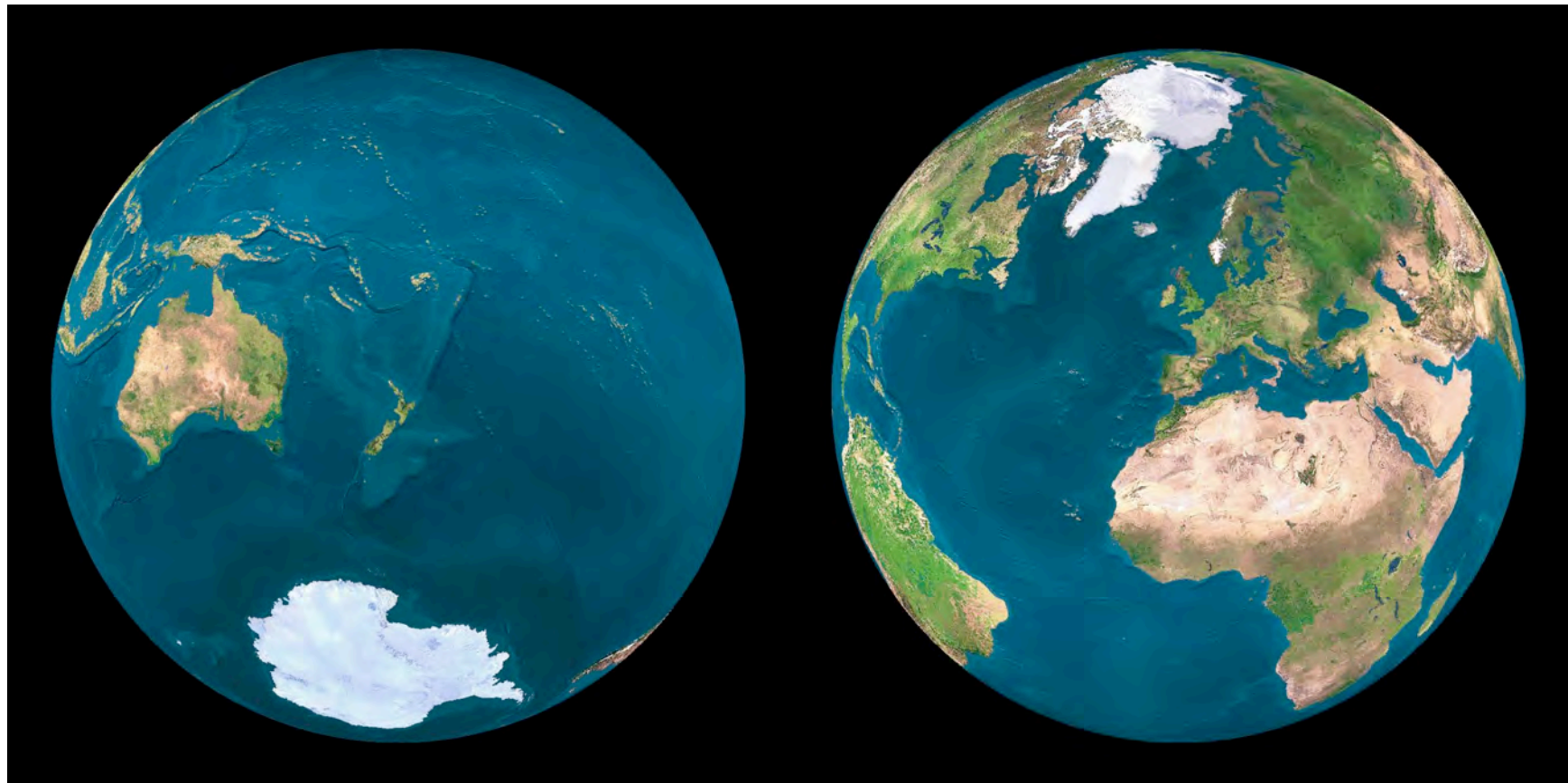


CellML SBGN SBO BioPAX MIASE Workshop 2009 Waiheke Island

Welcome



Haere mai



Organisers

- Catherine Lloyd
- Mike Hucka
- Nicolas Le Novère
- Peter Hunter
- Poul Nielsen



Sponsors

- Auckland Bioengineering Institute
- Maurice Wilkins Centre
- Virtual Physiological Human Network of Excellence
- IBM
- NIH



Programme

- Sunday
 - CellML: where it's been and where it's going
 - Events, time delays, and typing
 - Space and hierarchy
- Monday
 - Model exchange and interoperability
 - Minimum information about a simulation experiment
 - Encoding simulation description
- Tuesday
 - Ontological representation and visualisation
- Wednesday
 - Interoperability and integration
- Thursday



CellML...

- Designed to support the definition and sharing of models of biological processes.
- Intended to provide consistency in the mathematical representation.
- Encourages model evolution and reuse.
- Provides a representation in a form (XML) that is both human and computer readable.
- Started 1999, around same time as SBML.
- CellML and SBML have different emphases:
 - “SBML is designed for representing models of biochemical reaction networks”. (<http://www.sbml.org/>)
 - “The purpose of CellML is to store and exchange computer-based mathematical models”. (<http://www.cellml.org/>)



...CellML

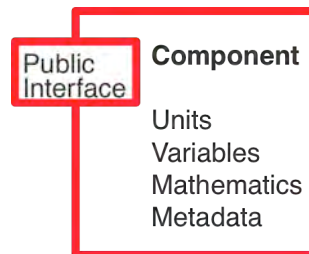
- CellML includes information about:
 - Model structure (how the parts of a model are organizationally related to one another);
 - Mathematics (equations describing the underlying biological processes);
 - Metadata (additional information about the model that allows scientists to search for specific models or model components in a database or other repository).
- CellML includes mathematics and metadata by leveraging existing XML-based languages, such as Content MathML, XML Linking Language (XLink), and Resource Description Framework (RDF).

