TUCS

Elena Czeizler | Eugen Czeizler | Bogdan Iancu | Ion Petre

Quantitative model refinement as a solution to the combinatorial size explosion of biomodels

TURKU CENTRE for COMPUTER SCIENCE

TUCS Technical Report

No 1015, June 2011

Quantitative model refinement as a solution to the combinatorial size explosion of biomodels

Elena Czeizler
> Research Programs Unit, Genome-Scale Biology, Faculty of Medicine,
> University of Helsinki, FI-00014 Helsinki, Finland
> `elena.czeizler@helsinki.fi`

Eugen Czeizler
> Department of Information and Computer Science,
> Aalto University School of Science, FI-00076 Aalto, Finland
> `eugen.czeizler@aalto.fi`

Bogdan Iancu
> Computational Biomodelling Laboratory, Åbo Akademi University
> and Turku Centre for Computer Science, FI-20520 Turku, Finland
> `biancu@abo.fi`

Ion Petre
> Computational Biomodelling Laboratory, Åbo Akademi University
> and Turku Centre for Computer Science, FI-20520 Turku, Finland
> `ipetre@abo.fi`

Abstract

Building a large system through a systematic, step-by-step refinement of an initial abstract specification is a well established technique in software engineering, not yet much explored in systems biology. In the case of systems biology, one starts from an abstract, high-level model of a biological system and aims to add more and more details about its reactants and/or reactions, through a number of consecutive refinement steps. The refinement should be done in a quantitatively correct way, so that (some of) the numerical properties of the model (such as the experimental fit and validation) are preserved. In this study, we focus on the data-refinement mechanism where the aim is to increase the level of details of some of the reactants of a given model. That is, we analyse the case when a model is refined by substituting a given species by several types of subspecies. We show in this paper how the refined model can be systematically obtained from the original one. As a case study for this methodology we choose a recently introduced model for the eukaryotic heat shock response, [19]. We refine this model by including details about the acetylation of the heat shock factors and its influence on the heat shock response. The refined model has a significantly higher number of kinetic parameters and variables. However, we show that our methodology allows us to preserve the experimental fit/validation of the model with minimal computational effort.

Keywords: Model refinement, quantitative analysis, heat shock response, acetylation.

TUCS Laboratory
Computational Biomodelling Laboratory

1 Introduction

Extensive experimental and computational effort is invested nowadays in compiling large, system-level models for complex biological systems, including regulatory networks, signaling pathways, metabolic pathways etc. The first step in developing such system-level models lies usually in creating an abstraction of the biological process consisting of a relatively small number of biochemical reactions describing the main mechanisms of the considered process. The chosen reactions can be abstract representations of some particular subprocess, encapsulating, in fact, many biochemical reactions from the considered system. A mathematical model is then associated to the molecular model comprising these chosen reactions. For this, one chooses an appropriate kinetic law, e.g., mass-action law or Michaelis-Menten kinetics, based on which one can then write the mathematical equations describing the dynamics of the system. The numerical setup of this mathematical model is either obtained from the literature, or, using available experimental data, it is derived through various computational model fit procedures.

Starting from this abstract model, the process of model development proceeds with a series of iterative steps involving hypothesis generation, experimental design, experimental analysis and model refinement, [2], [12]. Particularly, the simplifications and abstractions included in the initial model might be refined later on, including more accurate details of the process. One approach for this would be to simply repeat the whole model development procedure in order to include all the intended changes. However, this can be extremely inefficient since it requires to re-fit the model, a step which is both time-consuming and computationally-intensive, [3]. Another approach, that seems little investigated so far, is to refine the initial model step-by-step making sure that the experimental (numerical) model fit is preserved. In other words, the numerical setup of the refined model should be obtained from that of the initial model, see [16] for a recent case study regarding self-assembly models.

In this study, we focus on the model refinement step within the above model development cycle. In particular, we analyse the case when the model is adjusted by refining some of its reactants, i.e., by replacing a given species by some of its subspecies. This is the case, for instance, when more details about the post-translational modifications of proteins, e.g., acetylation or phosphorylation, are required. In such a case, the model is refined by replacing a given protein P with its variants indicating whether P withstood some post-translational modification or not, e.g., whether P was acetylated or not. This substitution of species also leads to a refinement of all the reactions in which protein P was involved. Then, we show how we can attain the values of the parameters of the refined model from those of the initial model. Hence, we make sure that we preserve all previously obtained systemic properties, such as numerical fit of the model.

The paper is organized as follows. First, we present the model refinement procedure consisting in replacing a species with several subspecies while still pre-

serving the previously obtained model fit with respect to some experimental data. Then, we illustrate this technique by considering as a case study the regulatory role of protein acetylation within the eukaryotic heat shock response. In particular, we refine a recent model for the eukaryotic heat shock response, see [19], by considering also the acetylation of some particular proteins.
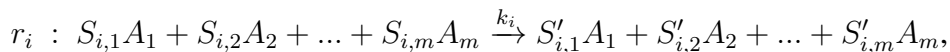
2 Quantitative model refinement

There are several types of refinement that can be applied on a given model. For instance, one could focus on the data of a given model and refine it by replacing one (or more) of its species with a number of subspecies. This way, the refined model would illustrate various differences in the behaviour of those subspecies. We call this *data refinement*. Another type of refinement, that we call *process refinement*, concerns the model reactions. That is, the model is refined by replacing a generic reaction describing a particular process with several reactions detailing on the intermediary steps of that process.

The problem of formal refinement has been considerably documented in the field of software engineering, particularly related to the concurrent computing paradigm [24], [1]. It aroused from the need to prove (in a formal/mathematical way) that the final implementation of a system corresponds to its original specifications. In systems biology, the problem of quantitative model refinement has already been investigated within the framework of rule-based modelling, see [17], [10], [5]. This approach is not concerned with the *data refinement* we have previously defined, since this type of refinement is already built in the system through agent resolution [10]. However, the main consideration in this respect is rule refinement, a method to refine set rules so that the dynamic behaviour of the model is preserved [17]. Rule based modelling allows the construction of more detailed models through resolution augmentation in the space of agents and rules, comprising model variants in more finely-grained hierarchical structures [10]. This type of framework enables writing feasible models and operates on their perturbation spaces in order to evaluate perturbation effects over a particular model [5].

However, independently of the type of refinement chosen, the refinement process should preserve the systemic properties of the original model, e.g., the model fit. We present here an approach for data model refinement which preserves the previously obtained numerical properties of the model.

Consider that a model M consists of a list of m species $\Sigma = \{A_1, A_2, ..., A_m\}$ and n reactions r_i, $1 \leq i \leq n$, of the form:

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

where $S_{i,1}, ..., S_{i,m}, S'_{i,1}, ..., S'_{i,m} \geq 0$ are some integers called *the stoichiometric coefficients* of r_i and $k_i \geq 0$ is *the kinetic rate constant* of r_i. A number of different mathematical formulations can be associated to model M, in terms of

continuous or discrete variables, deterministic or non-deterministic evolution etc. We choose in this study a continuous, mass-action formulation, where to each variable $A_i, 1 \leq i \leq n$, we associate a time-dependent function $[A_i] : \mathbb{R}_+ \to \mathbb{R}_+$ describing its concentration level. In particular, for each species A_i, we denote by $[A_i](t)$ its concentration at time t. The dynamics of the system is then described through a system of differential equations [13] in which, for each reaction, we assumed the principle of mass action, originally introduced in [8], [9].
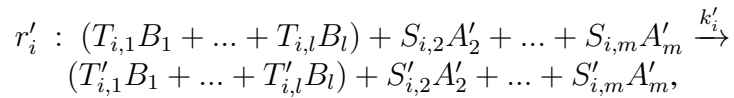
In particular, for model M we obtain the following system of ODE:

$$\frac{d[A_j]}{dt} = -\sum_{i=1}^{n} (k_i S_{i,j} \prod_{k=1}^{m} [A_k]^{S_{i,k}}) + \sum_{i=1}^{n} (k_i S'_{i,j} \prod_{k=1}^{m} [A_k]^{S_{i,k}}), \;\; 1 \leq j \leq m. \quad (1)$$

The next lemma tackles the existence and uniqueness of solutions of systems of ODEs derived following the mass action law.

