowska, M., Norman, G., Parker, D., Tymchyshyn, O.: Probabilistic model checking of complex biological pathways. In: Computational Methods in Systems Biology. pp. 32–47. Springer (2006)
9. Hinton, A., Kwiatkowska, M., Norman, G., Parker, D.: Prism: A tool for automatic verification of probabilistic systems. Tools and Algorithms for the Construction and Analysis of Systems pp. 441–444 (2006)
10. Iancu, B., Czeizler, E., Czeizler, E., Petre, I.: Quantitative refinement of reaction models. IJUC 8(5-6), 529–550 (2012)
11. Iancu, B., Gratie, D., Azimi, S., Petre, I.: Computational modeling of the eukaryotic heat shock response: the bionetgen implementation, the petri net implementation and the prism implementation (2013), http://combio.abo.fi/research/computational-modeling-of-the-eukaryotic-heat-shock-response/
12. Kitano, H.: Systems biology: a brief overview. Science 295(5560), 1662–1664 (2002)
13. Kline, M., Morimoto, R.: Repression of the heat shock factor 1 transcriptional activation domain is modulated by constitutive phosphorylation. Molecular and cellular biology 17(4), 2107–2115 (1997)
14. Klipp, E., Herwig, R., Kowald, A., Wierling, C., Lehrach, H.: Systems biology in practice: concepts, implementation and application. Wiley-Vch (2005)
15. Kwiatkowska, M., Norman, G., Parker, D.: Quantitative analysis with the probabilistic model checker prism. Electronic Notes in Theoretical Computer Science 153(2), 5–31 (2006)
16. Kwiatkowska, M., Norman, G., Parker, D.: PRISM 4.0: Verification of probabilistic real-time systems. In: Gopalakrishnan, G., Qadeer, S. (eds.) Proc. 23rd International Conference on Computer Aided Verification (CAV'11). LNCS, vol. 6806, pp. 585–591. Springer (2011)
17. Liu, F., Heiner, M., Yang, M.: An efficient method for unfolding colored petri nets. In: Laroque, C., Himmelspach, J., Pasupathy, R., Rose, O., Uhrmacher, A. (eds.) Proceedings of the Winter Simulation Conference. p. 295. Winter Simulation Conference (2012)
18. Mizera, A., Czeizler, E., Petre, I.: Self-assembly models of variable resolution. LNBI Transactions on Computational Systems Biology pp. 181–203 (2011)
19. Murphy, E., Danos, V., Feret, J., Krivine, J., Harmer, R.: Elements of Computational Systems Biology, chap. Rule Based Modelling and Model Refinement, pp. 83–114. Wiley Book Series on Bioinformatics, John Wiley & Sons, Inc. (2010)
20. Petre, I., Mizera, A., Hyder, C., Meinander, A., Mikhailov, A., Morimoto, R., Sistonen, L., Eriksson, J., Back, R.: A simple mass-action model for the eukaryotic heat shock response and its mathematical validation. Natural Computing 10(1), 595–612 (2011)
21. Raman, K., Chandra, N.: Systems biology. Resonance 15(2), 131–153 (2010)
22. Rieger, T., Morimoto, R., Hatzimanikatis, V.: Mathematical modeling of the eukaryotic heat-shock response: Dynamics of the hsp70 promoter. Biophysical journal 88(3), 1646–1658 (2005)
23. Shi, Y., Mosser, D., Morimoto, R.: Molecular chaperones as hsf1-specific transcriptional repressors. Genes & development 12(5), 654–666 (1998)
24. Westerheide, S., Anckar, J., Stevens Jr, S., Sistonen, L., Morimoto, R.: Stress-inducible regulation of heat shock factor 1 by the deacetylase sirt1. Science Signalling 323(5917), 1063–1066 (2009)