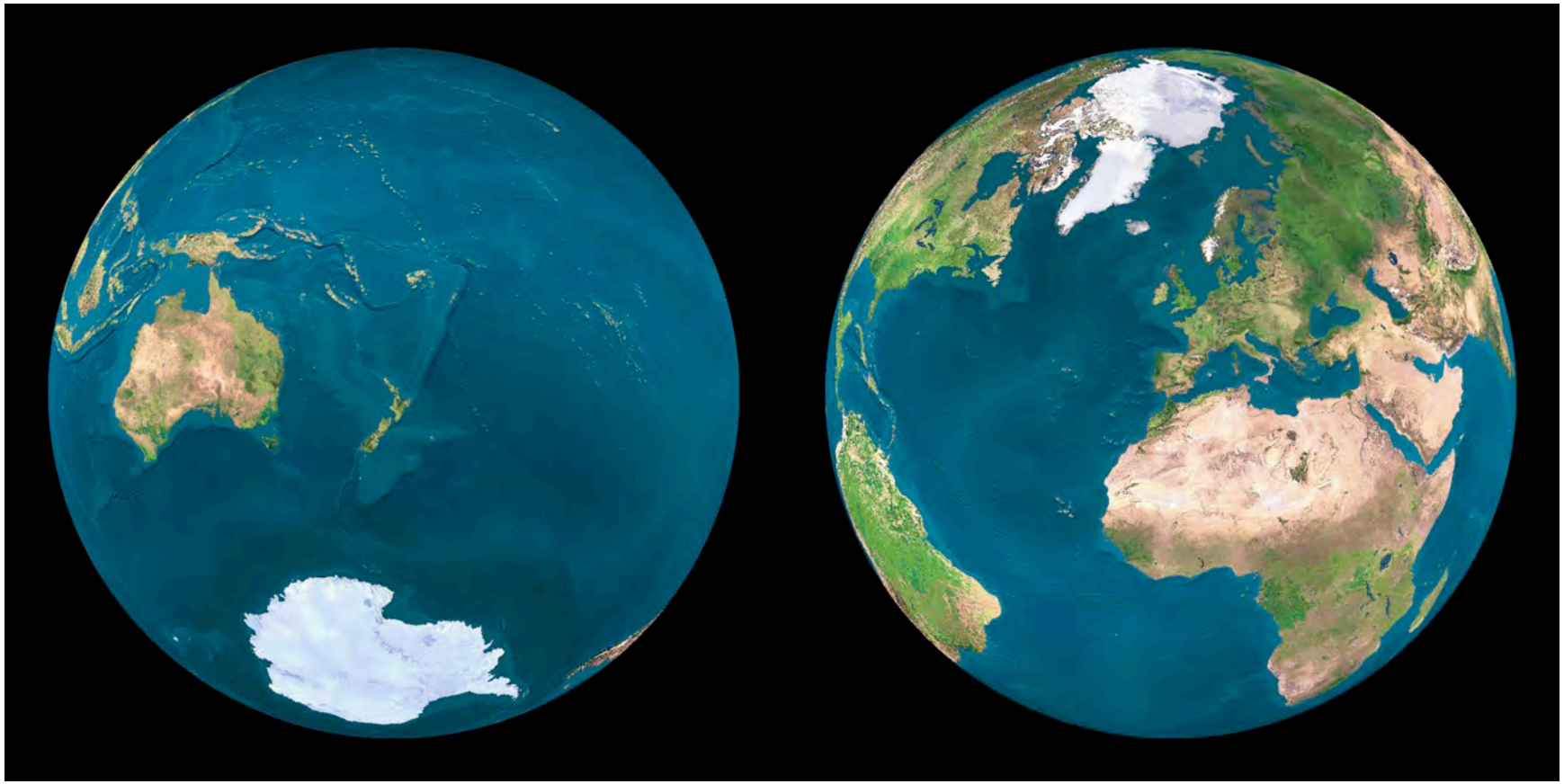
CellML SBGN SBO BioPAX MIASE Workshop 2009 Waiheke Island

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/)*

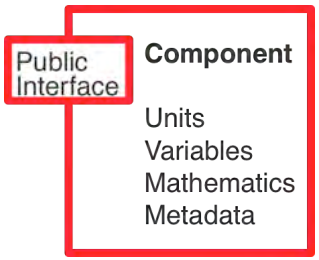
…CellML

- 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)*

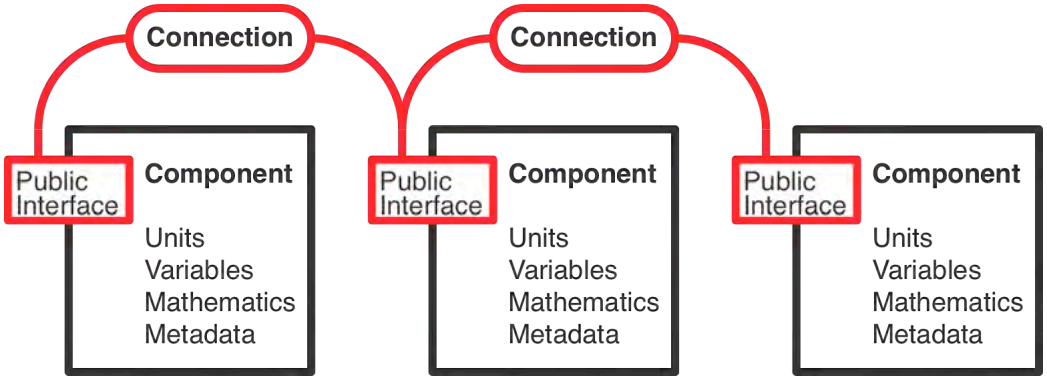
CellML components

- 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.