Lemma 1. *[11] Given a molecular model and its associated system of ODEs derived based on the principle of mass action, for any fixed initial condition, there is an open interval $\boldsymbol{I} \subset \mathbb{R}_+$, such that the system of ODEs has a unique solution on $\boldsymbol{I}$.*

Assume that the model M is to be detailed by distinguishing several subspecies of A_1. Such subspecies may be several different forms of A_1, several biochemical configurations of A_1 (e.g., caused by some post-translational modifications) etc. Each of these subspecies may participate in all reactions, where A_1 participated (in model M), possibly with different kinetics. If A_1 is to be replaced with subspecies $B_1, \ldots, B_l$, then we derive a new model M_R, in which the set of species is denoted through the new variables $\{A'_2, A'_3, ..., A'_m\} \cup \{B_1, ..., B_l\}$, for some $l \geq 2$. Variables $A'_i, 2 \leq i \leq m$, correspond to species A_i from model M, whereas $B_1, ..., B_l$ are to replace species A_1 in model M_R. Moreover, each reaction r_i from M is replaced in M_R by reaction r'_i of the following type:

$$r'_i : (T_{i,1}B_1 + ... + T_{i,l}B_l) + S_{i,2}A'_2 + ... + S_{i,m}A'_m \xrightarrow{k'_i}$$
$$(T'_{i,1}B_1 + ... + T'_{i,l}B_l) + S'_{i,2}A'_2 + ... + S'_{i,m}A'_m,$$

with k'_i its kinetic rate constant, and $T_{i,1}, ..., T_{i,l}, T'_{i,1}, ..., T'_{i,l}$ nonnegative integers such that $T_{i,1} + ... + T_{i,l} = S_{i,1}$ and $T'_{i,1} + ... + T'_{i,l} = S'_{i,1}$. We say now that *the model M_R is a data refinement of M on variable A_1* if and only if the following two conditions are satisfied:

$$[A_i](t) = [A'_i](t), \text{ for all } 2 \leq i \leq m, \quad (2)$$
$$[A_1](t) = [B_1](t) + ... + [B_l](t), \text{ for all } t \geq 0. \quad (3)$$

The refined model M_R has $m + l - 1$ species, whereas M consists of m species; thus, M_R has a linear-increase in the size of its data set. The number of reactions

replacing in M_R the reaction r_i of M is given by the number of non-negative integer solutions of the following system of equations:

$$\begin{aligned} T_{i,1} + T_{i,2} + ... + T_{i,l} &= S_i; \\ T'_{i,1} + T'_{i,2} + ... + T'_{i,l} &= S'_i; \end{aligned}$$

over the independent unknowns $T_{i,j}, T'_{i,j}, 1 \leq j \leq l$. The number of solutions of the first equation is the *multinomial coefficient* "l multichooses S_i", see [6]:

$$\left(\binom{1}{S_i} \right) = \binom{l + S_i - 1}{S_i} = \frac{(l + S_i - 1)!}{S_i!(l - 1)!}.$$

Since the two equations in the system are independent, the number of solutions of the system is $\left(\binom{l}{S_i} \right) \cdot \left(\binom{l}{S'_i} \right)$. This gives the increase in the number of reactions in the model refinement from M to M_R. In terms of kinetic parameters, M_R will have $\left(\binom{l}{S_i} \right)$ new free parameters, given by the number of possible combinations of $B_1, ..., B_l$ as reactants.

Arguing on the basis of biokinetics, see [18], we may assume that $S_i \leq 2$: any reaction with a ternary (or higher) stoichiometric coefficient would be so slow that its effects may be ignored. In case $S_i = 1$, M_R will have $\left(\binom{l}{1} \right) = l$ new kinetic parameters, i.e., a linear increase in the parameter space. In case $S_i = 2$, M_R will have $\left(\binom{l}{2} \right) = l(l+1)/2$ new parameters, i.e., a quadratic increase in the parameter space.

Some of the new kinetic parameters of M_R may be known from the literature, or the they can be experimentally measured. For the rest of them, for which no previous knowledge and no *direct* kinetic measurements are available, a computational procedure is needed to calculate them so that (2) and (3) hold. Such a procedure should focus only on those parameters whose kinetic values are not known.

Re-running parameter estimation procedures when the parameter space witnesses a (potentially) quadratic increase in every step of the refinement is computationally very expensive. Moreover, such a procedure makes little sense since the fit of an intermediate model is lost in the next refinement step.

We propose in this paper an approach where we systematically set the values of the unknown kinetic parameters of the refined model so that relations (2) and (3) hold. Clearly, some of the potential choices are unreasonable, such as those where we would set to 0 the kinetic parameters of all reactions involving $B_2, B_3, ...,$ or B_l; such a choice would eliminate the idea of refinement and it would only rename all the variables of M in M_R. Instead, we take an approach where the refined subspecies $B_1, ..., B_l$ of species A_1 are not distinguishable through the kinetics of the reactions they participate in. Thus, in the absence of any biological data regarding differences between some of these reactants, our choice of kinetic parameters aims to make no numerical distinctions between their

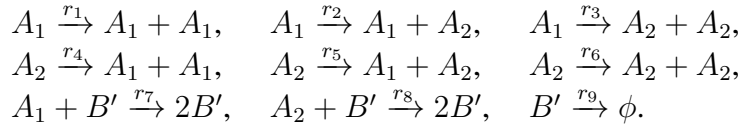reaction kinetics. As a side effect, this leads to simpler and more elegant mathematical considerations.

Consider next an example of such a data refinement procedure, in which we show how we can obtain the kinetic rate constants of the refined model depending on the type of equations included in the original model. We consider the Lotka-Volterra system, [15], [22], composed of the species A (the prey) and B (the predator), and the following reactions:

$$A \xrightarrow{k_1} 2A, \quad A + B \xrightarrow{k_2} 2B, \quad B \xrightarrow{k_3} \emptyset.$$
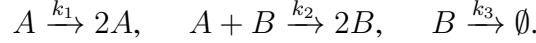
The set of mass-action based ODEs describing the dynamics of this system are:

$$d[A]/dt = k_1[A] - k_2[A][B] \qquad d[B]/dt = k_2[A][B] - k_3[B].$$

After refining the Lotka-Volterra model on variable A into subspecies A_1 and A_2, we obtain the following set of reactions:

$$A_1 \xrightarrow{r_1} A_1 + A_1, \quad A_1 \xrightarrow{r_2} A_1 + A_2, \quad A_1 \xrightarrow{r_3} A_2 + A_2,$$
$$A_2 \xrightarrow{r_4} A_1 + A_1, \quad A_2 \xrightarrow{r_5} A_1 + A_2, \quad A_2 \xrightarrow{r_6} A_2 + A_2,$$
$$A_1 + B' \xrightarrow{r_7} 2B', \quad A_2 + B' \xrightarrow{r_8} 2B', \quad B' \xrightarrow{r_9} \phi.$$

This leads to the following system of ODEs describing the dynamics of the variables A_1, A_2, and B':

$$d[A_1]/dt = (r_1 - r_3)[A_1] + (2r_4 + r_5)[A_2] - r_7[A_1][B'],$$
$$d[A_2]/dt = (r_2 + 2r_3)[A_1] + (-r_4 + r_6)[A_2] - r_8[A_2][B'],$$
$$d[B']/dt = r_7[A_1][B'] + r_8[A_2][B'] - r_9[B'].$$

Thus, $d([A_1]+[A_2])/dt = (r_1 - r_3 + r_2 + 2r_3)[A_1] + (2r_4 + r_5 - r_4 + r_6)[A_2] - (r_7 + r_8)[A_1][B']$. Consequently, if we choose $r_1 = r_2 = r_3 = r_4 = r_5 = r_6 = k_1/3$, $r_7 = r_8 = k_2$ and $r_9 = k_3$ then we obtain

$$d([A_1] + [A_2])/dt = k_1([A_1] + [A_2]) - k_2([A_1] + [A_2])[B'],$$
$$d[B']/dt = k_2([A_1] + [A_2])[B'] - k_3[B'],$$

which is identical to the initial system up to a renaming of variables where $[A]$ is replaced by $[A_1] + [A_2]$ and $[B]$ is replaced by $[B']$. For any $x_0 \geq 0$, if we set the initial values of the variables A_1, A_2, and B' such that $[A](x_0) = [A_1](x_0) + [A_2](x_0)$ and $[B](x_0) = [B'](x_0)$, it follows from Lemma 1 that there exists an open interval I including x_0 such that $[A](t) = [A_1](t) + [A_2](t)$ and $[B](t) = [B'](t)$, for all $t \in I$. That is, the second model is indeed a quantitative refinement of the initial one (on the interval I).
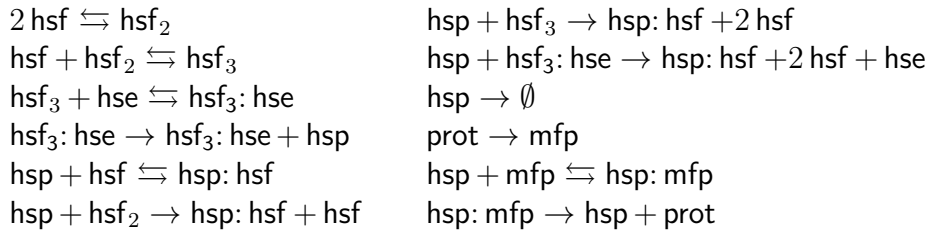
3 Models

3.1 The eukaryotic heat shock response: a molecular model

The heat shock response in eukaryotes is an evolutionarily conserved mechanism that controls the cellular response to proteotoxicity originating from environmental stressors such as elevated temperatures. When subjected to increased temperatures, proteins in the cell tend to misfold and accumulate in large aggregates that eventually induce cell death. Survival of the cell is promoted by a mechanism that restores protein homeostatis, i.e. the equilibrium between synthesis, folding and degradation of proteins.

