Åbo Akademi University

Large Scale Biological Models in Rodin

Usman Sanwal, Thai Son Hoang, Luigia Petre, and Ion Petre

Rodin Workshop

June 8, 2021

Modeling Biological Systems

- Modeling for biology is very complex problem
  - Many variables and reactions
  - Local interactions, emergent global behavior
  - Multi-scale data, phenomena
  - Combination of quantitative and qualitative data

- We recently used data refinement to model biological systems
  - Could not handle big models due to combinatorial explosion
  - Number of events increased significatly in different refinement steps

Our Current Approach

Åbo Akademi University

- We demonstrate how a particular modeling feature of Event-B – the common mathematical function – enables us to significantly reduce the concrete model's sizes.

- The relation between the abstract and the concrete forms of a reactant is captured with a function

- This enables us to model the concrete reactions more elegantly and concisely, and as a result, the total number of events in the refined model is reduced significantly.

Modeling metabolic networks in Rodin

- Metabolic networks can be seen as sets of biochemical reactions

- A reaction can be written as a rewriting rule of the form

$$r : m_1X_1 + m_2X_2 + \ldots + m_nX_n \rightarrow m'_1X_1 + m'_2X_2 + \ldots + m'_nX_n$$

- To build an Rodin model for a given metabolic network, we choose as the set of variables of the model that of the species of metabolic network

- Each reaction of the metabolic network will be modeled in Event-B through an event

The general form of a Rodin model corresponding to a metabolic network

VARIABLES $X_1, X_2, ..., X_n$
INVARIANTS
@inv1 $X_1 \in \mathbb{N}$
@inv2 $X_2 \in \mathbb{N}$
...
@invn $X_n \in \mathbb{N}$

INITIALISATION
@act1 $X_1 = init_1$
@act2 $X_2 = init_2$
...
@actn $X_n = init_n$

Event r
WHERE
@grd1 $X_1 \geq m_1$
@grd2 $X_2 \geq m_2$
...
@grdn $X_n \geq m_n$
THEN
@act1 $X_1 := X_1 + (m_1' - m_1)$
@act2 $X_2 := X_2 + (m_2' - m_2)$
...
@actn $X_n := X_n + (m_n' - m_n)$
END

Example: Binding Event

Åbo Akademi University

- For example, we have a reaction (bind) of type:

$$A + B \rightarrow AB$$

AbstractBind event
WHERE
  @grd1 $A \geq 1$
  @grd2 $B \geq 1$
THEN
  @act1 $A := A - 1$
  @act2 $B := B - 1$
  @act3 $AB := AB + 1$
END

Refinement: Binding Event

Åbo Akademi University

- Further assume the A reactant is to be refined into two special cases, A_0 and A_1.

gluing1: $A = A_0 + A_1$

gluing2: $AB = A_0B + A_1B$

Context...
Constants A_0, A_1, A_0B, A_1B
Sets A_SET, AB_SET
Axioms
partition$(A_SET, \{A_0\}, \{A_1\})$
partition$(AB_SET, \{A_0B\}, \{A_1B\})$

ScalableConcreteBind event
ANY e, i
WHERE
@grd1 $A\_FUNC(e) \geq 1$
@grd2 $B \geq 1$
@grd3 $(e = A_0 \wedge i = A_0B) \vee (e = A_1 \wedge i = A_1B)$
THEN
@act1 $A\_FUNC(e) := A\_FUNC(e) - 1$
@act2 $B := B - 1$
@act3 $AB\_FUNC(i) := AB\_FUNC(i) + 1$
END

Example: Dimer Event

Åbo Akademi University

- For example, We have a reaction (bind) of type:

$$A + A \rightarrow AA$$

AbstractDimer event
WHERE
  @grd1 $A \geq 2$
THEN
  @act1 $A := A - 2$
  @act3 $AA := AA + 1$
END

Refinement: Binding Event

- Further assume the A reactant is to be refined into two special cases, A_0 and A_1.