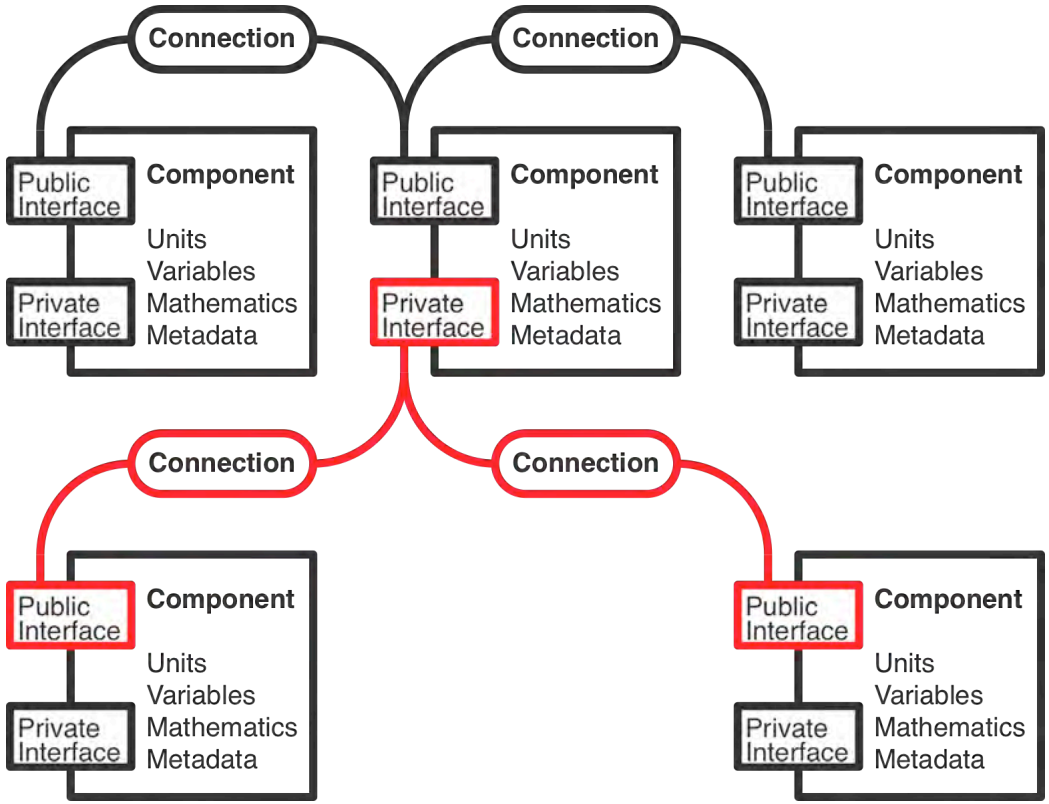
CellML connections

- *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*.