(C. M. Lloyd, M. D. B. Halstead, and P. F. Nielsen, "*CellML: its future, present and past*" *Progress in Biophysics & Molecular Biology*, vol. 85, pp. 433-450, June-July 2004)



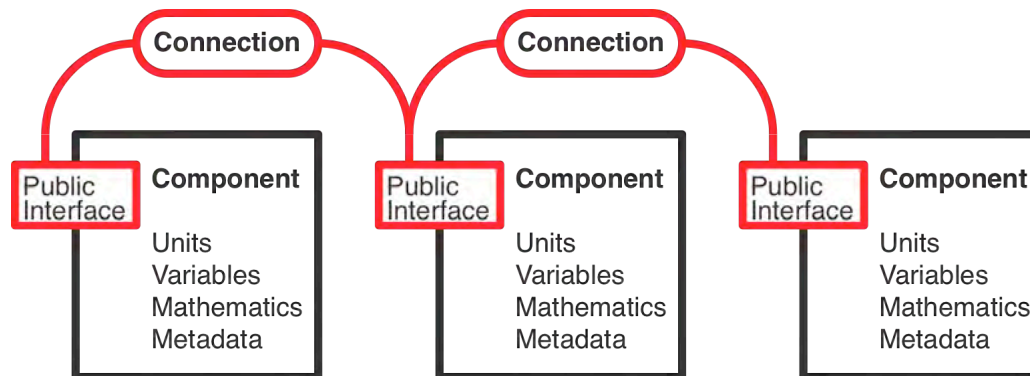
CellML components

- CellML has a simple structure based upon connected *components*.
- Components abstract concepts by providing well-defined interfaces to other components.
- Components encapsulate concepts by hiding details from other components.



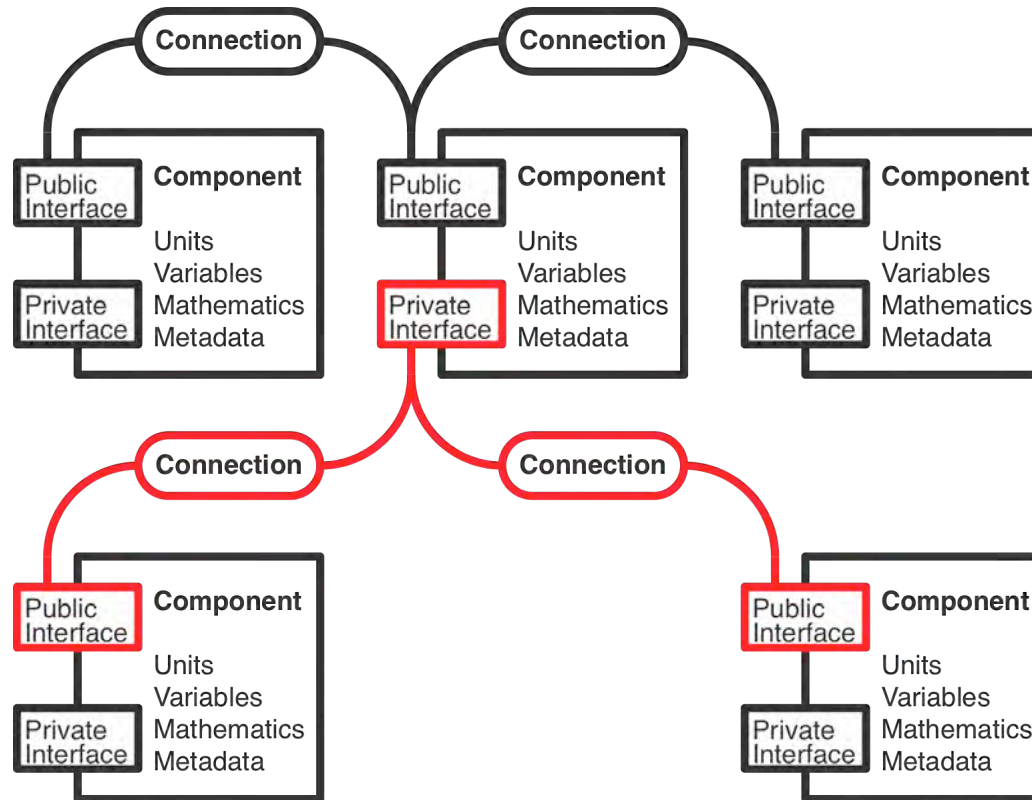
CellML connections

- *Connections* provide the means for sharing information by associating variables visible in the interface of one component with those in the interface of another component.
- Consistency is enforced by requiring that all variables be assigned appropriate physical *units*.