gluing1: $A = A_0 + A_1$

gluing2: $AA = AA_0 + AA_1 + AA_{01}$

Context...
Constants
$A_0, A_1, AA_0, AA_1, AA_{01}$
Sets
A_SET, AA_SET
Axioms
partition$(A_SET, \{A_0\}, \{A_1\})$
partition$(AA_SET, \{AA_0\}, \{AA_1\}, \{AA_{01}\})$

ScalableConcreteDimer event
ANY e_0, e_1, i
WHERE
@grd1 $e_0 \in \mathbb{N}$
@grd2 $e_1 \in \mathbb{N}$
@grd3 $e_0 + e_1 = 2$
@grd4 $A\_FUNC(A_0) \geq 1$
@grd5 $A\_FUNC(A_1) \geq 1$
@grd6 $e_0 = 2 \Rightarrow i = AA_0$
@grd7 $e_0 = 1 \Rightarrow i = AA_{01}$
@grd8 $e_0 = 0 \Rightarrow i = AA_1$
THEN
@act1 $A\_FUNC := A\_FUNC \Leftarrow \{A_0 \mapsto A\_FUNC(A_0) - e_0, A_1 \mapsto A\_FUNC(A_1) - e_1\}$
@act2 $AA\_FUNC(i) := AA\_FUNC(i) + 1$
END

Åbo Akademi University

Case studies

- We model two biological systems using refinement in Rodin:
  - Heat Shock Response Model
  - ErbB Signalling Pathway Model

Åbo Akademi University

Case studies

The heat shock response

- The heat shock response is a primordial defense mechanism against protein misfolding. It is a key cellular response against thermal (and other types of) stresses.
- We model all the reactions of molecular model for the heat shock response.
- We also introduce refinement for those reactions where heat shock factor is involved and model those reactions as well

Åbo Akademi University

The molecular model for the eukaryotic heat shock response

	Reaction	Description
(1)	$2\text{hsf} \rightleftharpoons \text{hsf}_2$	Dimerization
(2)	$\text{hsf} + \text{hsf}_2 \rightleftharpoons \text{hsf}_3$	Trimerization
(3)	$\text{hsf}_3 + \text{hse} \rightleftharpoons \text{hsf}_3 : \text{hse}$	DNA binding
(4)	$\text{hsf}_3 : \text{hse} \rightarrow \text{hsf}_3 : \text{hse} + \text{hsp}$	hsp synthesis
(5)	$\text{hsp} + \text{hsf} \rightleftarrows \text{hsp} : \text{hsf}$	hsf sequestration
(6)	$\text{hsp} + \text{hsf}_2 \rightarrow \text{hsp} : \text{hsf} + \text{hsf}$	Dimer dissipation
(7)	$\text{hsp} + \text{hsf}_3 \rightarrow \text{hsp} : \text{hsf} + 2\text{hsf}$	Trimer dissipation
(8)	$\text{hsp} + \text{hsf}_3 : \text{hse} \rightarrow \text{hsp} : \text{hsf} + 2\text{hsf} + \text{hse}$	DNA unbinding
(9)	$\text{hsp} \rightarrow \varnothing$	hsp degradation
(10)	$\text{prot} \rightarrow \text{mfp}$	Protein misfolding
(11)	$\text{hsp} + \text{mfp} \rightleftarrows \text{hsp} : \text{mfp}$	mfp sequestration
(12)	$\text{hsp} : \text{mfp} \rightarrow \text{hsp} + \text{prot}$	Protein refolding

Åbo Akademi University

HSF+HSF -> HSF_2

HSF_2 -> HSF+HSF

HSF+HSF_2 -> HSF_3

HSF_3 -> HSF+HSF_2

HSF_3+HSE -> HSF_3:HSE

HSF_3:HSE -> HSF_3+HSE

HSF_3:HSE -> HSF_3:HSE+HSP

HSP+HSF -> HSP:HSF

HSP:HSF -> HSP+HSF

HSP+HSF_2 -> HSP:HSF+HSF

HSP+HSF_3 -> HSP:HSF+2HSF

HSP+HSF_3:HSE ->HSP:HSF+2HSF+HSE

PROT -> MFP

HSP+MFP -> HSP:MFP

HSP:MFP -> HSP+MFP

HSP:MFP -> HSP+PROT

HSP → 0

HSFDimerForward
WHERE
  @grd1 $hsf \geq 2$
THEN
  @act1 $hsf := hsf - 2$
  @act2 $hsf_2 := hsf_2 + 1$
END