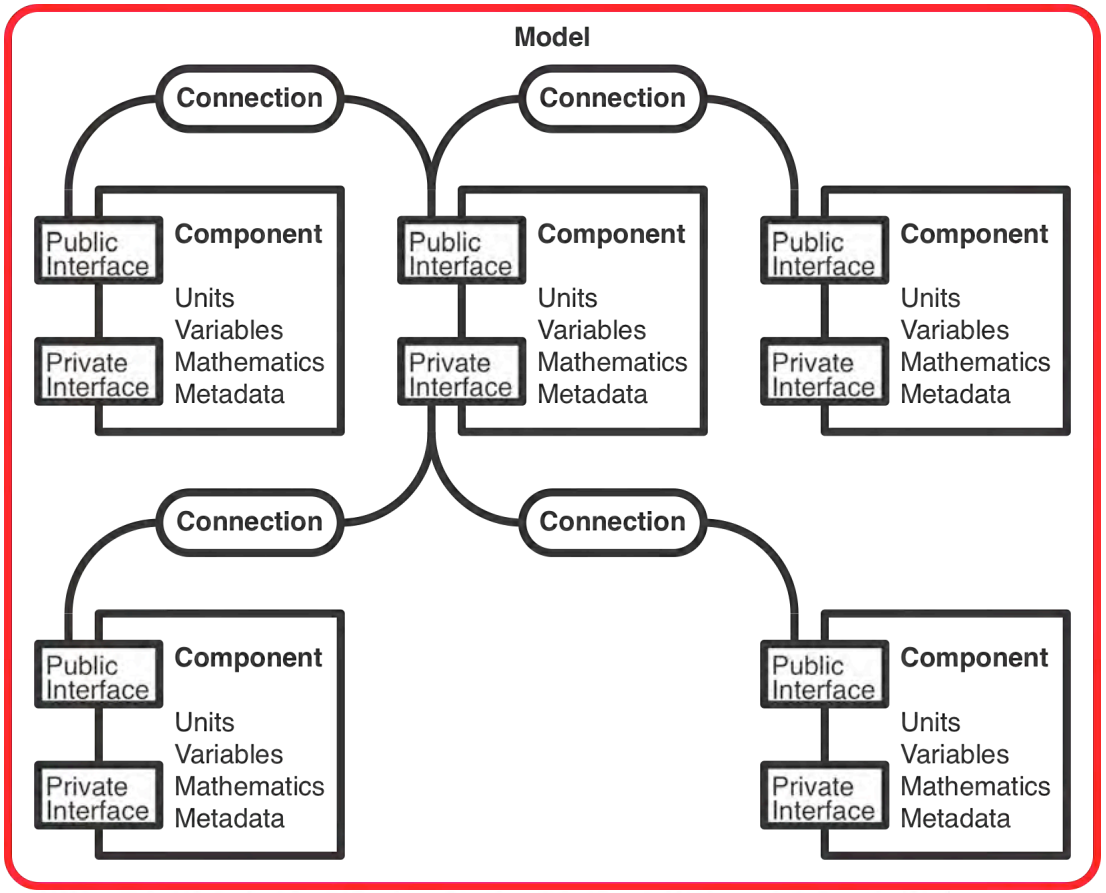
CellML encapsulation

- Encapsulation hierarchies are enabled using *private interfaces*.



