



<http://researchspace.auckland.ac.nz>

ResearchSpace@Auckland

Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

To request permissions please use the Feedback form on our webpage.

<http://researchspace.auckland.ac.nz/feedback>

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the [Library Thesis Consent Form](#) and [Deposit Licence](#).

Note : Masters Theses

The digital copy of a masters thesis is as submitted for examination and contains no corrections. The print copy, usually available in the University Library, may contain alterations requested by the supervisor.

A FRAMEWORK FOR ANNOTATING AND VISUALIZING CELLML MODELS

SARALA M. WIMALARATNE

Supervised by: Dr. Matt D. B. Halstead and Associate Professor Poul F. Nielsen

*A thesis submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Biomedical Engineering
The University of Auckland, New Zealand
May 2009*

To Sajith, mom, dad, and sister for their unconditional love

ABSTRACT

The Physiome Project was established to develop tools for international collaboration and sharing physiological knowledge in the form of biological models and experimental data. The CellML language was developed in response to the need for a high-level language to represent and exchange mathematical models of biological processes.

The language provides a flexible framework for describing the dynamics of biological processes but does not explicitly lend itself to capturing the underlying biological concepts such as the entities and processes that these models represent. The relationship between the biological process and the mathematical model describing the biological process is also often complex. This makes it difficult to see the biological concepts which the CellML structures represent. A framework which supports visualizing the biological concepts and its relationship to the underlying CellML model would provide a very useful toolset for understanding the biological concepts modeled in CellML.

The CellML models need to be annotated with biological concepts in order to provide the machine interpretable data for generating a visual representation. We have developed an ontological framework which can be used to explicitly annotate CellML models with physical and biological concepts, a method to derive a simplified biological view from the annotations, a visual language for representing all biophysical processes captured in the CellML models, and a method to map the visual language to the ontological framework in order to automate the generation of visual representations of a model.

The proposed method of model visualization produces a result that is dependent on the structure of the CellML models which requires modelers to structure the model in a way that best describes the biophysical concepts and abstractions they wish to demonstrate. Our argument is that this leads to a best practice approach to building and organizing models.

As a part of this research, a software tool for visualizing CellML models was developed. This tool combines the visual language and the ontologies to generate visualizations that depict the physical and biological concepts captured in CellML models and enables different communities in diverse disciplines to more easily understand CellML models within the biological domain they represent.

As research continues, with further improvement to the framework it would be possible to visually construct composite CellML models by selecting high level biological concepts.

ACKNOWLEDGEMENT

I express my sincere gratitude towards my supervisors Dr Matt Halstead and Associate Professor Poul Nielsen for their enormous efforts giving guidance and encouragement throughout the course of this project.

I wish to gratefully acknowledge, Professor Peter Hunter, Dr Edmund Crampin, Associate Professor Gill Dobbie, Dr Catherine Lloyd, Dr Mike Cooling, and Dr Vijayaraghavan Rajagopal for their advice and assistance.

I am also thankful to a number of international researchers for fruitful discussions and stimulating suggestions on several aspects of this work; Dr Hiroaki Kitano, Dr Nicolas Le Novère, Dr Michael Hucka, Dr Emek Demir, Ms Yukiko Matsuoka, and Dr Akira Funahashi.

I am grateful for financial assistance from the University of Auckland in the form of a doctoral scholarship. Also to the Education New Zealand Study Abroad Award to Japan and a number of conference and travel grants. Without these financial supports this work would have been impossible.

I wish to thank my fellow postgraduate students and administrative staff at the Auckland Bioengineering Institute for providing a friendly atmosphere.

Finally thanks to my in-laws for their support and encouragement during this study.