We describe the molecular model for the heat shock response proposed in [20] as follows. The key factors within this process are the heat shock proteins (hsp), that act as chaperones, assisting the misfolded proteins (mfp) in their refolding process. The response is regulated by the transactivation of the hsp-encoding genes. The transcription of the gene is mediated by specific heat shock transcription factors (hsf). The hsf's trimerize (hsf_3) and subsequently bind to a promoter site of the hsp-encoding gene, called heat shock element (hse). The trimerization phase of the heat shock factors is preceded by a dimerization stage resulting in the constitution of dimers (hsf_2). Hsf trimers bind to heat shock elements forming hsf_3: hse complexes. Therefore, protein synthesis is activated and new hsp molecules are ultimately formed. When the level of hsp's is sufficiently elevated, hsp synthesis is turned off through an ingenious mechanism [19, 20]. Heat shock proteins bind to the hsf molecules forming hsp: hsf complexes, thus impeding hsf's to trimerize and to bind to the heat shock elements. The sequestration of hsf's can be done in three different ways: by breaking dimers, trimers, and by unbinding hsf_3 from the heat shock elements combined with the simultaneous breaking of the trimer. However, an increase in the temperature causes some of the proteins (prot) to misfold, which drives hsp away from hsf. This in turn quickly turns on the heat shock response since the heat shock factors are again free and thus able to promote the synthesis of more heat shock proteins. The reactions of the molecular model in [19] are presented in Table 1.

Table 1: The molecular model for the eukaryotic heat shock response proposed in [19].

$2\,\mathsf{hsf} \leftrightarrows \mathsf{hsf}_2$	$\mathsf{hsp} + \mathsf{hsf}_3 \to \mathsf{hsp}\!:\!\mathsf{hsf} + 2\,\mathsf{hsf}$
$\mathsf{hsf} + \mathsf{hsf}_2 \leftrightarrows \mathsf{hsf}_3$	$\mathsf{hsp} + \mathsf{hsf}_3\!:\!\mathsf{hse} \to \mathsf{hsp}\!:\!\mathsf{hsf} + 2\,\mathsf{hsf} + \mathsf{hse}$
$\mathsf{hsf}_3 + \mathsf{hse} \leftrightarrows \mathsf{hsf}_3\!:\!\mathsf{hse}$	$\mathsf{hsp} \to \emptyset$
$\mathsf{hsf}_3\!:\!\mathsf{hse} \to \mathsf{hsf}_3\!:\!\mathsf{hse} + \mathsf{hsp}$	$\mathsf{prot} \to \mathsf{mfp}$
$\mathsf{hsp} + \mathsf{hsf} \leftrightarrows \mathsf{hsp}\!:\!\mathsf{hsf}$	$\mathsf{hsp} + \mathsf{mfp} \leftrightarrows \mathsf{hsp}\!:\!\mathsf{mfp}$
$\mathsf{hsp} + \mathsf{hsf}_2 \to \mathsf{hsp}\!:\!\mathsf{hsf} + \mathsf{hsf}$	$\mathsf{hsp}\!:\!\mathsf{mfp} \to \mathsf{hsp} + \mathsf{prot}$

6

3.2 Mathematical model

Associated to the molecular model in Table 1, we consider a mathematical model formulated in terms of mass-action based ODEs, see [13]. The model was originally considered in [19]. We include the system of ODEs in Table 2 and we refer the reader to [19] for more details.

Both the kinetic rate constants and the initial values of all reactants were estimated in [19], by imposing the following three conditions:

(i) the system is in a steady state if the temperature is $37°C$. This is a natural consequence of the fact that the model should not exhibit any response in the absence of the heat shock, i.e., at $37°C$;

(ii) the numerical predictions of the model for $[\mathsf{hsf}_3\!:\!\mathsf{hse}](t)$ should agree with the experimental data from [14], for a temperature of $42°C$;

(iii) the numerical prediction of the model for $[\mathsf{hsp}](t)$ should confirm the data obtained in [19] through a de-novo fluorescent reporter-based experiment, for a temperature of $42°C$.

3.3 The role of protein acetylation within the eukaryotic heat shock response

It has been recently shown that the acetylation of the heat shock factors (hsf), i.e., the transcription factors for the hsp-encoding genes, plays an important role in regulating the heat shock response [23]. The acetylation process consists in substituting an acetyl group for a hydrogen atom within a chemical compound. The reverse process, i.e., the deacetylation, represents the suppression of an acetyl group from a compound. Protein acetylation can occur at the alpha-amino group of the amino-terminus (N-terminal acetylation) or on the lysine residues at the epsilon-amino group (lysine acetylation) [7]. The lysine acetylation in particular is known to play a significant role in gene regulation by changing the charge of histone tails. Due to the neutralization of the positive charge of the histones, lysine acetylation diminishes their DNA binding affinity [4, 21].

3.4 Data refinement of the model

We discuss, in this section, how to extend the heat shock model of [19] to account for the acetylation of hsf and its influence on the response. For this, we refine all species and complexes involving hsf to account for two subtypes of hsf: one where its K80 residue is not acetylated, and one where it is acetylated. Consequently, the complex hsp: hsf will also be refined to two subtypes, depending on the acetylation of its hsf component. In the case of hsf_2, hsf_3 and $\mathsf{hsf}_3\!:\!\mathsf{hse}$, the refinement will be accomplished by counting how many of the hsf components in that complex

(two in the case of hsf_2 and three in the case of hsf_3 and hsf_3: hse) have their K80 residue acetylated. Thus, we perform the following data refinements:

$$hsf \rightarrow \{rhsf, rhsf^{(1)}\};$$
$$hsf_2 \rightarrow \{rhsf_2, rhsf_2^{(1)}, rhsf_2^{(2)}\};$$
$$hsf_3 \rightarrow \{rhsf_3, rhsf_3^{(1)}, rhsf_3^{(2)}, rhsf_3^{(3)}\};$$
$$hsf_3: hse \rightarrow \{rhsf_3: rhse, rhsf_3^{(1)}: rhse, rhsf_3^{(2)}: rhse, rhsf_3^{(3)}: rhse\};$$
$$hsp: hsf \rightarrow \{rhsp: rhsf, rhsp: rhsf^{(1)}\}.$$

These data refinements imply several changes in the list of reactions of our model. For example, the reaction $hsp + hsf \leftrightarrows hsp$: hsf is replaced by two reactions : $rhsp + rhsf \leftrightarrows rhsp$: rhsf, and $rhsp + rhsf^{(1)} \leftrightarrows rhsp$: $rhsf^{(1)}$. As another example, reaction $2\,hsf \leftrightarrows hsf_2$ is replaced by three reactions, based on the method described in Section 2:

$$2\,rhsf \leftrightarrows rhsf_2; \quad rhsf + rhsf^{(1)} \leftrightarrows rhsf_2^{(1)}; \quad 2rhsf^{(1)} \leftrightarrows rhsf_2^{(2)}.$$
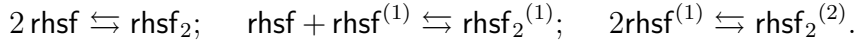
Due to space limitations, the complete list of reactions is given in Appendix B. The refined model consists of 39 reactions, some of which are reversible.

3.5 Quantitative equivalence of the basic and the refined model

We discuss now the numerical setup of the refined model for the heat shock response in such a way that the refinement procedure is quantitatively correct as defined in (2)-(3). The initial values of the refined variables are set so that the following 10 conditions (derived from the data refinement relationships) are satisfied:

$$[hsf](0) = [rhsf](0) + [rhsf^{(1)}](0);$$
$$[hsf_2](0) = [rhsf_2](0) + [rhsf_2^{(1)}] + [rhsf_2^{(2)}](0);$$
$$[hsf_3](0) = [rhsf_3](0) + [rhsf_3^{(1)}](0) + [rhsf_3^{(2)}](0) + [rhsf_3^{(3)}](0);$$
$$[hsp: hsf](0) = [hsp: rhsf](0) + [rhsp: rhsf^{(1)}](0);$$
$$[hsf_3: hse](0) = [rhsf_3: rhse](0) + [rhsf_3^{(1)}: rhse](0) + [rhsf_3^{(2)}: rhse](0) +$$
$$+ [rhsf_3^{(3)}: rhse](0);$$
$$[hsp](0) = [rhsp](0);$$
$$[hsp: mfp](0) = [rhsp: rmfp](0);$$
$$[mfp](0) = [rmfp](0);$$
$$[prot](0) = [rprot](0);$$
$$[hse](0) = [rhse](0).$$

The system of mass-action based ODEs for the refined model is in Table 4. The refined model consists of 20 species, 39 reactions (some of them reversible),

8

54 kinetic parameters. We recall that the basic model had 10 species, 12 reactions and 16 kinetic parameters. We focus now on the numerical setup of the refined model so that it is a quantitative refinement of the model in [19], as defined in Section 2. We first introduce the following notations:

$$\mathsf{Rhsf} = \mathsf{rhsf} + \mathsf{rhsf}^{(1)};$$
$$\mathsf{Rhsf}_2 = \mathsf{rhsf}_2 + \mathsf{rhsf}_2{}^{(1)} + \mathsf{rhsf}_2{}^{(2)};$$
$$\mathsf{Rhsf}_3 = \mathsf{rhsf}_3 + \mathsf{rhsf}_3{}^{(1)} + \mathsf{rhsf}_3{}^{(2)} + \mathsf{rhsf}_3{}^{(3)};$$
$$\mathsf{Rhsf}_3\!:\mathsf{Rhse} = \mathsf{rhsf}_3\!:\mathsf{rhse} + \mathsf{rhsf}_3{}^{(1)}\!:\mathsf{rhse} + \mathsf{rhsf}_3{}^{(2)}\!:\mathsf{rhse} + \mathsf{rhsf}_3{}^{(3)}\!:\mathsf{rhse};$$
$$\mathsf{Rhsp}\!:\mathsf{Rhsf} = \mathsf{rhsp}\!:\mathsf{rhsf} + \mathsf{rhsp}\!:\mathsf{rhsf}^{(1)}.$$

We aim to identify some values for the kinetic parameters of the refined model in such a way that its system of differential equations is identical to the system associated to the initial model, modulo a variable renaming where hsf, hsf_2, hsf_3, hsf_3: hse, and hsp: hsf are replaced by Rhsf, Rhsf_2, Rhsf_3, Rhsf_3: Rhse, and Rhsp: Rhsf, respectively. To drive this process, we write the ODEs for [Rhsf], [Rhsf_2],[Rhsf_3], [Rhsf_3: Rhse] and for [Rhsp: Rhsf], see Table 5, based on the system of ODEs in Table 4. We then choose the values of the kinetic parameters in such a way that the right hand side of each ODE in Table 5 becomes identical to the right hand side of the corresponding ODE in Table 2, modulo the variable renaming above. For example, we aim to choose the kinetic parameters of the refined model in such a way that the ODE corresponding to Rhsf in Table 5 is identical to the ODE corresponding to hsf in Table 2, modulo the variable renaming above. To identify the first term of the ODE for Rhsf (in the form written in Table 5) with the first term of the ODE for hsf (in the form written in Table 2), it is enough to set $r_1^+ = k_1^+$, $r_2^+ = 2k_1^+$, and $r_3^+ = k_1^+$. A similar reasoning for all terms of all ODEs leads to a solution.