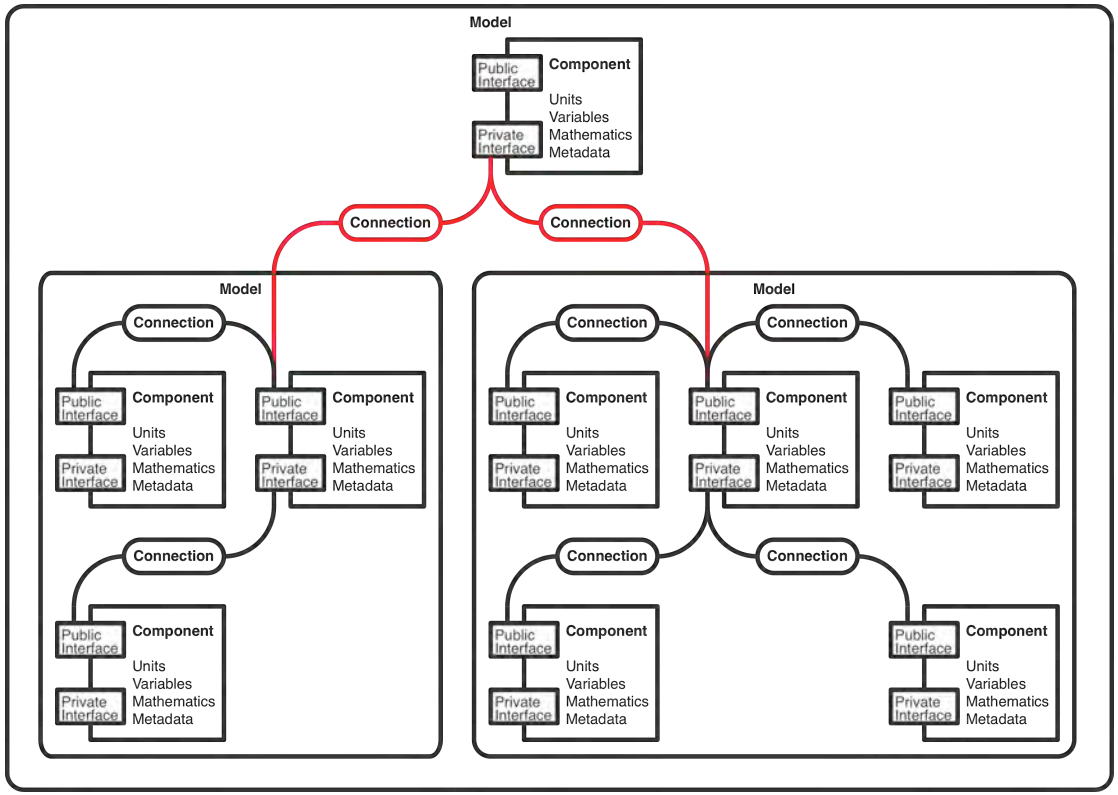
CellML model

- 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)*

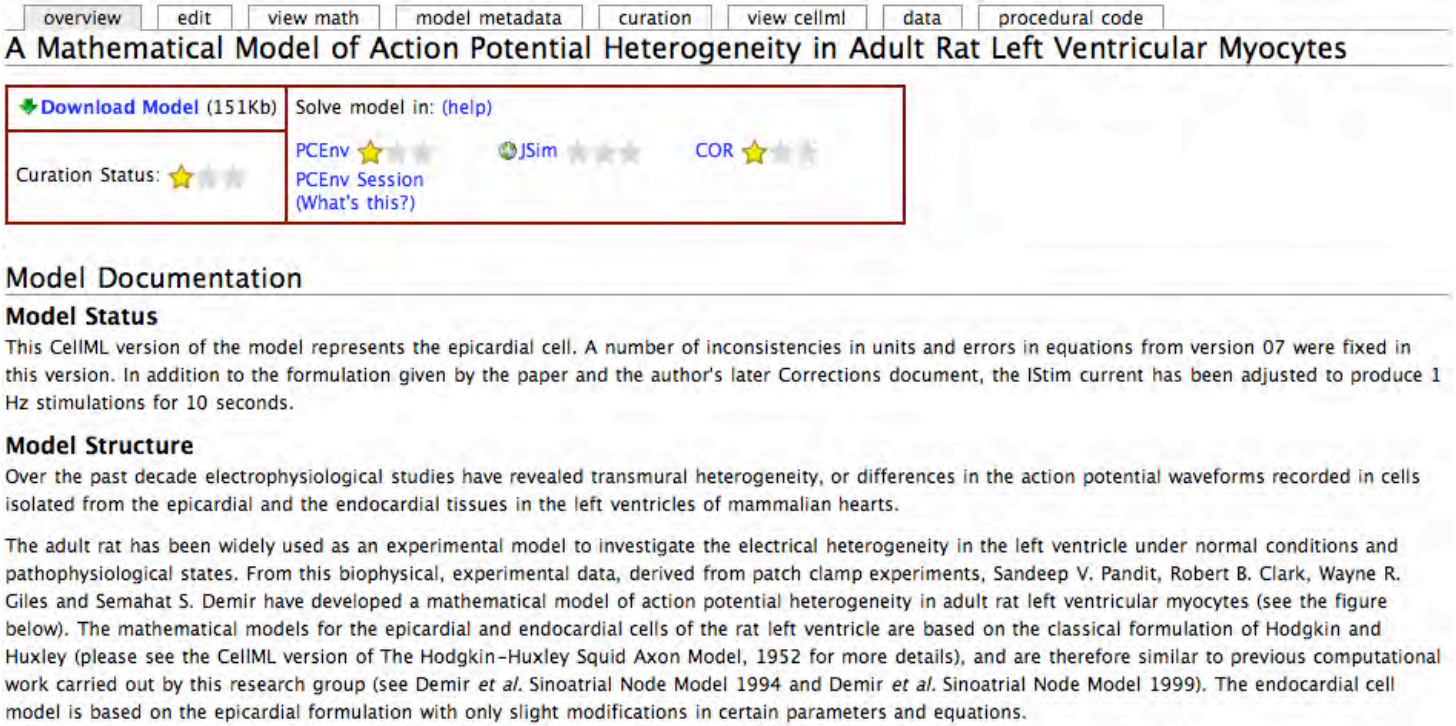
Model curation