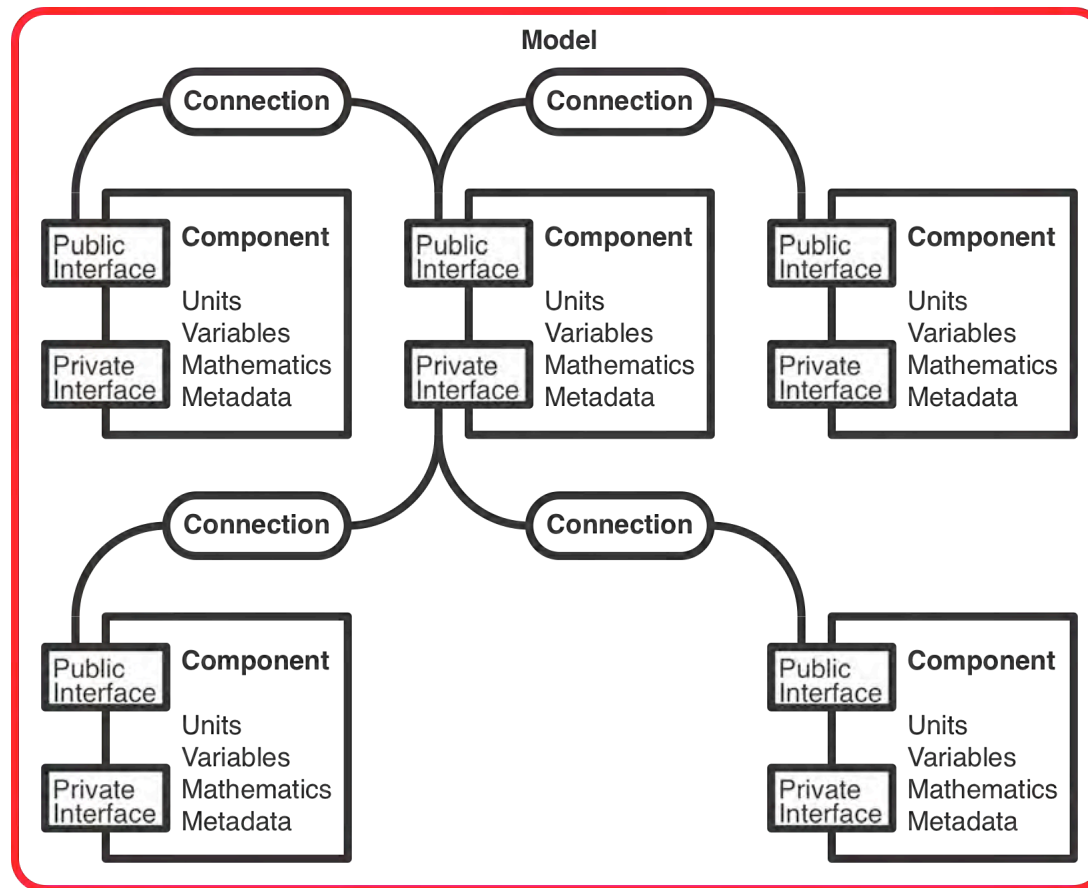
CellML encapsulation

- Encapsulation hierarchies are enabled using *private interfaces*.



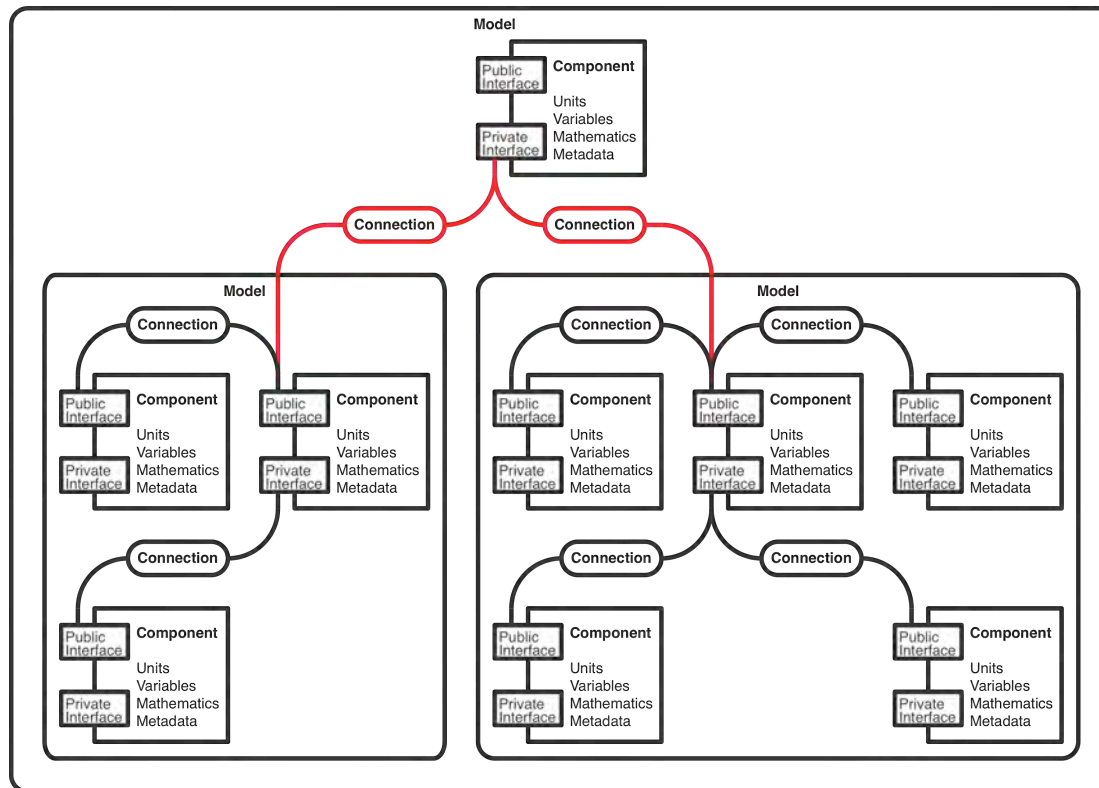
CellML model

- A *model* is the root element for a CellML document. It is a container for components, connections, units, and metadata.



CellML import

- Model reuse is enabled by the *import* element.
- New models may thus be constructed by combining existing models into model hierarchies.