Due to space limitations, we skip all details and only include the table listing the identified values for the parameters in Appendix D. Clearly, the solution is not unique. However, to find one, we cannot count on solving the systems of ODEs in Tables 2 and 5; in fact these systems cannot be solved analytically. Instead, we simply chose the values of the kinetic parameters of the refined model as expressions of the kinetic parameters of the original model, in such a way that the ODEs in Table 5 can be rewritten to the equations in Table 2, modulo the variable renaming above. Based also on the way we set up the initial values of the refined variables, it follows that the systems of differential equations associated to the initial model and to the refined one have identical initial conditions. Thus, it follows from Lemma 1 that conditions (2) and (3) are satisfied, i.e., the second model is indeed a quantitative refinement of the initial model in Table 1.

4 Discussion

We focus in this paper on quantitative model refinement as an essential stage within the complex process of model development in systems biology. In particular, we analyse the case when the model is refined by replacing one species with several subspecies. Specifically, we show that, by attaining the numerical setup of the refined model from that of the initial model, we obtain a quantitative refinement which preserves previously obtained numerical properties of the model, e.g., model fit and validation, see [3].

We refined the model from [19] by considering only one acetylation site for each hsf molecule. This led to a significant augmentation in the number of reactions of the refined model and, consequently, in the number of parameters. While the basic model comprises a set of 12 reactions involving 10 different species and 16 kinetic rate constants, the refined model contains a number of 39 reactions involving 20 species and 54 kinetic rate constants. Fitting a model of this proportion implies a lot of time and computational resources. However, with our approach, we were able to build a refined model, with a satisfactory numerical behaviour (as defined in our notion of quantitative refinement), avoiding any supplementary model fit.

Our solution to the problem of setting the kinetic parameters of the refined model is clearly not unique. Since the systems of ODEs corresponding to the original and to the refined model are in general impossible to solve analytically, we adopt a symbolic approach where we make sure that the two systems of ODEs are isomorphic through the variable renaming given by the data refinement relationship. In the absence of any biological knowledge regarding the kinetic parameters of the refined model, our solution is only driven by the symbolic approach described above. If there is biological knowledge about some of the values of the parameters of the refined model, then such knowledge can be taken into consideration in the form of constraint in our symbolic approach. The existence of a solution in such a case, as well as effectively constructing one appear as interesting problems in this context.

We only focused in this paper on refining the basic heat shock response model of [19] in a quantitatively correct way to include some of the details of the acetylation of hsf. Due to lack of space, we did not include in the refined model the details regarding the role that the acetylation of hsf plays in fine-tuning the heat shock response, as described in [23]. We plan to return to these aspects in a separate study.

5 Acknowledgements

References

[1] L. Aceto, *Action refinement in process algebras*, Distinguished Dissertations in Computer Science, Cambridge University Press, 1992.

[2] F. J. Bruggeman and H. V. Westerhoff. The nature of systems biology. Trends in Microbiology, 15(1):45-50, 2007.

[3] W. W. Chen, B. Schoeberl, P. J. Jasper, M. Niepel, U. B. Nielsen, D. A. Lauffenburger, P. K. Sorger. Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. Molecular Systems Biology, 5(239), 2009.

[4] C. Choudhary, C. Kumar, F. Gnad, M. L. Nielsen, M. Rehman, T. C. Walther, J. V. Olsen, M. Mann, Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. Science, 325(5942), 834-840, 2009.

[5] V. Danos, J. Feret, W. Fontana, R. Harmer, J. Krivine, Rule-based Modelling and Model Perturbation, Transactions on Computational Systems Biology XI, 116-137, 2009.

[6] I. M. Gessel and R. P. Stanley (1995) "Algebraic enumeration" in Graham, R. L., M. Grötschel, L. Lovász (Eds.), Handbook of combinatorics, Vol. 2. Elsevier: 10211061.

[7] M. A. Glozak, N. Sengupta, X. Zhang, E. Seto, Acetylation and deacetylation of non-histone proteins, In: Gene, 363: 15-23, 2005.

[8] C. M. Guldberg and P. Waage, Studies Concerning Affinity. *C. M. Forhandlinger: Videnskabs-Selskabet i Christiana* **35**, 1864.

[9] C. M. Guldberg and P. Waage, Concerning Chemical Affinity. *Erdmann's Journal für Practische Chemie* **127**: 69-114, 1879.

[10] R. Harmer, Rule-based modelling and tunable resolution, S. Barry Cooper & Vincent Danos, editors, Fifth Workshop on Developments in Computational Models - Computational Models from Nature EPTCS 9, 65-71, 2009.

[11] M. W. Hirsch, S. Smale, R. Devaney, *Differential equations, dynamical systems, and an introduction to chaos*, Second Edition, Academic Press, 2004.

[12] H. Kitano. Systems biology: A brief overview. Science, 295(5560):1662-1664, 2002.

[13] E. Klipp, R. Herwig, A. Kowald, C. Wierling, H. Lehrach, Systems Biology in Practice, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2005.

[14] M. P. Kline, R. I. Morimoto, Repression of the heat shock factor 1 transcriptional activation domain is modulated by constitutive phosphorylation. Molecular and cellular biology,17(4),21072115, 1997.

[15] A. J. Lotka, Elements of physical biology. Williams and Wilkins, Baltimore, 1925.

[16] A. Mizera, E. Czeizler, I. Petre, Self-assembly models of variable resolution, submitted, 2011.

[17] E. Murphy, V. Danos, J. Feret, J. Krivine, R. Harmer, Rule-based modelling and model refinement. In Huma M. Lodhi and Stephen H. Muggleton, editors, *Elements of Computational Systems Biology.*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2010.

[18] D. L. Nelson, M. M. Cox, *Lehninger Principles of Biochemistry* Third Edition, Worth Publishers, New York, 2000.

[19] I. Petre, A. Mizera, C. L. Hyder, A. Meinander, A. Mikhailov, R.I. Morimoto, L. Sistonen, J. E. Eriksson, R.-J. Back, A simple mass-action model for the eukaryotic heat shock response and its mathematical validation, Natural Computing, 10(1), 595-612, 2011.

[20] T. R. Rieger, R. I. Morimoto, V. Hatzimanikatis, Mathematical Modeling of the Eukariotic Heat-Shock Response: Dynamics of the hsp70 Promoter, Biophysical Journal, vol. 88, 1646-1658, 2005.

[21] K. Struhl, Histone acetylation and transcriptional regulatory mechanisms,Genes & Development, 12:599-606, 1998.

[22] V. Volterra, In R. N. Chapman 1931. Animal ecology. McGraw-Hill, New York, 409-448, 1926.

[23] S. D. Westerheide, J. Anckar, S. M. Stevens Jr., L. Sistonen, R. I. Morimoto, Stress-Inducible Regulation of Heat Shock Factor 1 by the Deactylase SIRT1, Science, vol. 323, 2009.

[24] N. Wirth, Program Development by Stepwise Refinement, Communications of the ACM, vol. 14, iss. 4, New York, USA, 1971.