- 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 Novere, 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)*

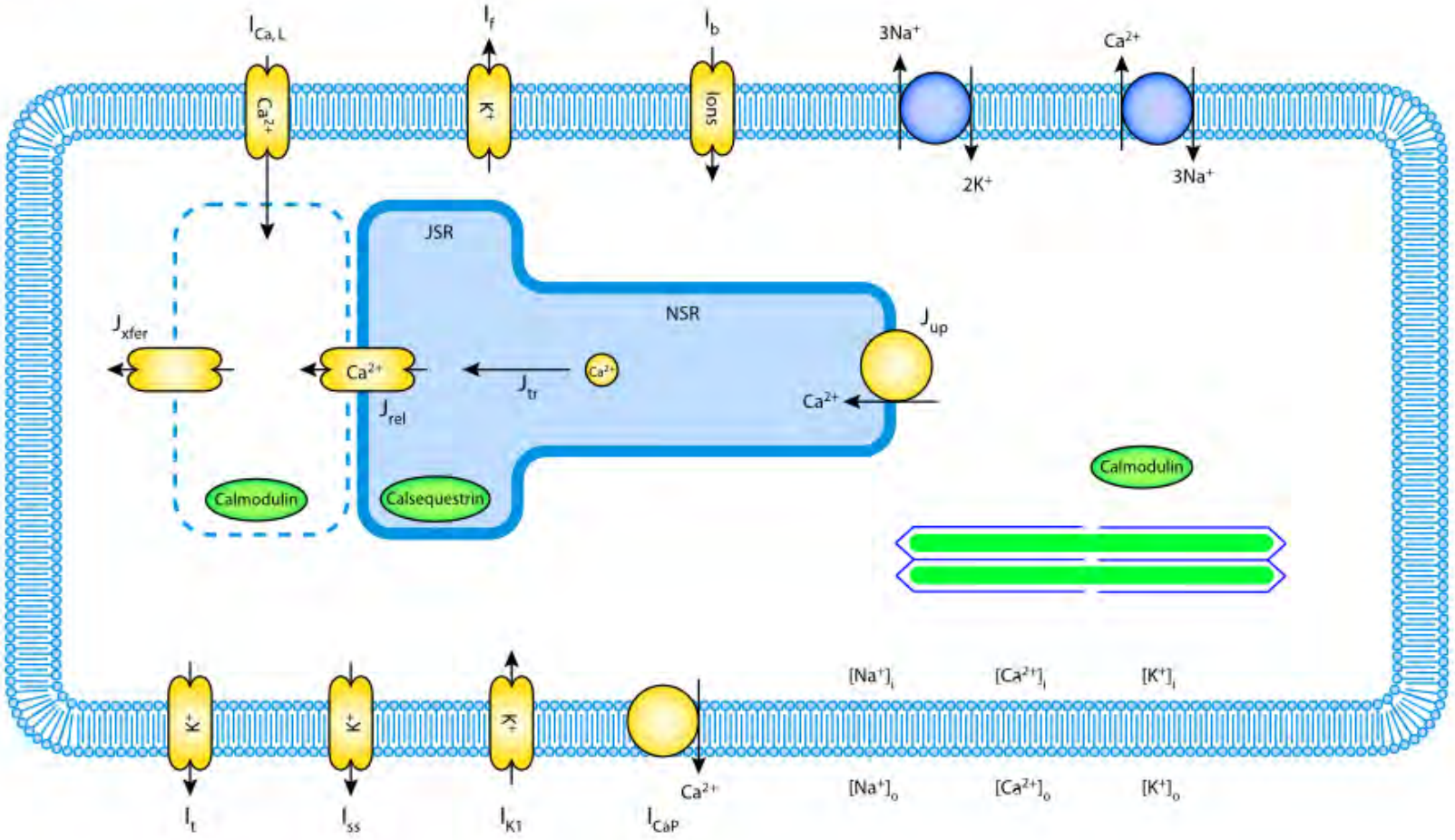
Pandit *et al*. model of cardiac action potential