CellML model repository

- CellML model repository has over 350 published models of:
 - Signal transduction pathways;
 - Metabolic pathways;
 - Electrophysiological;
 - Calcium dynamics;
 - Endocrine;
 - Cell cycle;
 - Smooth and skeletal muscle models;
 - Mechanical constitutive relationships...
- The BioModels database is a similar repository to support models created by the SBML community.
<http://www.biomodels.net/>
- **<http://www.cellml.org/models/>**

(C. M. Lloyd, J. R. Lawson, P. J. Hunter, and P. F. Nielsen, "*The CellML model repository*" *Bioinformatics*, 2008)



Model curation

- Actively curating models in the repository.
- Currently tag models with an attribute indicating level of curation:
 - Level 0 (☆☆☆) the model reflects the published version but not curated;
 - Level 1 (★☆☆) the model loads and runs in the specified simulation environment;
 - Level 2 (★★☆) the model produces results that are qualitatively similar to those previously published for the model;
 - Level 3 (★★★) the model has been quantitatively and rigorously verified as producing identical results to the original published model.
- Curation workflow being created for CellML 1.1 repository.
- Working towards MIRIAM compliance for all models.

(N. Le Novère, A. Finney, M. Hucka, U. S. Bhalla, F. Campagne, J. Collado-Vides, E. J. Crampin, M. Halstead, E. Klipp, P. Mendes, P. Nielsen, H. Sauro, B. Shapiro, J. L. Snoep, H. D. Spence, and B. L. Wanner, "**Minimum information requested in the annotation of biochemical models (MIRIAM)**" *Nature Biotechnology*, vol. 23, pp. 1509-1515, December 2005)



Pandit *et al.* model of cardiac action potential

- Pandit S.V., Clark R.B., Giles W.R., *et al.* A mathematical model of action potential heterogeneity in adult rat left ventricular myocytes. *Biophysical Journal* 2001;81(6):3029-51.
- www.cellml.org/models/pandit_clark_giles_demir_2001_version11

The screenshot shows the CellML model page for the Pandit et al. model. At the top, there are tabs for 'overview', 'edit', 'view math', 'model metadata', 'curation', 'view cellml', 'data', and 'procedural code'. The title is 'A Mathematical Model of Action Potential Heterogeneity in Adult Rat Left Ventricular Myocytes'. Below the title, there is a 'Download Model (151Kb)' button and a 'Solve model in: (help)' link. To the right of these are three star ratings: 'PCEnv' (3 stars), 'JSim' (4 stars), and 'COR' (3 stars). Below the ratings, there is a 'Curation Status' section with a star rating and a link to 'PCEnv Session (What's this?)'. The 'Model Documentation' section is divided into 'Model Status' and 'Model Structure'. The 'Model Status' section states that this CellML version represents the epicardial cell and that a number of inconsistencies in units and errors in equations from version 07 were fixed in this version. The 'Model Structure' section describes the model's development and its basis on the classical formulation of Hodgkin and Huxley.

overview edit view math model metadata curation view cellml data procedural code

A Mathematical Model of Action Potential Heterogeneity in Adult Rat Left Ventricular Myocytes

[Download Model \(151Kb\)](#) Solve model in: [\(help\)](#)

Curation Status: ★☆☆☆☆

PCEnv ★☆☆☆☆ JSim ★★★★★ COR ★☆☆☆☆

[PCEnv Session \(What's this?\)](#)

Model Documentation

Model Status

This CellML version of the model represents the epicardial cell. A number of inconsistencies in units and errors in equations from version 07 were fixed in this version. In addition to the formulation given by the paper and the author's later Corrections document, the Istim current has been adjusted to produce 1 Hz stimulations for 10 seconds.

Model Structure

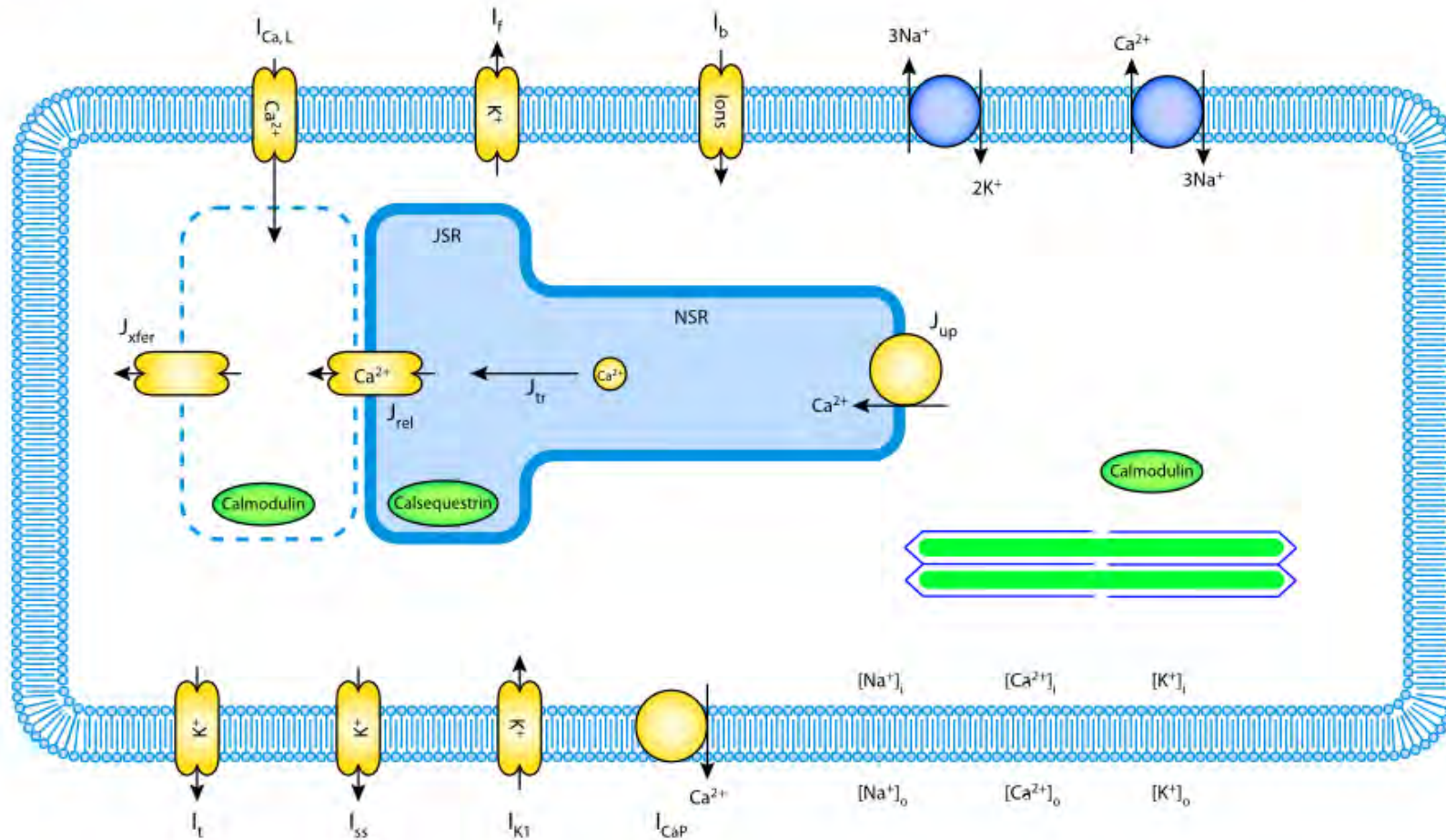
Over the past decade electrophysiological studies have revealed transmural heterogeneity, or differences in the action potential waveforms recorded in cells isolated from the epicardial and the endocardial tissues in the left ventricles of mammalian hearts.

