

## TUTORIAL

# BioModels: Content, Features, Functionality, and Use

N Juty<sup>1</sup>, R Ali<sup>1</sup>, M Glont<sup>1</sup>, S Keating<sup>1</sup>, N Rodriguez<sup>2</sup>, MJ Swat<sup>1</sup>, SM Wimalaratne<sup>1</sup>, H Hermjakob<sup>1</sup>,  
N Le Novère<sup>2</sup>, C Laibe<sup>1</sup> and V Chelliah<sup>1\*</sup>

**BioModels is a reference repository hosting mathematical models that describe the dynamic interactions of biological components at various scales. The resource provides access to over 1,200 models described in literature and over 140,000 models automatically generated from pathway resources. Most model components are cross-linked with external resources to facilitate interoperability. A large proportion of models are manually curated to ensure reproducibility of simulation results. This tutorial presents BioModels' content, features, functionality, and usage.**

*CPT Pharmacometrics Syst. Pharmacol.* (2015) 4, 55–68; doi:10.1002/psp4.3; published online 25 February 2015

## BACKGROUND

Mathematical models play an important role in the interpretation and investigation of mechanisms underlying complex biological systems, and thereby in the understanding of human diseases. Using the ever increasing amount of high quality quantitative data produced in the Life Sciences, and experimentally rooted mathematical models and simulations, it is now possible to validate our current understanding of biological processes and to generate meaningful hypotheses for ill-defined processes. This in turn drives further experimental design.

Mathematical modeling of biomolecular processes has become a standard part of the molecular and systems biologists' toolkits. Wide availability and interoperability of models is seen as hugely important in an increasingly collaborative field. However, models covering even the same domain can exhibit much variation since different scientists have developed them at different periods, coming from different perspectives and using different conventions. To increase the reusability of existing published models, there has been a significant drive toward standardization of model encoding, interoperability, distribution, and reuse.

A first step to encourage common development procedures in mathematical modeling was the definition and adoption of standard and machine-readable encoding formats. Efforts in this area include Systems Biology Markup Language (SBML),<sup>1</sup> which has become the most widely used markup language in systems biology. As with other types of scientific research data, the next step was to provide freely accessible repositories of existing models. Such repositories fulfill several functions: models can be deposited to provide access to interested parties, can be directly retrieved for use, or can be used as a starting point for further improvements and refinements. In addition, such model repositories can be viewed as providing reusable components, from which submodels can be extracted to compose novel models, or to aggregate entire models into supermodels.

In order to promote interoperability between models and data, it is necessary to unambiguously identify all model

components. One mechanism is to use standard cross-references<sup>2,3</sup> to external resources that describe the nature of the model component as intended by the model encoder. This allows, for example, the specification of a model element as being a particular metabolite, protein, or gene. Such cross-references facilitate model query, comparison, and processing, such as for model enrichment or merging.

BioModels<sup>4</sup> was created with many of these issues in mind. It is a repository for mathematical models describing biological, biochemical, or biomedical processes. It is easily accessible and allows researchers to re-use, analyze, and investigate the dynamic properties of the model system. While it currently uses SBML as its core model format, one can nevertheless retrieve models in many alternative model encoding formats.

This tutorial describes the domain content, access features, and use of BioModels. It is organized into three main sections as follows: (1) an overview of BioModels' content which provide the details and the generations of two groups of models, (a) models published in the literature and (b) Path2Models, hosted in the database; (2) a comprehensive guide to BioModels Web interface that describes the facilities available to access and retrieve models, functionalities available for each model, and additional services provided to facilitate the use of the repository by the community; and (3) a content usage section, illustrating how models are shared and re-used to answer different biological questions, and other general use of BioModels' content followed by conclusions. For easy navigation through the content of the tutorial, a structural overview is provided as a supplementary material SI-1.