- 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



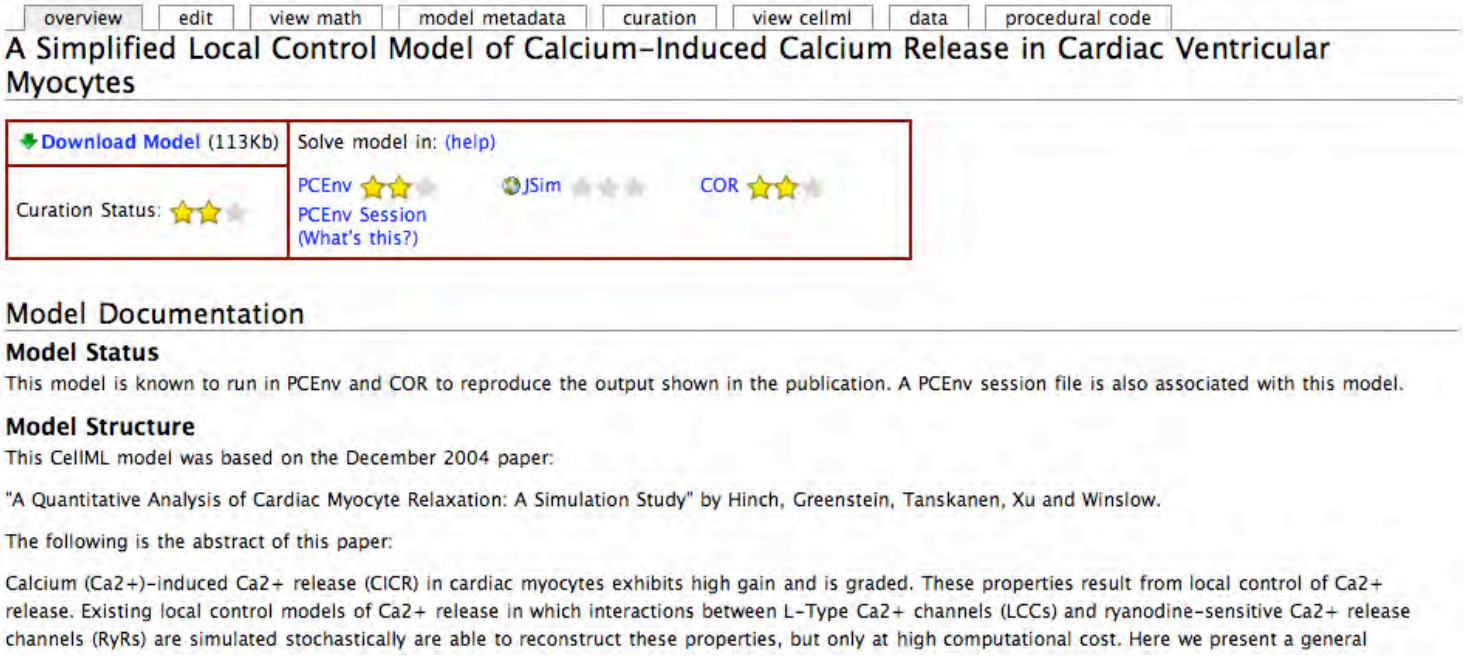
Pandit *et al*. model of cardiac action potential

