Semantic annotation in the Physiome Model Repository

David Nickerson
Auckland Bioengineering Institute
Auckland, New Zealand
Physiome Model Repository

- [https://models.physiomeproject.org](https://models.physiomeproject.org)
- Over 800 public workspaces
  - Each independently version controlled
  - Persistent releases of specific versions (exposure)
  - Many different types of biology and mathematics
  - Proteins through to whole organ and larger scale
- Historically CellML (+SED-ML) models
- Modularity and reuse
- Consistency between browser and tool integration
  - content type negotiation
  - REST
PMR Semantics

• User indicates the resources they would like indexed
  • Within a workspace (~COMBINE archive)
  • Some smarts in extracting RDF from CellML models
  • Various RDF serialization formats supported

• Versioning
  • Workspace – latest version
  • Exposures

• SPARQL endpoint
  • Read-only
  • Permissions filter
Why do we annotate?

• Comprehension
The ORd human ventricular action potential model

This workspace houses a CellML 1.0 encoding of the 2011 O’Hara, Virág, Varró, & Rudy 2011 human cardiac ventricular action potential model (ORd). The original article is available at: http://www.ncbi.nlm.nih.gov/pubmed/21637795. This model was encoded based on the Matlab version of the code available from: http://rudylab.wustl.edu/research/cellml/.

The CellML 1.0 encoding of the ORd model was contributed by Steven Niederer. While the units in the CellML encoding are not yet perfect, it is a match for the Matlab code and matches the simulation output for a single beat perfectly. The figure below shows the output of the simulation experiment action-potential.xml encoded in SED-ML using the original version of the model from Steve. This output is generated by running the simulation experiment using the SED-ML Web Tools.

Component: intracellular_ions

\[
\begin{align*}
\frac{d}{dt} (CaMK) &= aCaMK CaMKb (CaMKb + CaMK) - (bCaMK CaMK) \\
\text{cmdmax} &= \begin{cases} 
\text{cmdmax}_b.1.3 \text{ if celltype} = 1 \\
\text{cmdmax}_b \text{ otherwise} 
\end{cases} \\
\frac{d}{dt} (cai) &= \left( \frac{-(\text{Ca}^2+\text{Ca}^{-}+\text{Ca}^{2+}+\text{Ca}^{-}+\text{Ca}^{2+}) \text{ Acap}}{v \text{ cm Acap}} \right) \\
\frac{d}{dt} (nai) &= \left( \frac{-(\text{Na}^++\text{Na}^{-}+\text{Na}^{2+}+\text{Na}^{-}+\text{Na}^{2+}) \text{ Acap}}{v \text{ cm Acap}} \right) \\
\frac{d}{dt} (nass) &= \left( \frac{-(\text{Na}^++\text{Na}^{-}+\text{Na}^{2+}+\text{Na}^{-}+\text{Na}^{2+}) \text{ Acap}}{v \text{ cm Acap}} \right) - \text{JdiffNa} \\
\frac{d}{dt} (ki) &= \left( \frac{-(\text{K}^++\text{K}^{-}+\text{K}^{2+}+\text{K}^{-}+\text{K}^{2+}) \text{ Acap}}{v \text{ cm Acap}} \right) \\
\frac{d}{dt} (cass) &= \left( \frac{-(\text{Ca}^2+\text{Ca}^{-}+\text{Ca}^{2+}+\text{Ca}^{-}+\text{Ca}^{2+}) \text{ Acap}}{v \text{ cm Acap}} \right) + Bcais \\
\frac{d}{dt} (cansr) &= \text{Jup} - \left( \frac{Jr \text{ vcap}}{v \text{ cm Acap}} \right) \\
\frac{d}{dt} (cajsr) &= Bcajsr - Jrel
\end{align*}
\]
Generated Code

The following is MATLAB code generated by the CellML API from this CellML file. (Back to language selection)