A The original ODE-based model for the heat shock response

Table 2: The system of ODEs associated with the biochemical model proposed in [19].

$$d[\mathsf{hsf}]/dt = -2k_1^+[\mathsf{hsf}]^2 + 2k_1^-[\mathsf{hsf}_2] - k_2^+[\mathsf{hsf}][\mathsf{hsf}_2] + k_2^-[\mathsf{hsf}_3]$$
$$-k_5^+[\mathsf{hsf}][\mathsf{hsp}] + k_5^-[\mathsf{hsp}\!:\!\mathsf{hsf}] + k_6[\mathsf{hsf}_2][\mathsf{hsp}]$$
$$+2k_7[\mathsf{hsf}_3][\mathsf{hsp}] + 2k_8[\mathsf{hsf}_3\!:\!\mathsf{hse}][\mathsf{hsp}]$$
$$d[\mathsf{hsf}_2]/dt = k_1^+[\mathsf{hsf}]^2 - k_1^-[\mathsf{hsf}_2] - k_2^+[\mathsf{hsf}][\mathsf{hsf}_2] + k_2^-[\mathsf{hsf}_3]$$
$$-k_6[\mathsf{hsf}_2][\mathsf{hsp}]$$
$$d[\mathsf{hsf}_3]/dt = k_2^+[\mathsf{hsf}][\mathsf{hsf}_2] - k_2^-[\mathsf{hsf}_3] - k_3^+[\mathsf{hsf}_3][\mathsf{hse}] + k_3^-[\mathsf{hsf}_3\!:\!\mathsf{hse}]$$
$$-k_7[\mathsf{hsf}_3][\mathsf{hsp}]$$
$$d[\mathsf{hse}]/dt = -k_3^+[\mathsf{hsf}_3][\mathsf{hse}] + k_3^-[\mathsf{hsf}_3\!:\!\mathsf{hse}] + k_8[\mathsf{hsf}_3\!:\!\mathsf{hse}][\mathsf{hsp}]$$
$$d[\mathsf{hsf}_3\!:\!\mathsf{hse}]/dt = k_3^+[\mathsf{hsf}_3][\mathsf{hse}] - k_3^-[\mathsf{hsf}_3\!:\!\mathsf{hse}] - k_8[\mathsf{hsf}_3\!:\!\mathsf{hse}][\mathsf{hsp}]$$
$$d[\mathsf{hsp}]/dt = k_4[\mathsf{hsf}_3\!:\!\mathsf{hse}] - k_5^+[\mathsf{hsf}][\mathsf{hsp}] + k_5^-[\mathsf{hsp}\!:\!\mathsf{hsf}] - k_6[\mathsf{hsf}_2][\mathsf{hsp}]$$
$$-k_7[\mathsf{hsf}_3][\mathsf{hsp}] - k_8[\mathsf{hsf}_3\!:\!\mathsf{hse}][\mathsf{hsp}] - k_{11}^+[\mathsf{hsp}][\mathsf{mfp}]$$
$$+(k_{11}^- + k_{12})[\mathsf{hsp}\!:\!\mathsf{mfp}] - k_9[\mathsf{hsp}]$$
$$d[\mathsf{hsp}\!:\!\mathsf{hsf}]/dt = k_5^+[\mathsf{hsf}][\mathsf{hsp}] - k_5^-[\mathsf{hsp}\!:\!\mathsf{hsf}] + k_6[\mathsf{hsf}_2][\mathsf{hsp}]$$
$$+k_7[\mathsf{hsf}_3][\mathsf{hsp}] + k_8[\mathsf{hsf}_3\!:\!\mathsf{hse}][\mathsf{hsp}]$$
$$d[\mathsf{mfp}]/dt = \phi_T[\mathsf{prot}] - k_{11}^+[\mathsf{hsp}][\mathsf{mfp}] + k_{11}^-[\mathsf{hsp}\!:\!\mathsf{mfp}]$$
$$d[\mathsf{hsp}\!:\!\mathsf{mfp}]/dt = k_{11}^+[\mathsf{hsp}][\mathsf{mfp}] - (k_{11}^- + k_{12})[\mathsf{hsp}\!:\!\mathsf{mfp}]$$
$$d[\mathsf{prot}]/dt = -\phi_T[\mathsf{prot}] + k_{12}[\mathsf{hsp}\!:\!\mathsf{mfp}]$$

B The refined model for the heat shock response

Table 3: The list of reactions for the refined model. For an irreversible reaction q_i, r_i denotes its kinetic rate constant. For a reversible reaction q_i, we denote by r_i^+ and r_i^- the kinetic rate constants of its 'left-to-right' and 'right-to-left' directions, respectively.

Reaction	Kinetic rate constants
$2\,\text{rhsf} \leftrightarrows \text{rhsf}_2$	r_1^+, r_1^-
$\text{rhsf} + \text{rhsf}^{(1)} \leftrightarrows \text{rhsf}_2^{(1)}$	r_2^+, r_2^-
$2\text{rhsf}^{(1)} \leftrightarrows \text{rhsf}_2^{(2)}$	r_3^+, r_3^-
$\text{rhsf} + \text{rhsf}_2 \leftrightarrows \text{rhsf}_3$	r_4^+, r_4^-
$\text{rhsf}^{(1)} + \text{rhsf}_2 \leftrightarrows \text{rhsf}_3^{(1)}$	r_5^+, r_5^-
$\text{rhsf} + \text{rhsf}_2^{(1)} \leftrightarrows \text{rhsf}_3^{(1)}$	r_6^+, r_6^-
$\text{rhsf}^{(1)} + \text{rhsf}_2^{(1)} \leftrightarrows \text{rhsf}_3^{(2)}$	r_7^+, r_7^-
$\text{rhsf} + \text{rhsf}_2^{(2)} \leftrightarrows \text{rhsf}_3^{(2)}$	r_8^+, r_8^-
$\text{rhsf}^{(1)} + \text{rhsf}_2^{(2)} \leftrightarrows \text{rhsf}_3^{(3)}$	r_9^+, r_9^-
$\text{rhsf}_3 + \text{rhse} \leftrightarrows \text{rhsf}_3 \colon \text{rhse}$	r_{10}^+, r_{10}^-
$\text{rhsf}_3^{(1)} + \text{rhse} \leftrightarrows \text{rhsf}_3^{(1)} \colon \text{rhse}$	r_{11}^+, r_{11}^-
$\text{rhsf}_3^{(2)} + \text{rhse} \leftrightarrows \text{rhsf}_3^{(2)} \colon \text{rhse}$	r_{12}^+, r_{12}^-
$\text{rhsf}_3^{(3)} + \text{rhse} \leftrightarrows \text{rhsf}_3^{(3)} \colon \text{rhse}$	r_{13}^+, r_{13}^-
$\text{rhsf}_3 \colon \text{rhse} \to \text{rhsf}_3 \colon \text{rhse} + \text{rhsp}$	r_{14}
$\text{rhsf}_3^{(1)} \colon \text{rhse} \to \text{rhsf}_3^{(1)} \colon \text{rhse} + \text{rhsp}$	r_{15}
$\text{rhsf}_3^{(2)} \colon \text{rhse} \to \text{rhsf}_3^{(2)} \colon \text{rhse} + \text{rhsp}$	r_{16}
$\text{rhsf}_3^{(3)} \colon \text{rhse} \to \text{rhsf}_3^{(3)} \colon \text{rhse} + \text{rhsp}$	r_{17}
$\text{rhsp} + \text{rhsf} \leftrightarrows \text{rhsp} \colon \text{rhsf}$	r_{18}^+, r_{18}^-
$\text{rhsp} + \text{rhsf}^{(1)} \leftrightarrows \text{rhsp} \colon \text{rhsf}^{(1)}$	r_{19}^+, r_{19}^-
$\text{rhsp} + \text{rhsf}_2 \to \text{rhsp} \colon \text{rhsf} + \text{rhsf}$	r_{20}
$\text{rhsp} + \text{rhsf}_2^{(1)} \to \text{rhsp} \colon \text{rhsf} + \text{rhsf}^{(1)}$	r_{21}
$\text{rhsp} + \text{rhsf}_2^{(1)} \to \text{rhsp} \colon \text{rhsf}^{(1)} + \text{rhsf}$	r_{22}
$\text{rhsp} + \text{rhsf}_2^{(2)} \to \text{rhsp} \colon \text{rhsf}^{(1)} + \text{rhsf}^{(1)}$	r_{23}
$\text{rhsp} + \text{rhsf}_3 \to \text{rhsp} \colon \text{rhsf} + 2 * \text{rhsf}$	r_{24}
$\text{rhsp} + \text{rhsf}_3^{(1)} \to \text{rhsp} \colon \text{rhsf} + \text{rhsf}^{(1)} + \text{rhsf}$	r_{25}
$\text{rhsp} + \text{rhsf}_3^{(1)} \to \text{rhsp} \colon \text{rhsf}^{(1)} + 2 * \text{rhsf}$	r_{26}
$\text{rhsp} + \text{rhsf}_3^{(2)} \to \text{rhsp} \colon \text{rhsf} + 2\text{rhsf}^{(1)}$	r_{27}
$\text{rhsp} + \text{rhsf}_3^{(2)} \to \text{rhsp} \colon \text{rhsf}^{(1)} + \text{rhsf}^{(1)} + \text{rhsf}$	r_{28}
$\text{rhsp} + \text{rhsf}_3^{(3)} \to \text{rhsp} \colon \text{rhsf}^{(1)} + 2\text{rhsf}^{(1)}$	r_{29}
$\text{rhsp} + \text{rhsf}_3 \colon \text{rhse} \to \text{rhsp} \colon \text{rhsf} + 2\,\text{rhsf} + \text{rhse}$	r_{30}
$\text{rhsp} + \text{rhsf}_3^{(1)} \colon \text{rhse} \to \text{rhsp} \colon \text{rhsf}^{(1)} + 2\,\text{rhsf} + \text{rhse}$	r_{31}
$\text{rhsp} + \text{rhsf}_3^{(1)} \colon \text{rhse} \to \text{rhsp} \colon \text{rhsf} + \text{rhsf}^{(1)} + \text{rhsf} + \text{rhse}$	r_{32}
$\text{rhsp} + \text{rhsf}_3^{(2)} \colon \text{rhse} \to \text{rhsp} \colon \text{rhsf}^{(1)} + \text{rhsf}^{(1)} + \text{rhsf} + \text{rhse}$	r_{33}