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

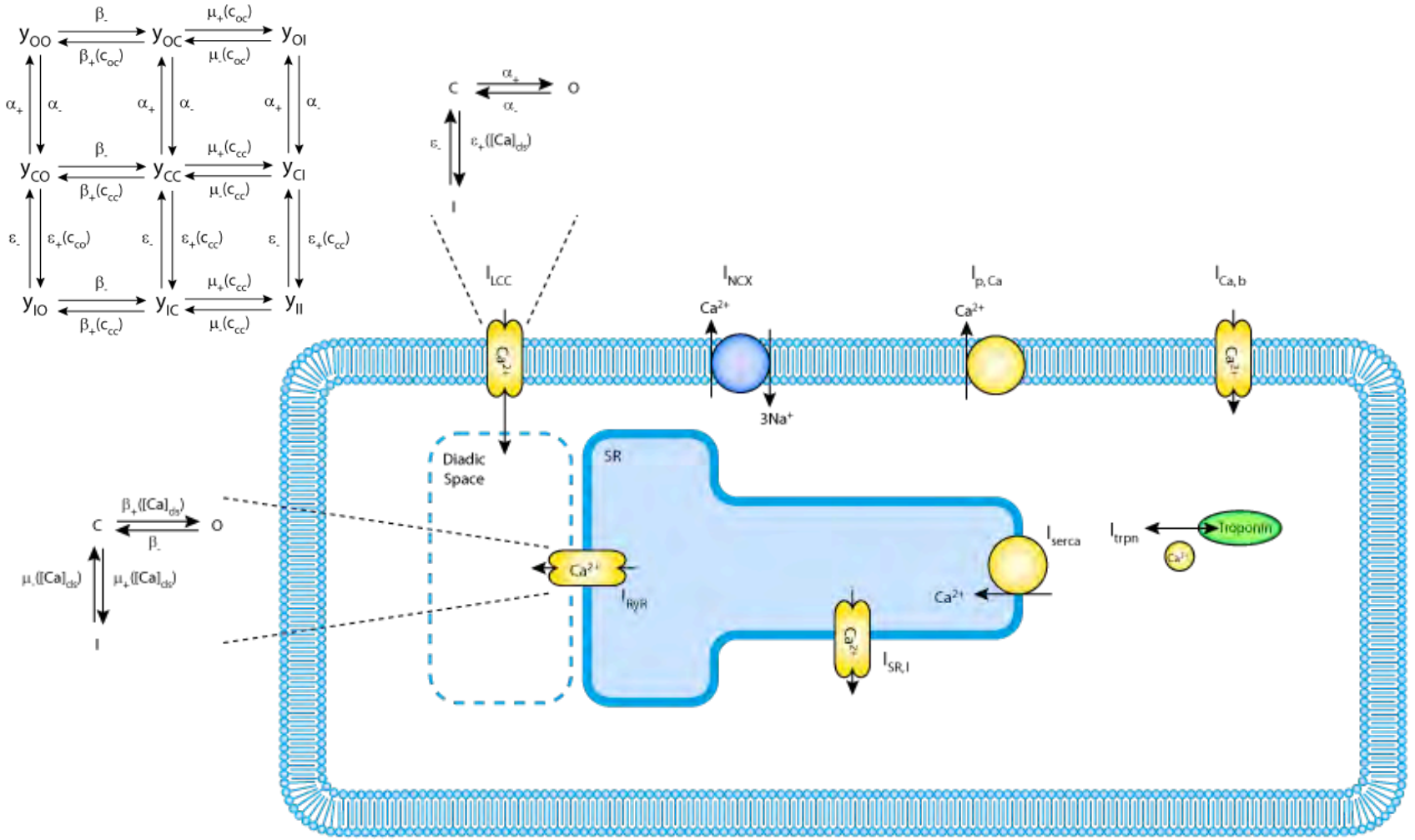


Hinch *et al*. model of Ca-induced Ca release

- 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

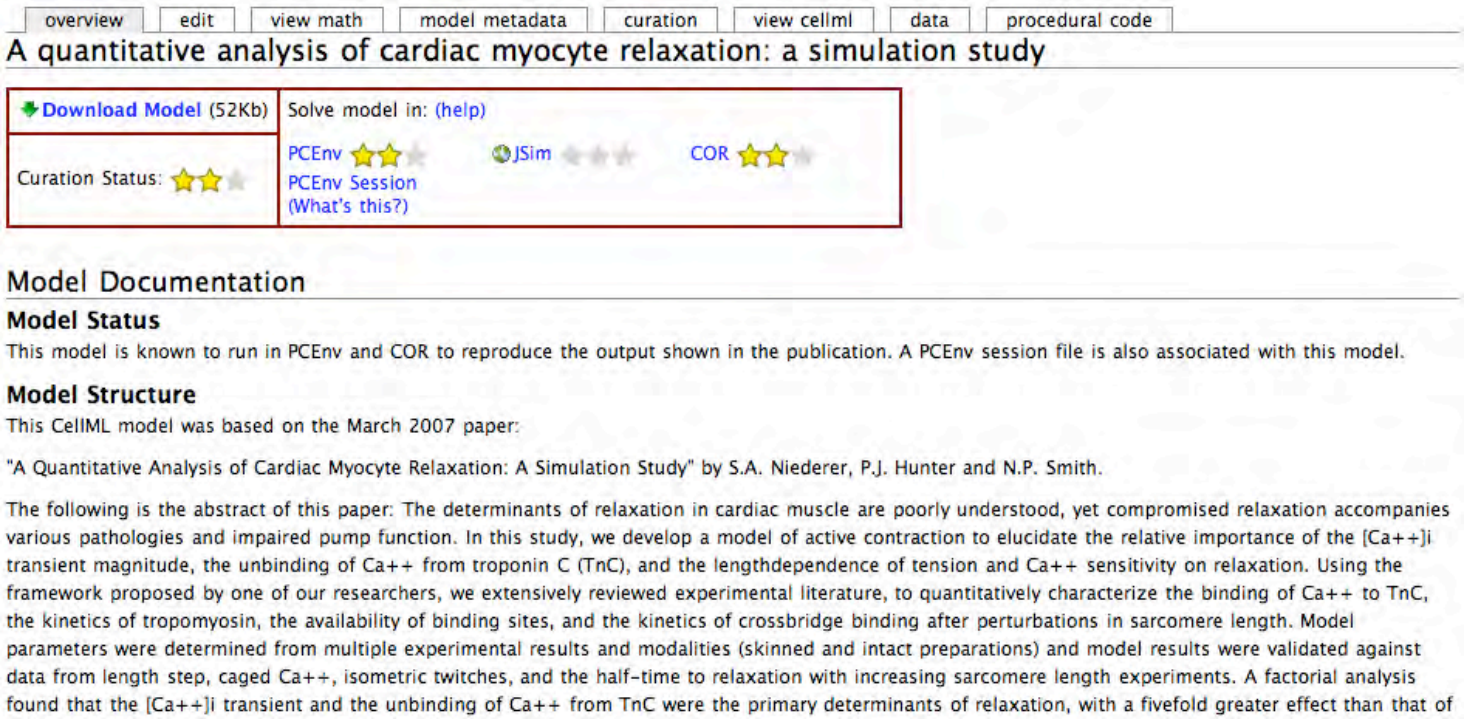


Hinch *et al*. model of Ca-induced Ca release

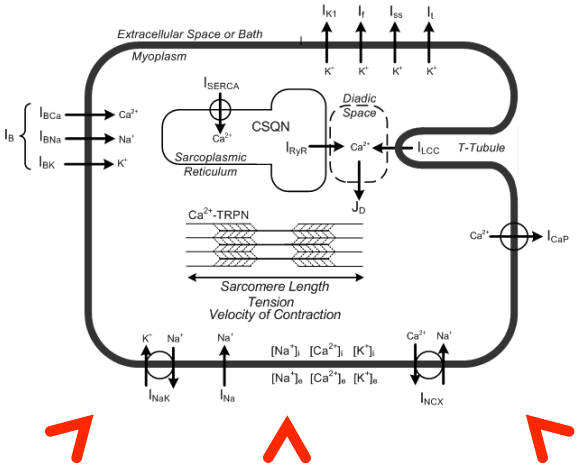


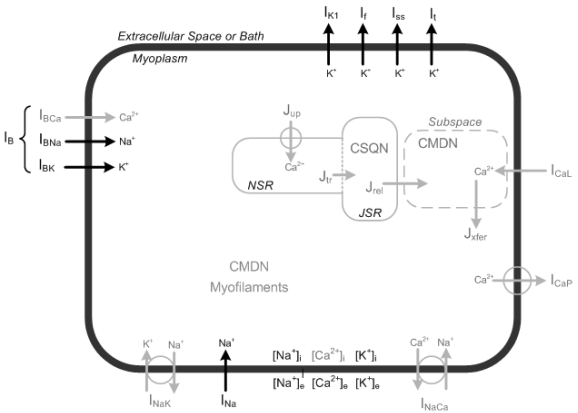
Niederer *et al*. model of myofilament mechanics

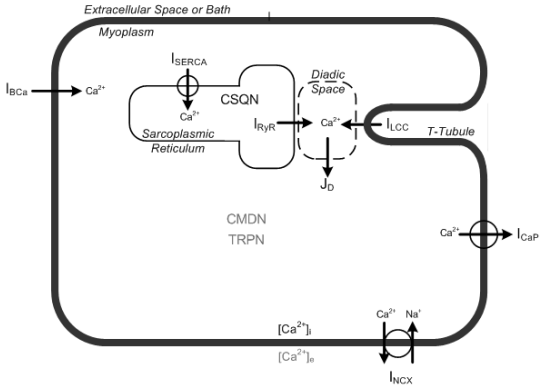
- Niederer S.A., Hunter P.J., Smith N.P. A quantitative analysis of cardiac myocyte relaxation: a simulation study. *Biophysical Journal* 2006;90(5):1697-722
- www.cellml.org/models/niederer_hunter_smith_2006_version02