The raw code is available.

```matlab
function [RATES, ALGEBRAIC] = computeRates(VO2, STATES, CONSTANTS)
    global algebraicVariableCount;
    statesSize = size(STATES);
    statesColumnCount = statesSize(2);
    if (statesColumnCount == 1)
        STATES = STATES';
        ALGEBRAIC = zeros(1, algebraicVariableCount);
    else
        statesRowCount = statesSize(1);
        ALGEBRAIC = zeros(statesRowCount, algebraicVariableCount);
        RATES = zeros(statesRowCount, statesColumnCount);
    end

    ALGEBRAIC(:,:,1) = 1.00000./(1.00000+exp((STATES(:,:,1)-87.6100)/7.48800));
    RATES(:,:,1) = (ALGEBRAIC(:,:,1) - STATES(:,:,1))./CONSTANTS(:,:,44);
    ALGEBRAIC(:,:,2) = 1.00000./(1.00000+exp((STATES(:,:,1)+93.8100)/7.48800));
    RATES(:,:,2) = (ALGEBRAIC(:,:,2) - STATES(:,:,1))./CONSTANTS(:,:,96);
    ALGEBRAIC(:,:,3) = 1.00000./(1.00000+exp(-(STATES(:,:,1)-CONSTANTS(:,:,32))./CONSTANTS(:,:,33)));
    ALGEBRAIC(:,:,4) = 1.00000./((STATES(:,:,1)-CONSTANTS(:,:,36))./CONSTANTS(:,:,34))/CONSTANTS(:,:,35)+CONSTANTS(:,:,37).*exp(-(STATES(:,:,1)+CONSTANTS(:,:,38))./CONSTANTS(:,:,39));
    RATES(:,:,3) = (ALGEBRAIC(:,:,1) - STATES(:,:,1))./ALGEBRAIC(:,:,14);
    ALGEBRAIC(:,:,4) = 1.00000./(1.00000+exp(-(STATES(:,:,1)+CONSTANTS(:,:,40))./CONSTANTS(:,:,41)));
    ALGEBRAIC(:,:,5) = 1.00000./((1.432000-0.5).*exp(-(STATES(:,:,1)+1.19600)/6.28500)+6.14000.*exp((STATES(:,:,1)+0.595600)/28.2700));
    RATES(:,:,4) = (ALGEBRAIC(:,:,2) - STATES(:,:,2))./ALGEBRAIC(:,:,15);
    ALGEBRAIC(:,:,6) = 1.00000./((0.00979400.*exp(-(STATES(:,:,1)+17.9500)/28.0500)+0.334300.*exp((STATES(:,:,1)+5.73000)/56.6600));
    RATES(:,:,6) = (ALGEBRAIC(:,:,2) - STATES(:,:,13))./ALGEBRAIC(:,:,16);
    ALGEBRAIC(:,:,7) = 1.00000./(1.00000+exp(-((STATES(:,:,1)-14.34000)/14.32000));
    ALGEBRAIC(:,:,8) = 1.05150./((1.00000./((1.28900-1.00000.*exp(-(STATES(:,:,1)-18.48099)/29.3814))+3.50000./((1.00000+exp((STATES(:,:,1)+100.000)/29.3814)));
    RATES(:,:,8) = (ALGEBRAIC(:,:,5) - STATES(:,:,20))./ALGEBRAIC(:,:,18);
    ALGEBRAIC(:,:,9) = 1.00000./(1.00000+exp(-(STATES(:,:,1)+3.94800)/4.23000));
```

Source

Derived from workspace An encoding of the human ORd model by Steve Neiderer at changeset 2593df010620.

Collaboration

To begin collaborating on this work, please use your git client and issue this command:

```bash
git clone https://models.physionet.org
```

Downloads

- Download This File
- Complete Archive as .tgz

Views Available

- Documentation
- Model Metadata
- Mathematics
- Generated Code
- Cite this model
- Source View
- Launch with OpenCOR

Tools

- Compare...
- Combine Archive Web

License

The terms of use/license for this work is unspecified.
Why do we annotate?

- Comprehension
- Modularity and reuse
Modular modelling with Physiome standards

Michael T. Cooling1, David P. Nickerson4, Paul M. F. Nielsen2 and Peter J. Hunter3
1Auckland Bioengineering Institute, the University of Auckland, New Zealand
2Department of Engineering Science, the University of Auckland, New Zealand

Key points
- The complexity of computational models is increasing, supported by research in modelling tools and frameworks. But relatively little thought has gone into design principles for complex models.
- We propose a set of design principles for complex model construction with the Physiome standard modelling protocol CaBBi.
- By following the principles, models are generated that are extensible and are suitable themselves for reuse in larger models of increasing complexity.
- We illustrate these principles with examples including an architectural prototype linking, for the first time, electrophysiology, thermodynamically compliant metabolism, signal transduction, gene regulation and synthetic biology.
- The design principles complement other Physiome research projects, facilitating the application of virtual experiment protocols and model analysis techniques to assist the modelling community in creating libraries of composable, characterised and simulatable simulation modules.