The adult rat has been widely used as an experimental model to investigate the electrical heterogeneity in the left ventricle under normal conditions and pathophysiological states. From this biophysical, experimental data, derived from patch clamp experiments, Sandeep V. Pandit, Robert B. Clark, Wayne R. Giles and Semahat S. Demir have developed a mathematical model of action potential heterogeneity in adult rat left ventricular myocytes (see the figure below). The mathematical models for the epicardial and endocardial cells of the rat left ventricle are based on the classical formulation of Hodgkin and Huxley (please see the CellML version of The Hodgkin-Huxley Squid Axon Model, 1952 for more details), and are therefore similar to previous computational work carried out by this research group (see Demir *et al.* Sinoatrial Node Model 1994 and Demir *et al.* Sinoatrial Node Model 1999). The endocardial cell model is based on the epicardial formulation with only slight modifications in certain parameters and equations.



Pandit *et al.* model of cardiac action potential

- CellML model repository has article, math, metadata, code, ...



Hinch *et al.* model of Ca-induced Ca release

- Hinch R., Greenstein J.R., Tanskanen A.J., *et al.* A simplified local control model of Ca-induced Ca release in cardiac ventricular myocytes. *Biophysical Journal* 2004;87:3723-3736.
- www.cellml.org/models/hinch_greenstein_tanskanen_xu_winslow_2004_version02

[overview](#) [edit](#) [view math](#) [model metadata](#) [curation](#) [view cellml](#) [data](#) [procedural code](#)

A Simplified Local Control Model of Calcium-Induced Calcium Release in Cardiac Ventricular Myocytes

[Download Model](#) (113Kb)

Solve model in: [\(help\)](#)

Curation Status: ★★☆☆

PCEnv ★★☆☆

PCEnv Session (What's this?)

JSim ★★☆☆

COR ★★☆☆

Model Documentation

Model Status

This model is known to run in PCEnv and COR to reproduce the output shown in the publication. A PCEnv session file is also associated with this model.

Model Structure

This CellML model was based on the December 2004 paper:

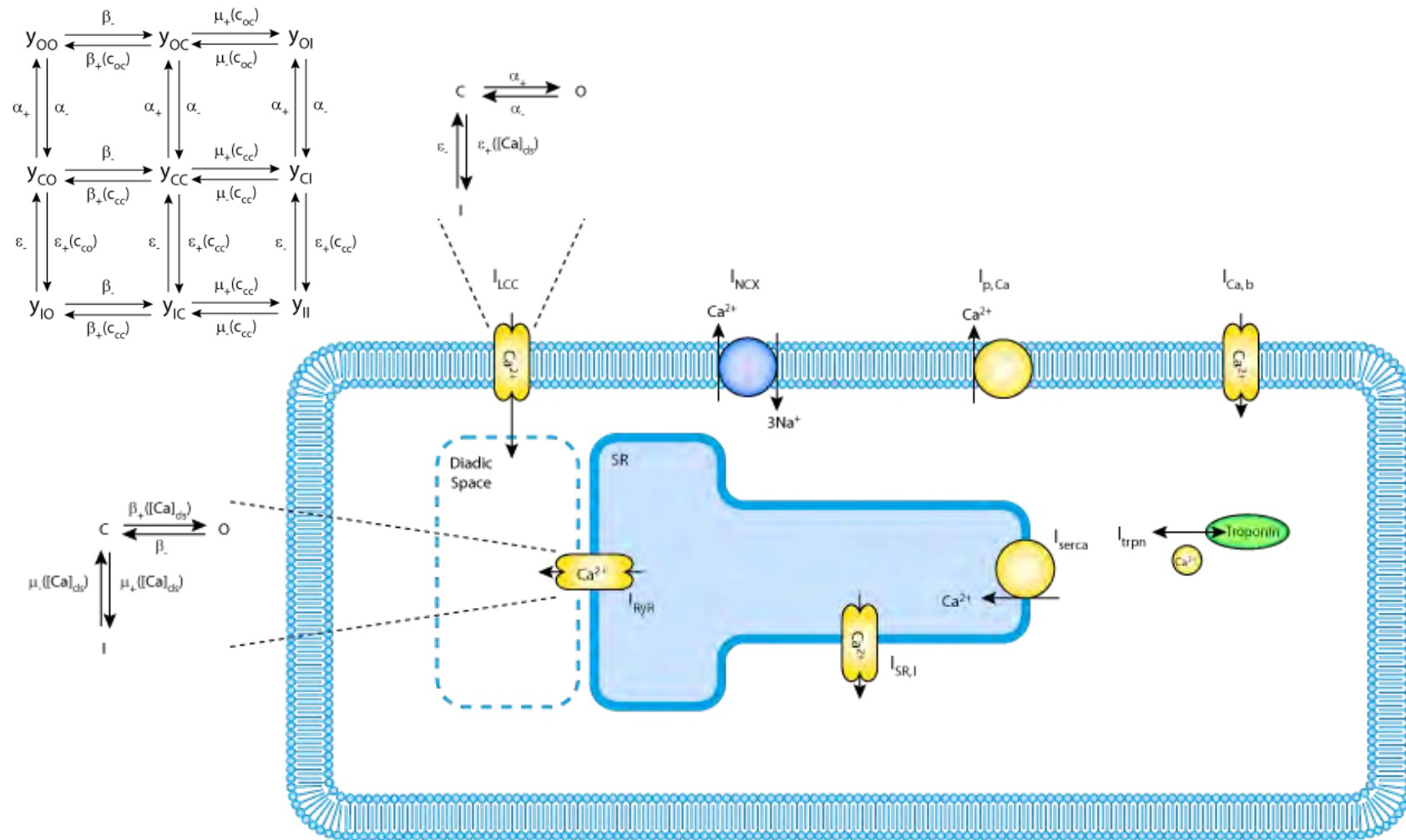
"A Quantitative Analysis of Cardiac Myocyte Relaxation: A Simulation Study" by Hinch, Greenstein, Tanskanen, Xu and Winslow.