HSFDimerBackward
WHERE
  @grd1 $hsf_2 \geq 1$
THEN
  @act1 $hsf_2 := hsf_2 - 1$
  @act2 $hsf := hsf + 2$
END

HSFTrimerForward
WHERE
  @grd1 $hsf \geq 1$
  @grd2 $hsf_2 \geq 1$
THEN
  @act1 $hsf := hsf - 1$
  @act2 $hsf_2 := hsf_2 - 1$
  @act3 $hsf_3 := hsf_3 + 1$
END

HSFTrimerBackward
WHERE
  @grd1 $hsf_3 \geq 1$
THEN
  @act1 $hsf_3 := hsf_3 - 1$
  @act2 $hsf := hsf + 1$
  @act3 $hsf_2 := hsf_2 + 1$
END

DNABindingForward
WHERE
  @grd1 $hsf_3 \geq 1$
  @grd2 $hse \geq 1$
THEN
  @act1 $hsf_3 := hsf_3 - 1$
  @act2 $hse := hse - 1$
  @act3 $hsf_3 : hse := hsf_3 : hse + 1$
END

DNABindingBackward
WHERE
  @grd1 $hsf_3 : hse \geq 1$
THEN
  @act1 $hsf_3 : hse := hsf_3 : hse - 1$
  @act2 $hsf_3 := hsf_3 + 1$
  @act3 $hse := hse + 1$
END

HSFDimerForward
WHERE
  @grd1 $hsf \geq 2$
THEN
  @act1 $hsf := hsf - 2$
  @act2 $hsf_2 := hsf_2 + 1$
END

HSFDimerBackward
WHERE
  @grd1 $hsf_2 \geq 1$
THEN
  @act1 $hsf_2 := hsf_2 - 1$
  @act2 $hsf := hsf + 2$
END

HSFTrimerForward
WHERE
  @grd1 $hsf \geq 1$
  @grd2 $hsf_2 \geq 1$
THEN
  @act1 $hsf := hsf - 1$
  @act2 $hsf_2 := hsf_2 - 1$
  @act3 $hsf_3 := hsf_3 + 1$
END

HSFTrimerBackward
WHERE
  @grd1 $hsf_3 \geq 1$
THEN
  @act1 $hsf_3 := hsf_3 - 1$
  @act2 $hsf := hsf + 1$
  @act3 $hsf_2 := hsf_2 + 1$
END

DNABindingForward
WHERE
  @grd1 $hsf_3 \geq 1$
  @grd2 $hse \geq 1$
THEN
  @act1 $hsf_3 := hsf_3 - 1$
  @act2 $hse := hse - 1$
  @act3 $hsf_3 : hse := hsf_3 : hse + 1$
END

DNABindingBackward
WHERE
  @grd1 $hsf_3 : hse \geq 1$
THEN
  @act1 $hsf_3 : hse := hsf_3 : hse - 1$
  @act2 $hsf_3 := hsf_3 + 1$
  @act3 $hse := hse + 1$
END

HSFTrimerUnbindingForward
WHERE
  @grd1 $hsp \geq 1$
  @grd2 $hsf_3 : hse \geq 1$
THEN
  @act1 $hsp := hsp - 1$
  @act2 $hsf_3 : hse := hsf_3 : hse - 1$
  @act3 $hsp: hsf := hsp: hsf + 1$
  @act4 $hse := hse + 1$
  @act5 $hsf := hsf + 2$
END

ProteinMisfoldingForward
WHERE
  @grd1 $prot \geq 1$
THEN
  @act1 $prot := prot - 1$
  @act2 $mfp := mfp + 1$
END