Perspective

A Reappraisal of How to Build Modular, Reusable Models of Biological Systems

Maxwell L. Neal1*, Michael T. Cooling2, Lucian P. Smith1, Christopher T. Thompson3, Herbert M. Sauro1, Brian E. Carlson4, Daniel L. Cook5, John H. Gennari6
1Department of Bioengineering, University of Washington, Seattle, Washington, United States of America, 2Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand, 3Department of Physiology, Medical College of Wisconsin, Milwaukee, Wisconsin, United States of America, 4Department of Molecular and Integrative Physiology, University of Michigan, Ann Arbor, Michigan, United States of America, 5Department of Physiology and Biophysics, University of Washington, Seattle, Washington, United States of America, 6Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, Washington, United States of America

SemGen: a tool for semantics-based annotation and composition of biosimulation models

Maxwell L. Neal1*, Christopher T. Thompson2, Karam G. Kim3, Ryan C. James1, Daniel L. Cook3, Brian E. Carlson4 and John H. Gennari3
1Seattle Children’s Research Institute, Center for Global Infectious Disease Research, Seattle, WA 98109, USA, 2Department of Molecular and Integrative Physiology, University of Michigan, Ann Arbor, MI, United States of America, 3Auckland Bioengineering Institute, The University of Auckland, Auckland, New Zealand, 4School of Math and Statistics, University of Melbourne, Victoria, Australia, 5ARC Centre of Excellence in Convergent Bio-Nano Science and Technology, Melbourne School of Engineering, University of Melbourne, Victoria, Australia, 6School of Mathematics and Statistics, University of Melbourne, Victoria, Australia

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on April 19, 2019; revised on September 9, 2019; editorial decision on September 18, 2019; accepted on September 24, 2019
Why do we annotate?

• Comprehension
• Modularity and reuse
• Search
Model Annotation and Discovery with the Physiome Model Repository

Dewan M. Sarwar 1, Reza Kalbasi 1, John H. Gennari 2, Brian E. Carlson 3, Maxwell L. Neal 4, Bernard de Bono 1, Koray Atalag 1, Peter J. Hunter 1 and David P. Nickerson 1,*

1 Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand
2 Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, Washington, USA
3 Molecular & Integrative Physiology, University of Michigan, Ann Arbor, Michigan, USA and
4 Center for Global Infectious Disease Research, Seattle Children’s Research Institute, Seattle, Washington, USA

https://github.com/dewancse/model-discovery-tool
Recommender System

sodium/hydrogen exchanger 3 is a Kidney model. It is located in proximal convoluted tubule, epithelial cell of proximal tubule, apical plasma membrane.

Biological Meaning: Flux of Na+ from luminal to cytosol through apical plasma membrane
Species: Rattus norvegicus
Gene: Slc9a3
Protein: sodium/hydrogen exchanger 3

Recommendations/suggestions based on existing models in PMR
Basolateral membrane model
- sodium/hydrogen exchanger 3 (human)
- low affinity sodium-glucose cotransporter (mouse)
- sodium/potassium-transporting ATPase subunit alpha-1 (rat)
Alternative model of sodium/hydrogen exchanger 3
Not Exist
Kidney model in PMR
- sodium/hydrogen exchanger 3 (human)
- low affinity sodium-glucose cotransporter (mouse)
- sodium/glucose cotransporter 1 (human)

Identity Matrix: #
#
# Percent Identity Matrix - created by Clustal2.1
#
#
1: sp|Q9ET37|S5A4A_MOUSE 100.00 22.86 17.86 21.86
2: sp|P48764|SL9A3_HUMAN 22.86 100.00 19.15 89.49
3: sp|P06685|AT1A1_RAT 17.86 19.15 100.00 18.20
4: sp|P26433|SL9A3_RAT 21.86 89.49 18.20 100.00

COMBINE 2019
https://doi.org/10.17608/k6.auckland.8858654
Why do we annotate?