Table 3: The list of reactions for the refined model - Continued

$\text{rhsp} + \text{rhsf}_3^{(2)} \colon \text{rhse} \to \text{rhsp} \colon \text{rhsf} + 2\text{rhsf}^{(1)} + \text{rhse}$ r_{34}

$\text{rhsp} + \text{rhsf}_3^{(3)} \colon \text{rhse} \to \text{rhsp} \colon \text{rhsf}^{(1)} + 2\text{rhsf}^{(1)} + \text{rhse}$ r_{35}

$\text{rhsp} \to \emptyset$ r_{36}

$\text{rprot} \to \text{rmfp}$ r_{37}

$\text{rhsp} + \text{rmfp} \leftrightarrows \text{rhsp} \colon \text{rmfp}$ r_{38}^+, r_{38}^-

$\text{rhsp} \colon \text{rmfp} \to \text{rhsp} + \text{rprot}$ r_{39}

C The ODE-based model of the refined heat shock response model

Table 4: The system of differential equations of the mathematical model associated with the refined heat shock response model

$$d[\text{rhsf}]/dt = -2r_1^+[\text{rhsf}]^2 + 2r_1^-[\text{rhsf}_2] - r_2^+[\text{rhsf}][\text{rhsf}^{(1)}] + r_2^-[\text{rhsf}_2^{(1)}]$$
$$-r_4^+[\text{rhsf}][\text{rhsf}_2] + r_4^-[\text{rhsf}_3] - r_6^+[\text{rhsf}][\text{rhsf}_2^{(1)}] + r_6^-[\text{rhsf}_3^{(1)}]$$
$$-r_8^+[\text{rhsf}][\text{rhsf}_2^{(2)}] + r_8^-[\text{rhsf}_3^{(2)}] - r_{18}^+[\text{rhsp}][\text{rhsf}]$$
$$+r_{18}^-[\text{rhsp:rhsf}] + r_{20}[\text{rhsp}][\text{rhsf}_2] + r_{22}[\text{rhsp}][\text{rhsf}_2^{(1)}]$$
$$+2r_{24}[\text{rhsp}][\text{rhsf}_3] + r_{25}[\text{rhsp}][\text{rhsf}_3^{(1)}] + 2r_{26}[\text{rhsp}][\text{rhsf}_3^{(1)}]$$
$$+r_{28}[\text{rhsp}][\text{rhsf}_3^{(2)}] + 2r_{30}[\text{rhsp}][\text{rhsf}_3:\text{rhse}]$$
$$+2r_{31}[\text{rhsp}][\text{rhsf}_3^{(1)}:\text{rhse}] + r_{32}[\text{rhsp}][\text{rhsf}_3^{(1)}:\text{rhse}]$$
$$+r_{33}[\text{rhsp}][\text{rhsf}_3^{(2)}:\text{rhse}]$$

$$d[\text{rhsf}^{(1)}]/dt = -r_2^+[\text{rhsf}][\text{rhsf}^{(1)}] + r_2^-[\text{rhsf}_2^{(1)}] - 2r_3^+[\text{rhsf}^{(1)}]^2$$
$$+2r_3^-[\text{rhsf}_2^{(2)}] - r_5^+[\text{rhsf}^{(1)}][\text{rhsf}_2] + r_5^-[\text{rhsf}_3^{(1)}]$$
$$-r_7^+[\text{rhsf}^{(1)}][\text{rhsf}_2^{(1)}] + r_7^-[\text{rhsf}_3^{(2)}] - r_9^+[\text{rhsf}^{(1)}][\text{rhsf}_2^{(2)}]$$
$$+r_9^-[\text{rhsf}_3^{(3)}] - r_{19}^+[\text{rhsp}][\text{rhsf}^{(1)}] + r_{19}^-[\text{rhsp:rhsf}^{(1)}]$$
$$+r_{21}[\text{rhsp}][\text{rhsf}_2^{(1)}] + r_{23}[\text{rhsp}][\text{rhsf}_2^{(2)}] + r_{25}[\text{rhsp}][\text{rhsf}_3^{(1)}]$$
$$+2r_{27}[\text{rhsp}][\text{rhsf}_3^{(2)}] + r_{28}[\text{rhsp}][\text{rhsf}_3^{(2)}] + 2r_{29}[\text{rhsp}][\text{rhsf}_3^{(3)}]$$
$$+r_{32}[\text{rhsp}][\text{rhsf}_3^{(1)}:\text{rhse}] + r_{33}[\text{rhsp}][\text{rhsf}_3^{(2)}:\text{rhse}]$$
$$+2r_{34}[\text{rhsp}][\text{rhsf}_3^{(2)}:\text{rhse}] + 2r_{35}[\text{rhsp}][\text{rhsf}_3^{(3)}:\text{rhse}]$$

$$d[\text{rhsf}_2]/dt = r_1^+[\text{rhsf}]^2 - r_1^-[\text{rhsf}_2] - r_4^+[\text{rhsf}][\text{rhsf}_2] + r_4^-[\text{rhsf}_3]$$
$$-r_5^+[\text{rhsf}^{(1)}][\text{rhsf}_2] + r_5^-[\text{rhsf}_3^{(1)}] - r_{20}[\text{rhsp}][\text{rhsf}_2]$$

$$d[\text{rhsf}_2^{(1)}]/dt = r_2^+[\text{rhsf}][\text{rhsf}^{(1)}] - r_2^-[\text{rhsf}_2^{(1)}] - r_6^+[\text{rhsf}][\text{rhsf}_2^{(1)}]$$
$$+r_6^-[\text{rhsf}_3^{(1)}] - r_7^+[\text{rhsf}^{(1)}][\text{rhsf}_2^{(1)}] + r_7^-[\text{rhsf}_3^{(2)}]$$
$$-r_{21}[\text{rhsp}][\text{rhsf}_2^{(1)}] - r_{22}[\text{rhsp}][\text{rhsf}_2^{(1)}]$$

$$d[\text{rhsf}_2^{(2)}]/dt = r_3^+[\text{rhsf}^{(1)}]^2 - r_3^-[\text{rhsf}_2^{(2)}] - r_8^+[\text{rhsf}][\text{rhsf}_2^{(2)}]$$
$$+r_8^-[\text{rhsf}_3^{(2)}] - r_9^+[\text{rhsf}^{(1)}][\text{rhsf}_2^{(2)}] + r_9^-[\text{rhsf}_3^{(3)}]$$
$$-r_{23}[\text{rhsp}][\text{rhsf}_2^{(2)}]$$

$$d[\text{rhsf}_3]/dt = r_4^+[\text{rhsf}][\text{rhsf}_2] - r_4^-[\text{rhsf}_3] - r_{10}^+[\text{rhsf}_3][\text{rhse}]$$
$$+r_{10}^-[\text{rhsf}_3:\text{rhse}] - r_{24}[\text{rhsp}][\text{rhsf}_3]$$

$$d[\text{rhsf}_3^{(1)}]/dt = r_5^+[\text{rhsf}^{(1)}][\text{rhsf}_2] - r_5^-[\text{rhsf}_3^{(1)}] + r_6^+[\text{rhsf}][\text{rhsf}_2^{(1)}]$$
$$-r_6^-[\text{rhsf}_3^{(1)}] - r_{11}^+[\text{rhsf}_3^{(1)}][\text{rhse}] + r_{11}^-[\text{rhsf}_3^{(1)}:\text{rhse}]$$
$$-r_{25}[\text{rhsp}][\text{rhsf}_3^{(1)}] - r_{26}[\text{rhsp}][\text{rhsf}_3^{(1)}]$$

$$d[\text{rhsf}_3^{(2)}]/dt = r_7^+[\text{rhsf}^{(1)}][\text{rhsf}_2^{(1)}] - r_7^-[\text{rhsf}_3^{(2)}] + r_8^+[\text{rhsf}][\text{rhsf}_2^{(2)}]$$
$$-r_8^-[\text{rhsf}_3^{(2)}] - r_{12}^+[\text{rhsf}_3^{(2)}][\text{rhse}] + r_{12}^-[\text{rhsf}_3^{(2)}:\text{rhse}]$$
$$-r_{27}[\text{rhsp}][\text{rhsf}_3^{(2)}] - r_{28}[\text{rhsp}][\text{rhsf}_3^{(2)}]$$

$$d[\text{rhsf}_3^{(3)}]/dt = r_9^+[\text{rhsf}^{(1)}][\text{rhsf}_2^{(2)}] - r_9^-[\text{rhsf}_3^{(3)}] - r_{13}^+[\text{rhsf}_3^{(3)}][\text{rhse}]$$
$$+r_{13}^-[\text{rhsf}_3^{(3)}:\text{rhse}] - r_{29}[\text{rhsp}][\text{rhsf}_3^{(3)}]$$