ProteinChaperoningForward
WHERE
  @grd1 $hsp \geq 1$
  @grd2 $mfp \geq 1$
THEN
  @act1 $hsp := hsp - 1$
  @act2 $mfp := mfp - 1$
  @act3 $hsp: mfp := hsp: mfp + 1$
END

ProteinChaperoningBackward
WHERE
  @grd1 $hsp: mfp \geq 1$
THEN
  @act1 $hsp: mfp := hsp: mfp - 1$
  @act2 $hsp := hsp + 1$
  @act3 $mfp := mfp + 1$
END

ProteinRefoldingForward
WHERE
  @grd1 $hsp: mfp \geq 1$
THEN
  @act1 $hsp: mfp := hsp: mfp - 1$
  @act2 $hsp := hsp + 1$
  @act3 $prot := prot + 1$
END

Refinement for HSR Model

- Once *hsr* model is build, we refined all compounds comprising *hsf*

- We perform the refinement counting the number of *hsf* constituents of the complex

$$
\begin{aligned}
&\text{(i)}\ \ \mathsf{hsf} \rightarrow \{\mathsf{rhsf}^{(0)}, \mathsf{rhsf}^{(1)}\};\\
&\text{(ii)}\ \ \mathsf{hsf}_2 \rightarrow \{\mathsf{rhsf}_2^{(0)}, \mathsf{rhsf}_2^{(1)}, \mathsf{rhsf}_2^{(2)}\};\\
&\text{(iii)}\ \ \mathsf{hsf}_3 \rightarrow \{\mathsf{rhsf}_3^{(0)}, \mathsf{rhsf}_3^{(1)}, \mathsf{rhsf}_3^{(2)}, \mathsf{rhsf}_3^{(3)}\};\\
&\text{(iv)}\ \ \mathsf{hsf}_3\mathsf{:hse} \rightarrow \{\mathsf{rhsf}_3^{(0)}\mathsf{:hse}, \mathsf{rhsf}_3^{(1)}\mathsf{:hse}, \mathsf{rhsf}_3^{(2)}\mathsf{:hse}, \mathsf{rhsf}_3^{(3)}\mathsf{:hse}\};\\
&\text{(v)}\ \ \mathsf{hsp:hsf} \rightarrow \{\mathsf{hsp:rhsf}^{(0)}, \mathsf{hsp:rhsf}^{(1)}\}.
\end{aligned}
$$

- We completed the refinement in Event-B through 5 successive refinements: for hsf, then for hsf_2, then for hsf_3, then for hsf_3:hse, and finally for hsp:hsf

Old Refinement Approach
hsf refinement

Åbo Akademi University

Initial Reaction

HSFDimerForward
WHERE
@grd1 hsf ≥ 2
THEN
@act1 hsf := hsf -2
@act2 $hsf_2 := hsf_2 + 1$
END

$$gluing\colon\ rhsf^{(0)} + rhsf^{(1)} = hsf$$

Refined Reaction

Dimerization Forward-i.1
WHERE
@grd1 $rhsf^{(0)} \geq 2$
THEN
@act1 $rhsf^{(0)} := rhsf^{(0)} - 2$
@act2 $hsf_2 := hsf_2 + 1$
END

Dimerization Forward-i.2
WHERE
@grd1 $rhsf^{(0)} \geq 1 \wedge rhsf^{(1)} \geq 1$
THEN
@act1 $rhsf^{(0)} := rhsf^{(0)} - 1$
@act2 $rhsf^{(1)} := rhsf^{(1)} - 1$
@act3 $hsf_2 := hsf_2 + 1$
END

Dimerization Forward-i.3
WHERE
@grd1 $rhsf^{(1)} \geq 2$
THEN
@act1 $rhsf^{(1)} := rhsf^{(1)} - 2$
@act2 $hsf_2 := hsf_2 + 1$
END

After the first refinement, the model got from 17 events to 30 events

Old Refinement Approach: hsf_2 refinement

Dimerization Forward-i.1
WHERE
@grd1 $rhsf^{(0)} \geq 2$
THEN
@act1 $rhsf^{(0)} := rhsf^{(0)} - 2$
@act2 $hsf_2 := hsf_2 + 1$
END