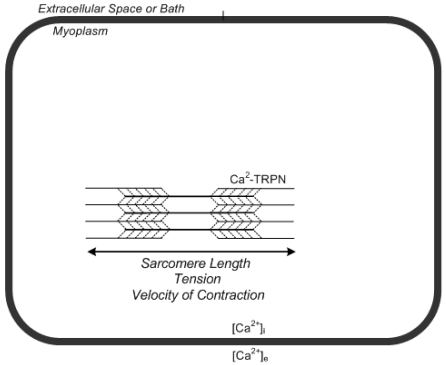


Combine models using CellML import



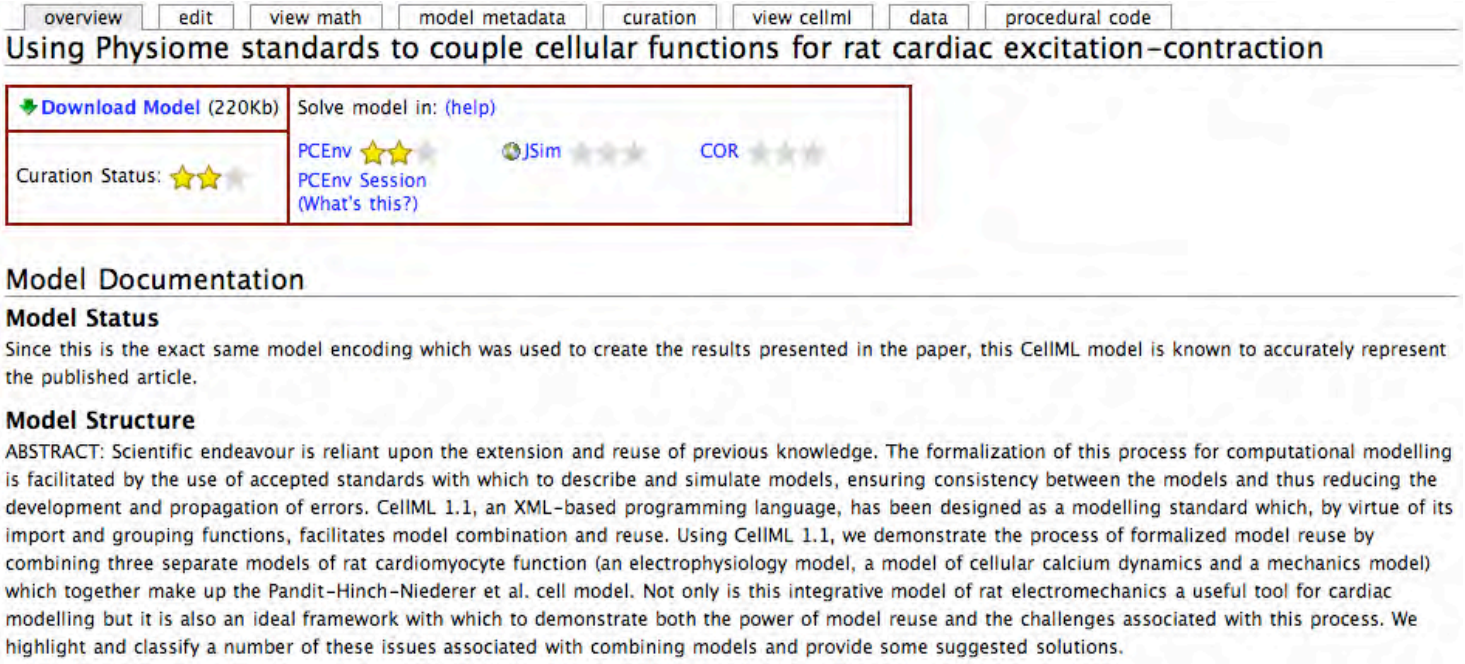




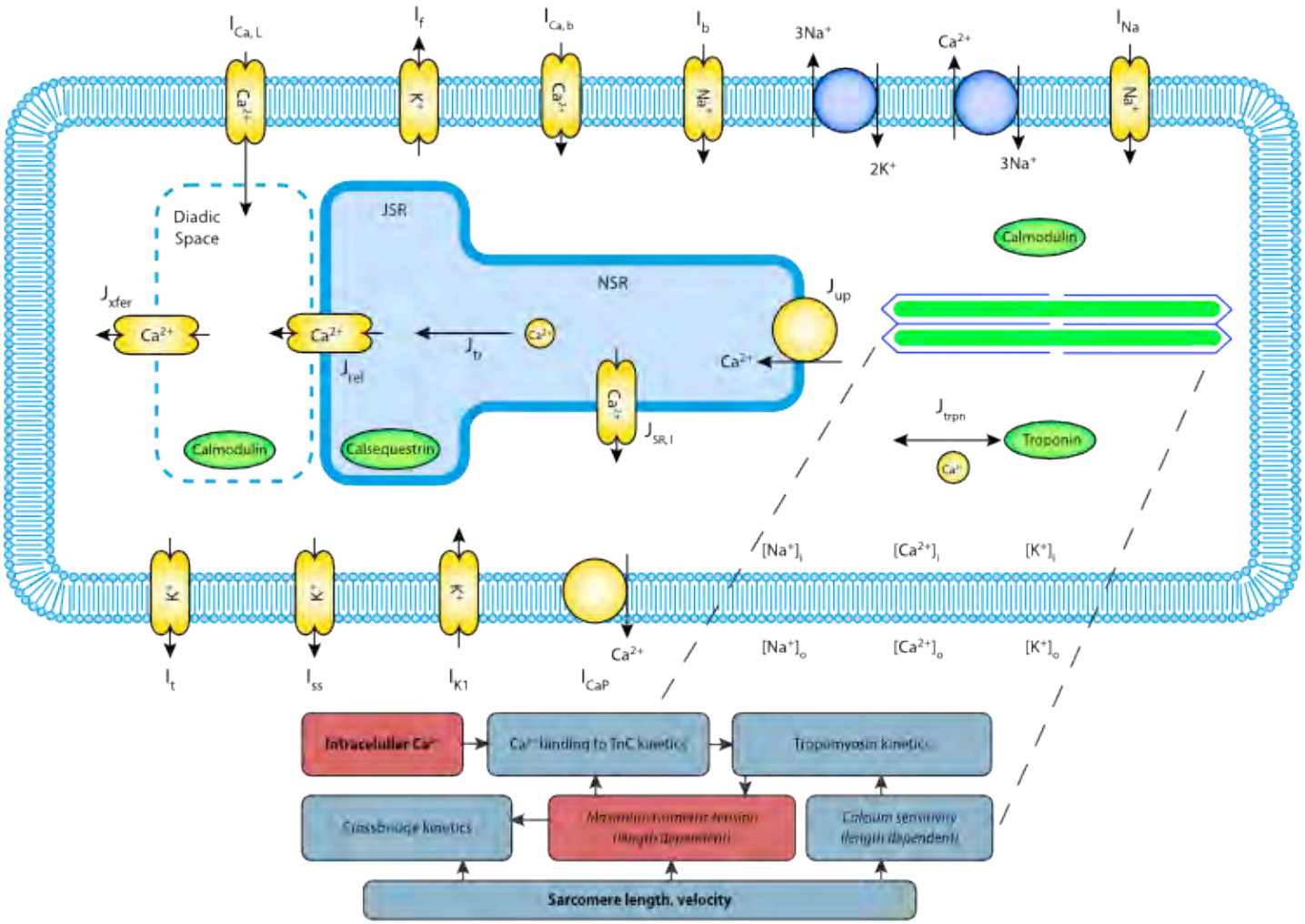


Terkildsen *et al*. Integrated model of e-c coupling

- 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



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 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)