Table 4: The system of differential equations of the mathematical model associated with the biochemical model - Continued

$$d[\text{rhse}]/dt = -r_{10}^{+}[\text{rhsf}_3][\text{rhse}] + r_{10}^{-}[\text{rhsf}_3\text{: rhse}] - r_{11}^{+}[\text{rhsf}_3^{(1)}][\text{rhse}]$$
$$+r_{11}^{-}[\text{rhsf}_3^{(1)}\text{: rhse}] - r_{12}^{+}[\text{rhsf}_3^{(2)}][\text{rhse}] + r_{12}^{-}[\text{rhsf}_3^{(2)}\text{: rhse}]$$
$$-r_{13}^{+}[\text{rhsf}_3^{(3)}][\text{rhse}] + r_{13}^{-}[\text{rhsf}_3^{(3)}\text{: rhse}] + r_{30}[\text{rhsp}][\text{rhsf}_3\text{: rhse}]$$
$$+r_{31}[\text{rhsp}][\text{rhsf}_3^{(1)}\text{: rhse}] + r_{32}[\text{rhsp}][\text{rhsf}_3^{(1)}\text{: rhse}]$$
$$+r_{33}[\text{rhsp}][\text{rhsf}_3^{(2)}\text{: rhse}] + r_{34}[\text{rhsp}][\text{rhsf}_3^{(2)}\text{: rhse}]$$
$$+r_{35}[\text{rhsp}][\text{rhsf}_3^{(3)}\text{: rhse}]$$

$$d[\text{rhsf}_3\text{: rhse}]/dt = r_{10}^{+}[\text{rhsf}_3][\text{rhse}] - r_{10}^{-}[\text{rhsf}_3\text{: rhse}]$$
$$-r_{30}[\text{rhsp}][\text{rhsf}_3\text{: rhse}]$$

$$d[\text{rhsf}_3^{(1)}\text{: rhse}]/dt = r_{11}^{+}[\text{rhsf}_3^{(1)}][\text{rhse}] - r_{11}^{-}[\text{rhsf}_3^{(1)}\text{: rhse}]$$
$$-r_{31}[\text{rhsp}][\text{rhsf}_3^{(1)}\text{: rhse}] - r_{32}[\text{rhsp}][\text{rhsf}_3^{(1)}\text{: rhse}]$$

$$d[\text{rhsf}_3^{(2)}\text{: rhse}]/dt = r_{12}^{+}[\text{rhsf}_3^{(2)}][\text{rhse}] - r_{12}^{-}[\text{rhsf}_3^{(2)}\text{: rhse}]$$
$$-r_{33}[\text{rhsp}][\text{rhsf}_3^{(2)}\text{: rhse}] - r_{34}[\text{rhsp}][\text{rhsf}_3^{(2)}\text{: rhse}]$$

$$d[\text{rhsf}_3^{(3)}\text{: rhse}]/dt = r_{13}^{+}[\text{rhsf}_3^{(3)}][\text{rhse}] - r_{13}^{-}[\text{rhsf}_3^{(3)}\text{: rhse}]$$
$$-r_{35}[\text{rhsp}][\text{rhsf}_3^{(3)}\text{: rhse}]$$

$$d[\text{rhsp}]/dt = r_{14}[\text{rhsf}_3\text{: rhse}] + r_{15}[\text{rhsf}_3^{(1)}\text{: rhse}] + r_{16}[\text{rhsf}_3^{(2)}\text{: rhse}]$$
$$+r_{17}[\text{rhsf}_3^{(3)}\text{: rhse}] - r_{18}^{+}[\text{rhsp}][\text{rhsf}] + r_{18}^{-}[\text{rhsp: rhsf}]$$
$$-r_{19}^{+}[\text{rhsp}][\text{rhsf}^{(1)}] + r_{19}^{-}[\text{rhsp: rhsf}^{(1)}] - r_{20}[\text{rhsp}][\text{rhsf}_2]$$
$$-r_{21}[\text{rhsp}][\text{rhsf}_2^{(1)}] - r_{22}[\text{rhsp}][\text{rhsf}_2^{(1)}] - r_{23}[\text{rhsp}][\text{rhsf}_2^{(2)}]$$
$$-r_{24}[\text{rhsp}][\text{rhsf}_3] - r_{25}[\text{rhsp}][\text{rhsf}_3^{(1)}] - r_{26}[\text{rhsp}][\text{rhsf}_3^{(1)}]$$
$$-r_{27}[\text{rhsp}][\text{rhsf}_3^{(2)}] - r_{28}[\text{rhsp}][\text{rhsf}_3^{(2)}] - r_{29}[\text{rhsp}][\text{rhsf}_3^{(3)}]$$
$$-r_{30}[\text{rhsp}][\text{rhsf}_3\text{: rhse}] - r_{31}[\text{rhsp}][\text{rhsf}_3^{(1)}\text{: rhse}]$$
$$-r_{32}[\text{rhsp}][\text{rhsf}_3^{(1)}\text{: rhse}] - r_{33}[\text{rhsp}][\text{rhsf}_3^{(2)}\text{: rhse}]$$
$$-r_{34}[\text{rhsp}][\text{rhsf}_3^{(2)}\text{: rhse}] - r_{35}[\text{rhsp}][\text{rhsf}_3^{(3)}\text{: rhse}]$$
$$-r_{36}[\text{rhsp}] - r_{38}^{+}[\text{rhsp}][\text{rmfp}] + r_{38}^{-}[\text{rhsp: rmfp}]$$
$$+r_{39}[\text{rhsp}][\text{rmfp}]$$

$$d[\text{rhsp: rhsf}]/dt = r_{18}^{+}[\text{rhsp}][\text{rhsf}] - r_{18}^{-}[\text{rhsp: rhsf}] + r_{20}[\text{rhsp}][\text{rhsf}_2]$$
$$+r_{21}[\text{rhsp}][\text{rhsf}_2^{(1)}] + r_{24}[\text{rhsp}][\text{rhsf}_3] + r_{25}[\text{rhsp}][\text{rhsf}_3^{(1)}]$$
$$+r_{27}[\text{rhsp}][\text{rhsf}_3^{(2)}] + r_{30}[\text{rhsp}][\text{rhsf}_3\text{: rhse}]$$
$$+r_{32}[\text{rhsp}][\text{rhsf}_3^{(1)}\text{: rhse}] + r_{34}[\text{rhsp}][\text{rhsf}_3^{(2)}\text{: rhse}]$$

$$d[\text{rhsp: rhsf}^{(1)}]/dt = r_{19}^{+}[\text{rhsp}][\text{rhsf}^{(1)}] - r_{19}^{-}[\text{rhsp: rhsf}^{(1)}]$$
$$+r_{22}[\text{rhsp}][\text{rhsf}_2^{(1)}] + r_{23}[\text{rhsp}][\text{rhsf}_2^{(2)}]$$
$$+r_{26}[\text{rhsp}][\text{rhsf}_3^{(1)}] + r_{28}[\text{rhsp}][\text{rhsf}_3^{(2)}]$$
$$+r_{29}[\text{rhsp}][\text{rhsf}_3^{(3)}] + r_{31}[\text{rhsp}][\text{rhsf}_3^{(1)}\text{: rhse}]$$
$$+r_{33}[\text{rhsp}][\text{rhsf}_3^{(2)}\text{: rhse}] + r_{35}[\text{rhsp}][\text{rhsf}_3^{(3)}\text{: rhse}]$$

$$d[\text{rhsp: rmfp}]/dt = r_{38}^{+}[\text{rhsp}][\text{rmfp}] - (r_{38}^{-} + r_{39})[\text{rhsp: rmfp}]$$
$$d[\text{rmfp}]/dt = r_{37}[\text{rprot}] - r_{38}^{+}[\text{rhsp}][\text{rmfp}] + r_{38}^{-}[\text{rhsp: rmfp}]$$
$$d[\text{rprot}]/dt = -r_{37}[\text{rprot}] + r_{39}[\text{rhsp: rmfp}]$$

Table 5: The ODEs corresponding to Rhsf, Rhsf$_2$, Rhsf$_3$, Rhsf$_3$: Rhse, and Rhsp: Rhsf in the refined model