The following is the abstract of this paper:

Calcium (Ca²⁺)-induced Ca²⁺ release (CICR) in cardiac myocytes exhibits high gain and is graded. These properties result from local control of Ca²⁺ release. Existing local control models of Ca²⁺ release in which interactions between L-Type Ca²⁺ channels (LCCs) and ryanodine-sensitive Ca²⁺ release channels (RyRs) are simulated stochastically are able to reconstruct these properties, but only at high computational cost. Here we present a general



Hinch *et al.* model of Ca-induced Ca release



Niederer *et al.* model of myofilament mechanics

- Niederer S.A., Hunter P.J., Smith N.P. A quantitative analysis of cardiac myocyte relaxation: a simulation study. *Biophysical Journal* 2006;90(5):1697-722
- www.cellml.org/models/niederer_hunter_smith_2006_version02

The screenshot shows the CellML model page for 'A quantitative analysis of cardiac myocyte relaxation: a simulation study'. At the top, there are navigation tabs: overview, edit, view math, model metadata, curation, view cellml, data, and procedural code. Below the title, there is a section for downloading the model (52Kb) and solving it in different environments. The curation status is shown as two stars. The model is known to run in PCEnv and COR. The model structure section provides a brief abstract of the paper.

overview edit view math model metadata curation view cellml data procedural code

A quantitative analysis of cardiac myocyte relaxation: a simulation study

[Download Model](#) (52Kb) Solve model in: ([help](#))

Curation Status: ★★

PCEnv ★★ JSim ★★ COR ★★

PCEnv Session (What's this?)

Model Documentation

Model Status

This model is known to run in PCEnv and COR to reproduce the output shown in the publication. A PCEnv session file is also associated with this model.

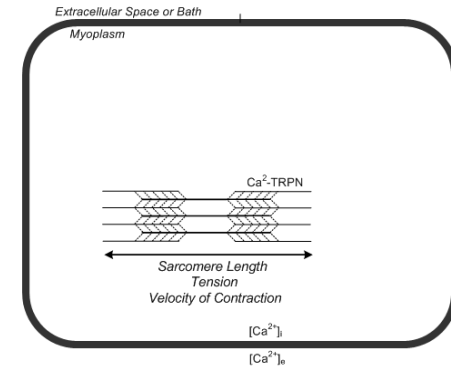
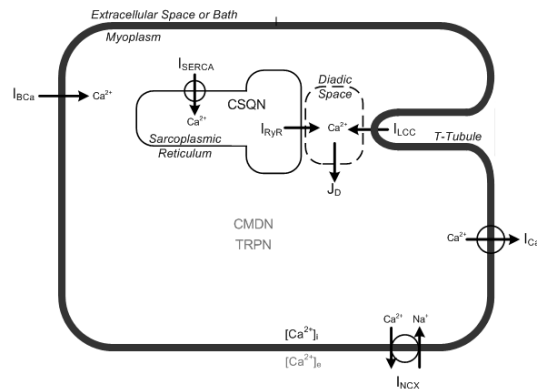
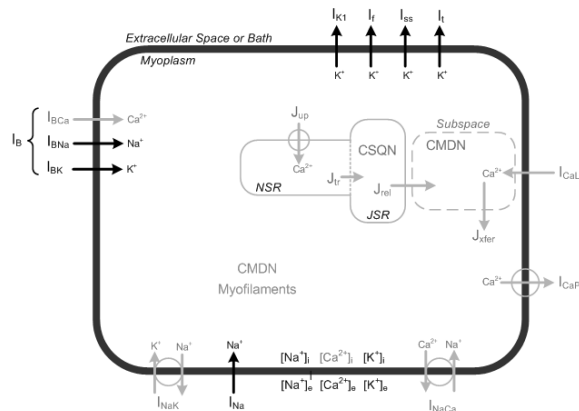
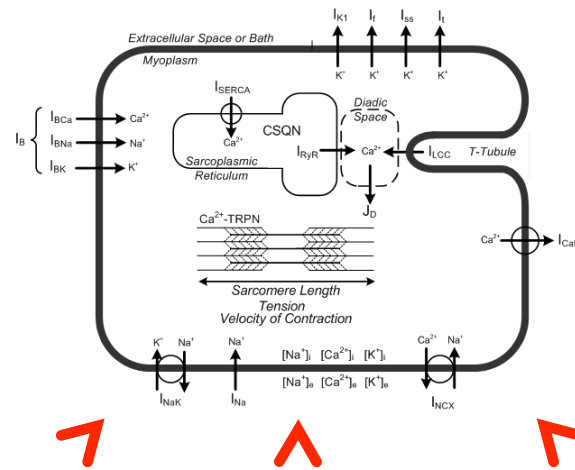
Model Structure

This CellML model was based on the March 2007 paper:

"A Quantitative Analysis of Cardiac Myocyte Relaxation: A Simulation Study" by S.A. Niederer, P.J. Hunter and N.P. Smith.