Dimerization Forward-i.2
WHERE
@grd1 $rhsf^{(0)} \geq 1 \wedge rhsf^{(1)} \geq 1$
THEN
@act1 $rhsf^{(0)} := rhsf^{(0)} - 1$
@act2 $rhsf^{(1)} := rhsf^{(1)} - 1$
@act3 $hsf_2 := hsf_2 + 1$
END

Dimerization Forward-i.3
WHERE
@grd1 $rhsf^{(1)} \geq 2$
THEN
@act1 $rhsf^{(1)} := rhsf^{(1)} - 2$
@act2 $hsf_2 := hsf_2 + 1$
END

$$gluing\colon\ rhsf_2^{(0)} + rhsf_2^{(1)} + rhsf_2^{(2)} = hsf_2$$

We continued similarly with all the other events and the other 3 refinements and ended up with 57 events in the final refined model.

Dimerization Forward-ii.1
WHERE
@grd1 $rhsf^{(0)} \geq 2$
THEN
@act1 $rhsf^{(0)} := rhsf^{(0)} - 2$
@act2 $rhsf_2^{(0)} := rhsf_2^{(0)} + 1$
END

Dimerization Forward-ii.2
WHERE
@grd1 $rhsf^{(0)} \geq 1 \wedge rhsf^{(1)} \geq 1$
THEN
@act1 $rhsf^{(0)} := rhsf^{(0)} - 1$
@act2 $rhsf^{(1)} := rhsf^{(1)} - 1$
@act3 $rhsf_2^{(1)} := rhsf_2^{(1)} + 1$
END

Dimerization Forward-ii.3
WHERE
@grd1 $rhsf^{(1)} \geq 2$
THEN
@act1 $rhsf^{(1)} := rhsf^{(1)} - 2$
@act2 $rhsf_2^{(2)} := rhsf_2^{(2)} + 1$
END

Current Aproach

Åbo Akademi University

Initial Reaction

HSFDimerForward
WHERE
@grd1 $\text{hsf} \geq 2$
THEN
@act1 $\text{hsf} := \text{hsf} - 2$
@act2 $\text{hsf}_2 := \text{hsf}_2 + 1$
END

Refined Reaction

Dimerization Refinement
ANY e_0, e_1, i
WHERE
@grd1 $e_0 \in \mathbb{N}$
@grd2 $e_1 \in \mathbb{N}$
@grd3 $e_0 + e_1 = 2$
@grd4 $\text{rhsf}(\text{HSF\_0}) \geq 1$
@grd5 $\text{rhsf}(\text{HSF\_1}) \geq 1$
@grd6 $e_0 = 2 \Rightarrow i = \text{HSF2\_0}$
@grd7 $e_0 = 1 \Rightarrow i = \text{HSF2\_1}$
@grd8 $e_0 = 0 \Rightarrow i = \text{HSF2\_2}$
THEN
@act1 $\text{rhsf} := \text{rhsf} \Leftarrow \{\text{HSF\_0} \mapsto \text{rhsf}(\text{HSF\_0}) - e_0, \text{HSF\_1} \mapsto \text{rhsf}(\text{HSF\_1}) - e_1\}$
@act2 $\text{rhsf2}(i) := \text{rhsf2}(i) + 1$
END

$$gluing\colon\ rhsf^{(0)} + rhsf^{(1)} = hsf$$
$$gluing\colon\ rhsf_2^{(0)} + rhsf_2^{(1)} + rhsf_2^{(2)} = hsf_2$$

Model Statistics

- For hsr model, number of events were reduced to 17 events from 57 events in the final model.

- For the ErbB model, number of events were reduced to 242 events from 1320 events in the final model.

Åbo Akademi University

Conclusions

- In this paper, we propose a modeling method that uses common mathematical concept of function to avoid combinatorial explosion.
- Our proposal is to go more abstract (‘higher-level’) and replace each species to be refined by a function defined on the constant set of all the subspecies.
- This simple artifact lets us express almost all the complexity in the event guards, where we can have many cases and combinations of parameters.