## OVERVIEW OF BIOMODELS' CONTENT

The content of the database can be divided into two categories: a) models that have been described in the scientific literature, and b) models that have been generated through automated processing of pathway resources (Path2Models).<sup>5,6</sup>

The diversity of models within BioModels that are stored in SBML format illustrates that SBML provides a means of

<sup>1</sup>European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK; <sup>2</sup>Babraham Institute, Babraham Research Campus, Cambridge, UK. \*Correspondence: V Chelliah (viji@ebi.ac.uk)

Received 8 August 2014; accepted 21 October 2014; published online on 25 February 2015. doi:10.1002/psp4.3

encoding models from a wide range of fields and disciplines within the broader biological context.<sup>5,7</sup> Accepted as a *lingua franca* for exchange of such models, SBML is supported by a large number of software applications,<sup>8</sup> providing a variety of simulation and analysis techniques, as well as enabling modelers to easily create or expand models without any need to be familiar with the underlying format.

SBML continues to be developed in stages, with a specification released at the end of each development cycle. Major editions of SBML are termed *Levels* and represent substantial changes to the composition and structure of the language. Within each given level, smaller refinements are made and released as new *Versions* of the given level. The latest level currently available is Level 3 (<http://sbml.org/Documents/Specifications>) which is being developed as a modular language, with a central core based on SBML Level 2, and extension packages layered on top of this core to provide additional, optional, sets of features. **Table 1** illustrates the features supported by the different levels and versions of SBML. Although it is usually possible to convert from an older level/version combination to a later one, some of the features introduced in later levels/versions cannot be represented in earlier levels of the format.

The next section describes the content of the database and the process leading to publication of models.

## MODELS PUBLISHED IN THE LITERATURE

The hosting of models described in literature within BioModels was launched in April 2005, and currently contains over 1,200 models described in approximately 160 different organisms. Most models in this category are kinetic models that describe and quantify interactions between biological components. Such models are subjected to extensive curation and annotation procedures.

### Model submission and provenance

Submission of models to BioModels is open to all users. Models can be submitted *via* an online interface and are accepted in two formats: SBML<sup>1</sup> and CellML.<sup>9</sup> On submission, each model is assigned a unique and stable submission identifier (of the form “MODEL” followed by a 10 digit timestamp), which allows users to access and retrieve it. Publication authors can use this identifier to reference their model in BioModels.

Models published in the literature may either be encoded directly by curators or submitted by modelers and author(s) of the publications. Additionally, models were contributed to the resource through collaborations with other model repositories such as the former SBML model repository (California Institute of Technology, Pasadena, CA), JWS Online,<sup>10</sup> the Database of Quantitative Cellular Signalling,<sup>11</sup> and the CellML Repository.<sup>12</sup> Several publishers of scientific journals recommend deposition of models to BioModels in their authors’ submission guidelines. This includes journals from the EMBO press, Public Library of Science, Royal Society of Chemistry, BioMed Central, Elsevier, and FEBS Publishers. As a result, and due to the increasing awareness of

**Table 1** SBML components and their availability in different levels, versions, or packages

	Level 1	Level 2		Level 3	
	Versions 1–2	Version 1	Versions 2–4	Version 1	Package
Compartment	X	X	X	X	
Species	X	X	X	X	
Parameter	X	X	X	X	
UnitDefinition	X	X	X	X	
Rule	X	X	X	X	
Reaction	X	X	X	X	
Event		X	X	X	
Events with priority					X
FunctionDefinition		X	X	X	
InitialAssignment			X	X	
Constraint			X	X	
CompartmentType			X		
SpeciesType			X		
QualitativeSpecies					Qual
Transition					Qual
SubModel					Comp
FluxBound					Fbc
Objective					Fbc
Layout					Layout

Model components that are used in SBML together with the SBML level, version, or package for which these components are available.