• Comprehension
• Modularity and reuse
• Search
• Do cool stuff
Machine Learning

• Add something about how cool ML is and why we should be using it.
• Insert latest buzz word here.
• TensorFlow.
• ICSB tutorial.
• How does this relate to annotation?
Link to (clinical) data

Computational Systems / Modelling and Simulation

- Modelling Formalism
  - CellML, FieldML, SBML, SED-ML, BioSignalML

Tooling & Software
- openCOR
- OpenCMRIS
- SemGen/SemSim
- RICORDO/ApiNATOMY

Physiome Model Repository
- CellML, FieldML
  - biological/biophysical concepts
  - maths
  - experiment protocols
  - annotations
  - versioning / exposures
  - web view / API

Measurement Data
- >Reference phenotypes/
  - study specific biomarkers
- >Experimental data
- >Simulation results

Imaging Data
- >CT, MRI, X-Ray etc.
- >Biosignals (ECG, EEG)

Semantic Web / Knowledge Engineering

- Genomic / Molecular
- Clinical terminology
- Environmental
- Physical / Chemical
- Bio-ontologies

Ontologies

IDentifiers
- URI, GUID, OID

Information Systems / Clinical Data

- Modelling Formalism
  - openEHR, HL7 FHIR

Tooling & Software
- Archetype Editors
- Template Designers
- Software frameworks
- Data transformation
- Terminology service

Clinical Model Repository
- Archetypes, Terminology
  - clinical/administrative objects
  - provenance / peer review
  - annotations and term bindings
  - versioning
  - web view / API

Tooling & Software
- XML tools
- Ontology editors
- Inference engines
- Lexical tools
- Open Data frameworks
- Discovery/visualisation

W3C Standards
- XML, RDF, OWL, SPARQL, LOD

Clinical Systems
- Biobanks / Research Data
- EMR / EHR
- National / Regional Clinical Data Repositories
- Personal Health Records

https://cellml.org/2019/05/03/auckland-cellml
Clinical data Repository → Clinical data annotations (SNOMED CT, LOINC) → Manual Ontology Mapping → Model annotation (composite annotation) → Model Repository

- OPB: Chemical concentration
  property_of

- CHEBI: potassium (1+)
  part_of

- FMA: Portion of blood
Epithelial Modelling Platform

https://github.com/dewancse/epithelial-modelling-platform
sodium/hydrogen exchanger 3 is a Kidney model. It is located in proximal convoluted tubule, epithelial cell of proximal tubule, apical plasma membrane.

Model: uniprot_1996.cellMTHED1_NHE3_Na

Biological meaning: Flux of Na+ from luminal to cytosol through apical plasma membrane

Species: Rattus norvegicus
Gene: SLC9A3
Protein: sodium/hydrogen exchanger 3

Recommendations/suggestions based on existing models in PMR
Basolateral membrane model
- Sodium/hydrogen exchanger 3 (human)
- Low affinity sodium-glucose cotransporter (mouse)
- Sodium/potassium-transporting ATPase subunit alpha-1 (rat)

Alternative model of sodium/hydrogen exchanger 3
Not Exist

Kidney model in PMR
- Sodium/hydrogen exchanger 3 (human)
- Low affinity sodium-glucose cotransporter (mouse)
Model + Data = Verification?

• Given a semantic description of the protocol used to generate some data
  • Experimental context
  • Simulation experiments
• Known model characteristics (and capabilities?)
  • Query for data that has similar characteristics
• Generate SED-ML to apply dataset’s protocol to model and execute it
  • Can compare model predictions to data
  • Gives some measure of confidence that the model might not be unusable?
CRBM journal curation service

• Work with journals to improve reproducibility
• Help develop common curation practices
  • Domain specific curation vs general curation
  • Curation != validation
• Measuring reproducibility
  • FAIR metrics and associated pitfalls?
• Annotation
  • Non-standard model formats
  • Simulation results (SourceData figure panels?)
Acknowledgements

- Dewan Sarwar
- Koray Atalag
- Bernard de Bono
- Tommy Yu
- Peter Hunter

- ABI Physiome Group