The following is the abstract of this paper: The determinants of relaxation in cardiac muscle are poorly understood, yet compromised relaxation accompanies various pathologies and impaired pump function. In this study, we develop a model of active contraction to elucidate the relative importance of the $[Ca^{++}]_i$ transient magnitude, the unbinding of Ca^{++} from troponin C (TnC), and the lengthdependence of tension and Ca^{++} sensitivity on relaxation. Using the framework proposed by one of our researchers, we extensively reviewed experimental literature, to quantitatively characterize the binding of Ca^{++} to TnC, the kinetics of tropomyosin, the availability of binding sites, and the kinetics of crossbridge binding after perturbations in sarcomere length. Model parameters were determined from multiple experimental results and modalities (skinned and intact preparations) and model results were validated against data from length step, caged Ca^{++} , isometric twitches, and the half-time to relaxation with increasing sarcomere length experiments. A factorial analysis found that the $[Ca^{++}]_i$ transient and the unbinding of Ca^{++} from TnC were the primary determinants of relaxation, with a fivefold greater effect than that of

Combine models using CellML import



Terkildsen *et al.* Integrated model of e-c coupling

- Terkildsen J.R., Niederer S., Crampin E.J., *et al.* Using Physiome standards to couple cellular functions for cardiac excitation-contraction. *Experimental Physiology* 93, pp919-929, 2008.
- www.cellml.org/models/terkildsen_niederer_crampin_hunter_smith_2008_version02

The screenshot shows the top navigation bar of the CellML website with tabs: overview, edit, view math, model metadata, curation, view cellml, data, and procedural code. The title of the model is 'Using Physiome standards to couple cellular functions for rat cardiac excitation-contraction'. Below the title, there is a section with a 'Download Model (220Kb)' button and a 'Solve model in: (help)' link. To the right of these are three icons with star ratings: PCEnv (3 stars), JSim (4 stars), and COR (4 stars). Below the icons, the 'Curation Status' is shown as two stars, followed by links for 'PCEnv Session' and 'What's this?'.

Model Documentation

Model Status

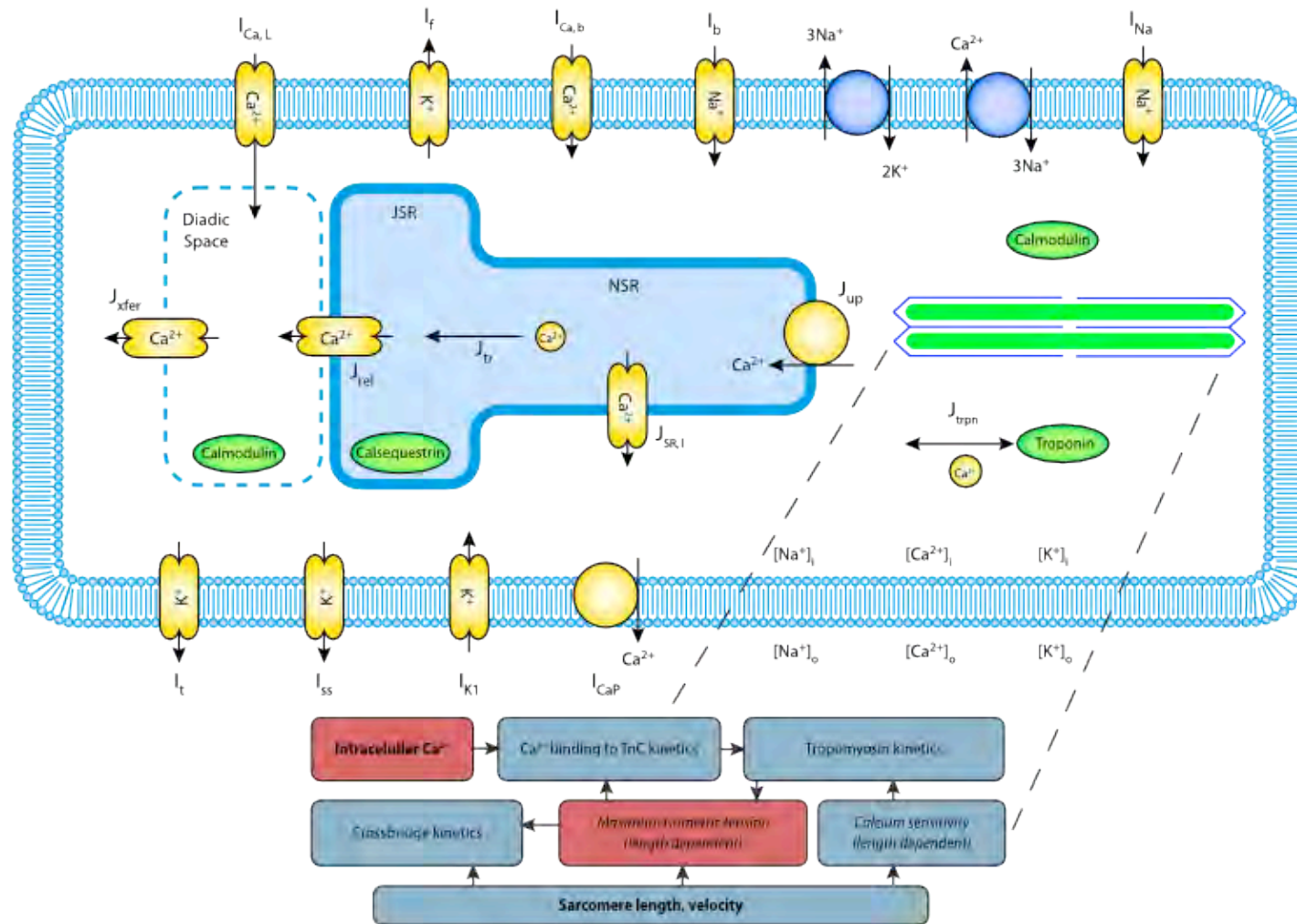
Since this is the exact same model encoding which was used to create the results presented in the paper, this CellML model is known to accurately represent the published article.