TABLE OF CONTENT

ABSTRACT	II
ACKNOWLEDGEMENT	III
TABLE OF CONTENT	IV
LIST OF FIGURES	VII
LIST OF TABLES.....	IX
ABBREVIATIONS	X
1 INTRODUCTION	1
1.1 Modeling languages for describing mathematical models of biology	2
1.1.1 Markup languages for describing mathematical models of biology	2
1.1.1.1 CellML.....	3
1.1.1.2 Systems biology markup language (SBML)	4
1.1.2 Ontological languages for describing mathematical models of biology	5
1.2 Identifying the problem in detail.....	6
1.2.1 Prototype development.....	7
1.3 Journal publications from this thesis	13
2 GUIDELINES FOR STRUCTURING CELLML MODELS.....	15
2.1 Introduction	15
2.2 Methods	19
2.2.1 Identification and representation of biophysical concepts and common mathematical constructs.....	21
2.2.2 Reconstruction of the original biological concepts by combining the components, providing model-specific values and using generic components	23
2.2.3 Use of encapsulation to partition the details of a model into a hierarchy of components	24
2.3 Results.....	25
2.3.1 Modularization of the G protein-coupled receptor (GPCR) cycle.....	25
2.3.2 Modularization of the Hodgkin–Huxley model.....	30
2.3.3 Modularization of the Noble model.....	34
2.4 Discussion	36
3 BIOPHYSICAL ANNOTATION AND REPRESENTATION OF CELLML MODELS	40

3.1	Introduction	40
3.2	Methods	44
3.2.1	Transformation of a CellML/XML model into a CellML/OWL model	45
3.2.2	Annotation of a CellML/OWL model to a CellMLBiophysical/OWL model ..	48
3.2.2.1	Annotating physical information	49
3.2.2.2	Annotating biological information.....	51
3.2.3	Simplification of a CellMLBiophysical/OWL model.....	53
3.3	Results.....	55
3.3.1	Representing the K_Ionic_Flow model in CellML	55
3.3.2	Translating the K_Ionic_Flow CellML model into a CellML/OWL model	56
3.3.3	Annotating the K_Ionic_Flow CellML/OWL instances to CellMLBiophysical/OWL instances	56
3.3.4	Simplifying the K_Ionic_Flow CellMLBiophysical/OWL model to show the biological view	57
3.4	Discussion	62
4	A METHOD FOR VISUALIZING CELLML MODELS	66
4.1	Introduction	66
4.2	Methods	70
4.2.1	The development of a standardized visual language for representing the physical and biological processes captured in the CellML models	71
4.2.1.1	A visual language for representing physical concepts	71
4.2.1.2	A visual language for representing the biological concepts	72
4.2.2	The representation of the visual language in computer readable form.....	74
4.2.3	The mapping of the visual language to the CellMLBiophysical/OWL ontology..	76
4.2.4	The development of an algorithm for generating visualizations of CellML models	78
4.2.4.1	Generating visualizations of the physical concepts in a CellML model	78
4.2.4.2	Generating visualizations of the biological concepts in a CellML model.....	79
4.3	Results.....	81
4.4	Discussion	86
5	A SOFTWARE TOOL FOR VISUALIZING CELLML MODELS	90
5.1	Introduction	90

5.2	Implementation.....	92
5.2.1	Application of the CellMLViewer	93
5.2.1.1	Visualizing CellML models without annotated biophysical data	95
5.2.1.2	Annotating CellML models with biophysical concepts and visual language data	96
5.2.1.3	Visualizing the annotated CellML models capturing the underlying physical and biological concepts	97
5.2.1.4	Layout the diagrams to illustrate the sequence of biological interactions.....	98
5.2.1.5	Storing visual representations.....	98
5.2.2	System architecture	99
5.3	Discussion	100
5.4	Availability and requirements.....	102
6	CONCLUSIONS.....	103
	REFERENCES	108