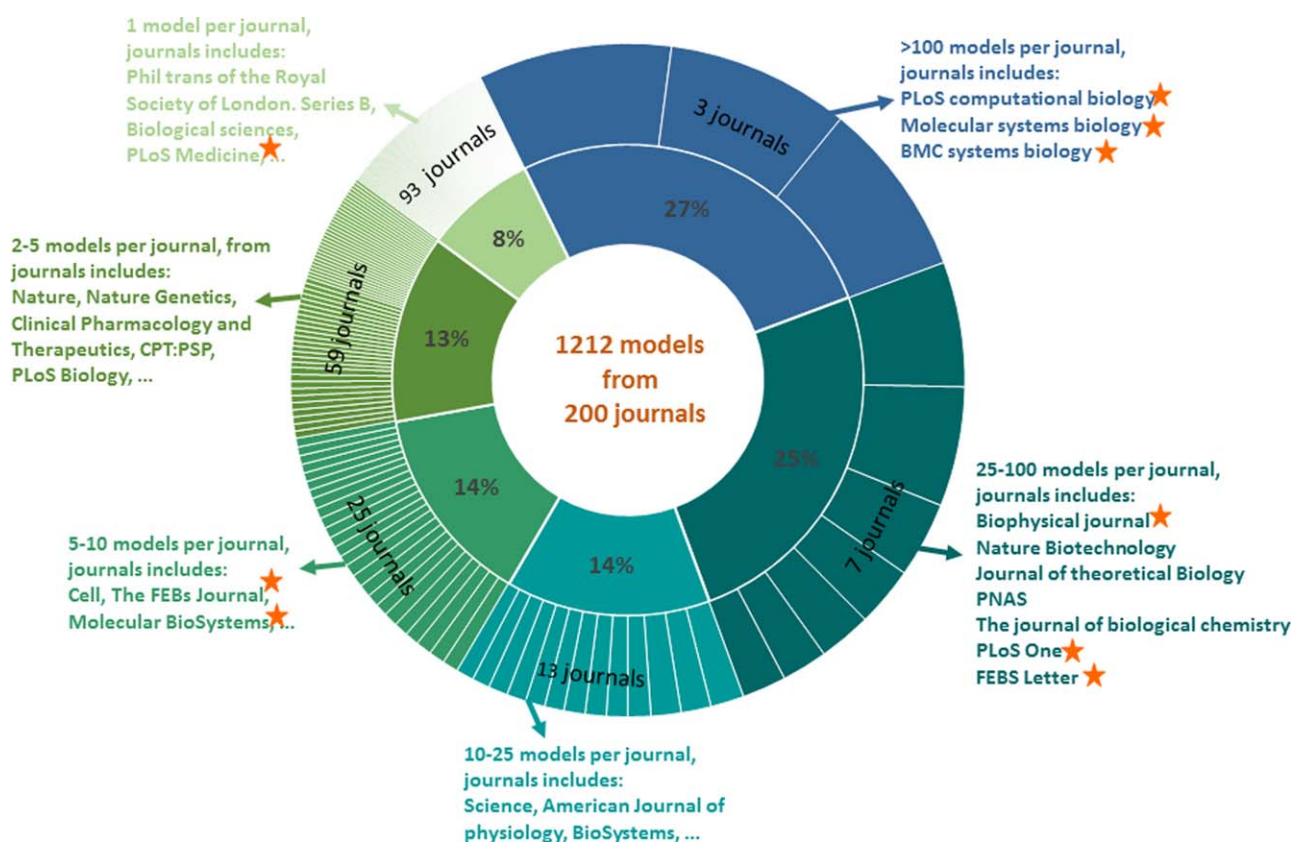
BioModels within the community, the direct submission of models by authors/modellers is increasing rapidly. The total number of literature models reached 1,212 as of September 2014. These models originate from 200 scientific journals (**Figure 1**). Approximately 50% (611) of the models are from just 10 journals, of which 6 journals recommend model submission to BioModels in their authors’ instructions.