Model Structure

ABSTRACT: Scientific endeavour is reliant upon the extension and reuse of previous knowledge. The formalization of this process for computational modelling is facilitated by the use of accepted standards with which to describe and simulate models, ensuring consistency between the models and thus reducing the development and propagation of errors. CellML 1.1, an XML-based programming language, has been designed as a modelling standard which, by virtue of its import and grouping functions, facilitates model combination and reuse. Using CellML 1.1, we demonstrate the process of formalized model reuse by combining three separate models of rat cardiomyocyte function (an electrophysiology model, a model of cellular calcium dynamics and a mechanics model) which together make up the Pandit-Hinch-Niederer et al. cell model. Not only is this integrative model of rat electromechanics a useful tool for cardiac modelling but it is also an ideal framework with which to demonstrate both the power of model reuse and the challenges associated with this process. We highlight and classify a number of these issues associated with combining models and provide some suggested solutions.



Terkildsen *et al.* Integrated model of e-c coupling



Next steps

- Decompose models into reusable functional units
- Place decomposed units into a library
- Import units from library to build new models
- e.g. Pandit model contains many reusable units...



New CellML model repository

- CellML 1.0 models are contained in a single file.
- CellML 1.1 allows a model to be decomposed into many smaller files by aggregating submodels.
- Managing shared component files is difficult.
e.g. updating a shared component for one model may break another, or it may be difficult to locate the correct component to be included into a model.
- The new CellML repository seeks address these issues by using a distributed version control system (DVCS).
- DVCS allows a group of modellers to work together to construct models, allowing multiple versions of shared components, without relying on a centralised server.



Problems

- CellML is useful to exchange implementation-independent descriptions of models.
- But higher-level knowledge is not captured in the process of constructing the CellML representation.
- The information may be included informally as metadata, but this approach is insufficiently rigorous.
- We need a *linking* mechanism to associate CellML entities to well-defined representations of knowledge (*ontologies*).



How are we using ontologies?

- We are developing ontologies for physiological form and function, and the CellML modelling language.
- Other groups are also developing ontologies relevant to biological modelling:
 - Anatomical (Gene Ontology/GONG, FMA);
 - Gene regulation pathway (Gene Ontology/GONG, BioPax);
 - Gene expression (Gene Ontology/GONG);
 - Common access to bioinformatics sources (TAMBIS/TaO);
 - Physical and mathematical (SBO, Stanford Knowledge Systems, OPB).
- We are linking model entities in the CellML repository to our own and other ontologies.
- Sarala Wimilaratne has been using ontologies to enable visualisation of CellML models.



Simulation and authoring

- Plugin packages for running CellML models:
 - CMISS/CMGUI (University of Auckland);
 - PCEnv (University of Auckland);
 - Cellular Open Resource (University of Oxford);
 - JSim (University of Washington);
 - Virtual Cell (University of Connecticut Health Center);
 - insilicoIDE (Osaka University).
- CellML tools are available at **<http://www.cellml.org/tools/>**
- COR and PCenv have combined their efforts to create **openCell**.



Programmable interfaces

- Query interfaces to databases
 - RDF, OWL, simple HTTP GET
- CellML API
 - This API can be accessed either directly from your C++ program, or from any language for which a CORBA language mapping is available (C, C++, Java, Python, Ada, COBOL, Lisp, PL I, Smalltalk, Tcl, Eiffel, Ruby, and probably more).
 - Math - generating code for simulation environments (OMG interface definition language)



Proposed changes for next version of CellML

- Rewrite of specifications to separate normative from informative text.
- Introduce the concept of secondary specifications restrict the general CellML specification to more specific subsets.
- Removal of ‘reaction’ element (domain-specific information represented by ontologies).
- New typing system to allow non-real and structured types.
Current proposal uses lambda calculus to integrate base types, structures, and units into a unified typing specification.



Summary

- CellML is a versatile format to define high-level representations of biological models.
- The CellML repository provides public access to a wide range of curated biological models.
- The addition of ontological links enables model entities to be associated with formal representations of relevant domains of knowledge.



People

- **Randall Britten** (software coordinator)
- **Mike Cooling** (model creation and reuse)
- **Sarala Dissanayake** (model visualisation)
- **Alan Garny** (software tools)
- **Matt Halstead** (ontologies)
- **Peter Hunter** (director)
- **James Lawson** (model creation and curation)
- **Catherine Lloyd** (model creation and curation);
- **Justin Marsh** (software tools)
- **Andrew Miller** (specifications, interfaces, tools);
- **David Nickerson** (model creation, specifications)
- **Poul Nielsen** (project leader, specifications)
- **Penny Noble** (model creation)
- **Tommy Yu** (model repository)