$$
\begin{aligned}
d[\text{Rhsf}]/dt = {}& -2(r_1^+[\text{rhsf}]^2 + r_2^+[\text{rhsf}][\text{rhsf}^{(1)}] + r_3{}^+[\text{rhsf}^{(1)}]^2) + 2(r_1^-[\text{rhsf}_2] \\
& + r_2^-[\text{rhsf}_2{}^{(1)}] + r_3{}^-[\text{rhsf}_2{}^{(2)}]) - (r_4{}^+[\text{rhsf}][\text{rhsf}_2] + r_6{}^+[\text{rhsf}][\text{rhsf}_2{}^{(1)}] \\
& + r_8{}^+[\text{rhsf}][\text{rhsf}_2{}^{(2)}] + r_5{}^+[\text{rhsf}^{(1)}][\text{rhsf}_2] + r_7{}^+[\text{rhsf}^{(1)}][\text{rhsf}_2{}^{(1)}] \\
& + r_9{}^+[\text{rhsf}^{(1)}][\text{rhsf}_2{}^{(2)}]) + (r_4{}^-[\text{rhsf}_3] + (r_5{}^- + r_6{}^-)[\text{rhsf}_3{}^{(1)}] \\
& + (r_7{}^- + r_8{}^-)[\text{rhsf}_3{}^{(2)}] + r_9{}^-[\text{rhsf}_3{}^{(3)}]) - [\text{rhsp}](r_{18}{}^+[\text{rhsf}] \\
& + r_{19}{}^+[\text{rhsf}^{(1)}]) + (r_{18}{}^-[\text{rhsp: rhsf}] + r_{19}{}^-[\text{rhsp: rhsf}^{(1)}]) \\
& + [\text{rhsp}](r_{20}[\text{rhsf}_2] + (r_{21} + r_{22})[\text{rhsf}_2{}^{(1)}] + r_{23}[\text{rhsf}_2{}^{(2)}]) \\
& + 2[\text{rhsp}](r_{24}[\text{rhsf}_3] + (r_{25} + r_{26})[\text{rhsf}_3{}^{(1)}] + (r_{27} + r_{28})[\text{rhsf}_3{}^{(2)}] \\
& + r_{29}[\text{rhsf}_3{}^{(3)}]) + 2[\text{rhsp}](r_{30}[\text{rhsf}_3: \text{rhse}] + (r_{31} + r_{32})[\text{rhsf}_3{}^{(1)}: \text{rhse}] \\
& + (r_{33} + r_{34})[\text{rhsf}_3{}^{(2)}: \text{rhse}] + r_{35}[\text{rhsf}_3{}^{(3)}: \text{rhse}]) \\[4pt]
d[\text{Rhsf}_2]/dt = {}& (r_1^+[\text{rhsf}]^2 + r_2^+[\text{rhsf}][\text{rhsf}^{(1)}] + r_3^+[\text{rhsf}^{(1)}]^2) - (r_1^-[\text{rhsf}_2] \\
& + r_2^-[\text{rhsf}_2{}^{(1)}] + r_3^-[\text{rhsf}_2{}^{(2)}]) - (r_4^+[\text{rhsf}][\text{rhsf}_2] + r_6^+[\text{rhsf}][\text{rhsf}_2{}^{(1)}] \\
& + r_8^+[\text{rhsf}][\text{rhsf}_2{}^{(2)}] + r_5^+[\text{rhsf}^{(1)}][\text{rhsf}_2] + r_7^+[\text{rhsf}^{(1)}][\text{rhsf}_2{}^{(1)}] \\
& + r_9^+[\text{rhsf}^{(1)}][\text{rhsf}_2{}^{(2)}]) + (r_4^-[\text{rhsf}_3] + (r_5^- + r_6^-)[\text{rhsf}_3{}^{(1)}]) \\
& + (r_7^- + r_8^-)[\text{rhsf}_3{}^{(2)}] + r_9^-[\text{rhsf}_3{}^{(3)}]) - [\text{rhsp}](r_{20}[\text{rhsf}_2] \\
& + (r_{21} + r_{22})[\text{rhsf}_2{}^{(1)}] + r_{23}[\text{rhsf}_2{}^{(2)}]) \\[4pt]
d[\text{Rhsf}_3]/dt = {}& (r_4^+[\text{rhsf}][\text{rhsf}_2] + r_6^+[\text{rhsf}][\text{rhsf}_2{}^{(1)}] + r_8^+[\text{rhsf}][\text{rhsf}_2{}^{(2)}] \\
& + r_5^+[\text{rhsf}^{(1)}][\text{rhsf}_2] + r_7^+[\text{rhsf}^{(1)}][\text{rhsf}_2{}^{(1)}] + r_9^+[\text{rhsf}^{(1)}][\text{rhsf}_2{}^{(2)}]) \\
& - (r_4^-[\text{rhsf}_3] + (r_5^- + r_6^-)[\text{rhsf}_3{}^{(1)}] + (r_7^- + r_8^-)[\text{rhsf}_3{}^{(2)}] + r_9^-[\text{rhsf}_3{}^{(3)}]) \\
& - [\text{rhse}](r_{10}^+[\text{rhsf}_3] + r_{11}^+[\text{rhsf}_3{}^{(1)}] + r_{12}^+[\text{rhsf}_3{}^{(2)}] + r_{13}^+[\text{rhsf}_3{}^{(3)}]) \\
& + (r_{10}^-[\text{rhsf}_3: \text{rhse}] + r_{11}^-[\text{rhsf}_3{}^{(1)}: \text{rhse}] + r_{12}^-[\text{rhsf}_3{}^{(2)}: \text{rhse}] \\
& + r_{13}^-[\text{rhsf}_3{}^{(3)}: \text{rhse}]) - [\text{rhsp}](r_{24}[\text{rhsf}_3] + (r_{25} + r_{26})[\text{rhsf}_3{}^{(1)}] \\
& + (r_{27} + r_{28})[\text{rhsf}_3{}^{(2)}] + r_{29}[\text{rhsf}_3{}^{(3)}]) \\[4pt]
d[\text{Rhsf}_3: \text{Rhse}]/dt = {}& [\text{rhse}](r_{10}^+[\text{rhsf}_3] + r_{11}^+[\text{rhsf}_3{}^{(1)}] + r_{12}^+[\text{rhsf}_3{}^{(2)}] \\
& + r_{13}^+[\text{rhsf}_3{}^{(3)}]) - (r_{10}^-[\text{rhsf}_3: \text{rhse}] + r_{11}^-[\text{rhsf}_3{}^{(1)}: \text{rhse}] + r_{12}^-[\text{rhsf}_3{}^{(2)}: \text{rhse}] \\
& + r_{13}^-[\text{rhsf}_3{}^{(3)}: \text{rhse}]) - [\text{rhsp}](r_{30}[\text{rhsf}_3: \text{rhse}] + (r_{31} + r_{32})[\text{rhsf}_3{}^{(1)}: \text{rhse}] \\
& + (r_{33} + r_{34})[\text{rhsf}_3{}^{(2)}: \text{rhse}] + r_{35}[\text{rhsf}_3{}^{(3)}: \text{rhse}]) \\[4pt]
d[\text{Rhsp}: \text{Rhsf}]/dt = {}& [\text{rhsp}](r_{18}^+[\text{rhsf}] + r_{19}{}^+[\text{rhsf}^{(1)}]) - (r_{18}^-[\text{rhsp: rhsf}] \\
& + r_{19}{}^-[\text{rhsp: rhsf}^{(1)}]) + [\text{rhsp}](r_{20}[\text{rhsf}_2] + (r_{21} + r_{22})[\text{rhsf}_2{}^{(1)}] \\
& + r_{23}[\text{rhsf}_2{}^{(2)}]) + [\text{rhsp}](r_{24}[\text{rhsf}_3] + (r_{25} + r_{26})[\text{rhsf}_3{}^{(1)}] \\
& + (r_{27} + r_{28})[\text{rhsf}_3{}^{(2)}] + r_{29}[\text{rhsf}_3{}^{(3)}]) \\
& + [\text{rhsp}](r_{30}[\text{rhsf}_3: \text{rhse}] + (r_{31} + r_{32})[\text{rhsf}_3{}^{(1)}: \text{rhse}] \\
& (r_{33} + r_{34})[\text{rhsf}_3{}^{(2)}: \text{rhse}] + r_{35}[\text{rhsf}_3{}^{(3)}: \text{rhse}])
\end{aligned}
$$

D The numerical setup of the refined model

Table 6: The numerical values of the parameters of the refined model

$$r_1^+ = k_1^+; \qquad r_8^+ = k_2^+; \qquad r_{16} = k_4; \qquad r_{28} = k_7/2;$$
$$r_1^- = k_1^-; \qquad r_8^- = k_2^-/2; \qquad r_{17} = k_4; \qquad r_{29} = k_7;$$
$$r_2^+ = 2 \cdot k_1^+; \qquad r_9^+ = k_2^+; \qquad r_{18}^{+} = k_5^+; \qquad r_{30} = k_8;$$
$$r_2^- = k_1^-; \qquad r_9^- = k_2^-; \qquad r_{18}^{-} = k_5^-; \qquad r_{31} = k_8/2;$$
$$r_3^+ = k_1^+; \qquad r_{10}^{+} = k_3^+; \qquad r_{19}^{+} = k_5^+; \qquad r_{32} = k_8/2;$$
$$r_3^- = k_1^-; \qquad r_{10}^{-} = k_3^-; \qquad r_{19}^{-} = k_5^-; \qquad r_{33} = k_8/2;$$
$$r_4^+ = k_2^+; \qquad r_{11}^{+} = k_3^+; \qquad r_{20} = k_6; \qquad r_{34} = k_8/2;$$
$$r_4^- = k_2^-; \qquad r_{11}^{-} = k_3^-; \qquad r_{21} = k_6/2; \qquad r_{35} = k_8;$$
$$r_5^+ = k_2^+; \qquad r_{12}^{+} = k_3^+; \qquad r_{22} = k_6/2; \qquad r_{36} = k_9;$$
$$r_5^- = k_2^-/2; \qquad r_{12}^{-} = k_3^-; \qquad r_{23} = k_6; \qquad r_{37} = \Phi_T;$$
$$r_6^+ = k_2^+; \qquad r_{13}^{+} = k_3^+; \qquad r_{24} = k_7; \qquad r_{38}^+ = k_{11}^{+};$$
$$r_6^- = k_2^-/2; \qquad r_{13}^{-} = k_3^-; \qquad r_{25} = k_7/2; \qquad r_{38}^- = k_{11}^{-};$$
$$r_7^+ = k_2^+; \qquad r_{14} = k_4; \qquad r_{26} = k_7/2; \qquad r_{39} = k_{12}$$
$$r_7^- = k_2^-/2; \qquad r_{15} = k_4; \qquad r_{27} = k_7/2;$$
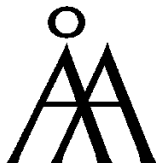
Turku
Centre *for*
Computer
Science

Joukahaisenkatu 3-5 B, 20520 Turku, Finland │ www.tucs.fi

University of Turku
Faculty of Mathematics and Natural Sciences
- Department of Information Technology
- Department of Mathematics
Turku School of Economics
- Institute of Information Systems Science

Åbo Akademi University
Division for Natural Sciences and Technology
- Department of Information Technologies