A reference to the publication describing the model is always stored within the model file itself. It can be provided at the time of submission either as a PubMed Identifier, a digital object identifier, or as a uniform resource locator. While models may be submitted prior to publication of the associated paper, they are only made publicly available afterwards.

### Curation phase

Following submission, models pass sequentially through several steps in a curation procedure, which ends in their publication through BioModels (**Figure 2A(a)**). Most of the literature models are quantitative kinetic models, while some are qualitative models such as metabolic network, flux balance analysis, logical, and Petri-net based models.

- Models submitted in older Levels or Versions of SBML are converted to the latest Level and Version of SBML, unless such conversion would result in information loss or introduce inaccuracies to the model.
- SBML consistency checks are performed using libSBML.<sup>13</sup> This includes checks for unit consistency and mathematical expressions, etc.



**Figure 1** Source of literature models. As of September 2014, 1,212 models originate from 200 scientific journals. The inner layer of the chart gives the percentage of models that originate from journals listed in the second layer. For example, from the blue segment of the chart it should be understood that 27% of the models come from three journals (i.e., more than 100 models from each of the three journals). Approximately 50% (611 models) of the models are from 10 journals, of which 6 journals recommend model submission to BioModels in their authors' instructions (indicated by stars). This diagram not only illustrates the diversity of journals from which models are obtained, but also the importance of journal's support to the resource through encouraging authors to submit their models to BioModels, during the paper submission process.