LIST OF FIGURES

Figure 1.1: A signaling cascade regulating L-type calcium channel	6
Figure 1.2: Structuring and storing biological knowledge using XML	8
Figure 1.3: Notation for visualizing the underlying biology modeled in CellML.....	9
Figure 1.4: Glyphs represented in SVG	10
Figure 1.5: Sequence diagram illustrating the functionality supported by the prototype tool	11
Figure 1.6: Visualization of cAMP/PKA cascade regulating L-type calcium channel generated from the prototype tool.....	12
Figure 2.1: CellML structure	17
Figure 2.2: The three modularizing steps applied to an example, formation of receptor– ligand–G protein phosphorylated complex	20
Figure 2.3: Schematic diagram of the GPCR pathway	26
Figure 2.4: The set of CellML models developed to describe the GPCR cycle	27
Figure 2.5: The GPCR cycle modeled in CellML	29
Figure 2.6: A set of models describing several electrophysiological concepts	31
Figure 2.7: Description of the Hodgkin–Huxley model in CellML.....	33
Figure 2.8: Describing the Noble model in CellML	35
Figure 3.1: Modeling the formation of cyclic adenosine monophosphate (cAMP) in BioPAX	43
Figure 3.2: CellML structures.....	46
Figure 3.3: Example of a metadata definition	48
Figure 3.4: CellMLBiophysical/OWL ontology top-level class structure	49
Figure 3.5: A physical-entity graph generated for a CellML model.....	50
Figure 3.6: The mapping between Physical instances and Biological instances	52
Figure 3.7: Application of the reducing rules	54
Figure 3.8: Modeling the potassium ionic current described in Hodgkin-Huxley model.....	56
Figure 3.9: The K_Ionic_Flow model.....	57
Figure 3.10: Annotating the imported K_Channel shown in Figure 3.9.....	59
Figure 3.11: Annotated Hodgkin-Huxley model	61
Figure 4.1: Schematic diagram illustrating a CellMLBiophysical/OWL model	70
Figure 4.2: A notation for visualizing physical concepts	72
Figure 4.3: A notation for representing biological concepts	73

Figure 4.4: Template codes describing the visual language	75
Figure 4.5: Class structure of the VisualTemplate/OWL ontology	77
Figure 4.6: Mappings for generating the physical view	79
Figure 4.7: Mappings for generating the biological view	80
Figure 4.8: Visualizing the K_Ionic_Flow model	83
Figure 4.9: Physical view generated for the Hodgkin–Huxley model.....	84
Figure 4.10: Biological view generated for the Hodgkin-Huxley model.....	85
Figure 5.1: Sequence diagram for user activity flow for visualizing and annotating CellML models	94
Figure 5.2: A visualization generated for a CellML model without annotations	95
Figure 5.3: A CellMLBiophysical/OWL model loaded in Protégé	96
Figure 5.4: Biological view generated from an annotated CellMLBiophysical/OWL model	98
Figure 5.5: Schematic diagram of the system architecture underlying the model visualization tool CellMLViewer	99
Figure 5.6: Model associations	100

LIST OF TABLES

Table 4.1: Relationship between the GenericNode instances and the glyph types used in the K_Ionic_Flow model	82
Table 4.2: Relationship between the connections and the glyph types used in the K_Ionic_Flow model	82
Table 4.3: Relationship between the GenericNode instances and the glyph types used in the Hodgkin-Huxley model.....	85
Table 4.4: Relationship between the connections and the glyph types used in the Hodgkin-Huxley model.....	85

ABBREVIATIONS

AC	Adenylyl cyclase
ATP	Adenosine triphosphate
BioPAX	Biological pathways exchange
cAMP	Cyclic adenosine monophosphate
COPASI	Complex pathway simulator
COR	Cellular open resource
CSML	Cell system markup language
CSO	Cell system ontology
DOM	Document object model
FMA	Foundational model of anatomy ontology
Gd	G Protein with attached guanosine diphosphate
Gi	Inhibitory G protein
GO	Gene ontology
GPCR	G protein-coupled receptor
Gs	Stimulatory G protein
Gt	G protein with attached guanosine triphosphate
GUI	Graphical user interface
IP3	Inositol 1,4,5-trisphosphate
K	Potassium
L	Ligand
MathML	Mathematical markup language
MATLAB	Matrix laboratory
MIM	Molecular interaction maps
Na	Sodium
OBO	Open biological ontology
OWL	Web ontology language
OWL-DL	Web ontology language – descriptive logic
PATIKA	Pathway analysis tools for integration and knowledge acquisition
PCEnv	Physiome CellML environment
PKA	Protein kinase A
PMR	Physiome model repository

R	Receptor
Rg	Receptor-G protein complex
Rl	Receptor-ligand complex
Rlg	Receptor–ligand–G protein complex
Rlgp	Receptor–ligand–G protein phosphorylated complex
RDF	Resource description framework
RDFS	Resource description framework schema
SBML	Systems biology markup language
SBO	Systems biology ontology
SVG	Scalable vector graphics
XLink	XML linking language
XML	Extensible markup language