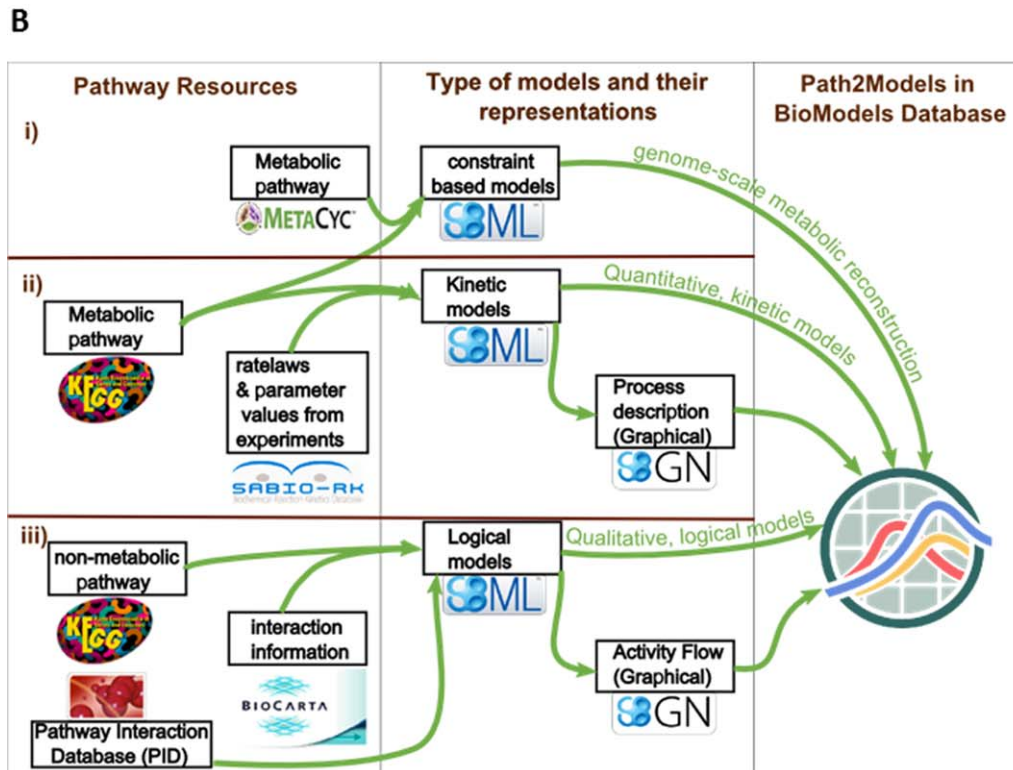
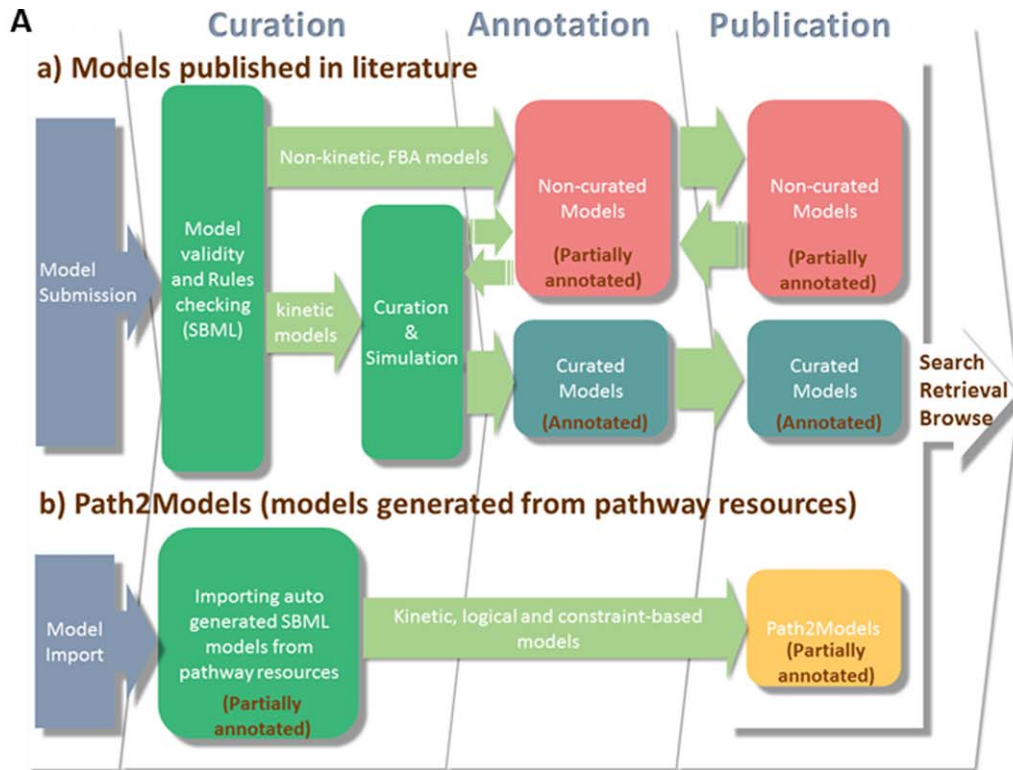
Models at this stage are fully SBML compliant but are then segregated into one of two branches based on the subsequent curation status. Generally, models that satisfy the MIRIAM (Minimum Information Required in the Annotation of Models) guidelines,<sup>14</sup> progress to the curated branch. In particular, these guidelines require that models (1) be encoded in a standard format, (2) be clearly related to a single reference from a scientific journal, (3) correspond to the processes listed in the reference publication, and (4) reproduce the simulation results under the conditions described in the paper. Models that do not satisfy the MIRIAM guidelines are moved to the noncurated branch.

- A large proportion of the kinetic models (45%: 548 out of 1,212 models) are manually curated (fully MIRIAM compliant), a process that includes checking the behavior of the model when instantiated in a simulation. The dynamic behaviors are verified with a tool that is different from the one used in the original paper, thereby precluding tool-specific errors or hidden dependencies. The tools most commonly used for simulation checks are COMPLEX PATHway Simulator,<sup>15</sup> the SBMLodeSolver,<sup>16</sup> SBMLSimulator,<sup>17</sup> or the facilities provided by the Systems Biology

Workbench.<sup>18</sup> A representative figure or table from the paper is reproduced and is made available with the model, together with a description of how it was obtained. If the simulation fails to reproduce the output present in the original publication, the authors are contacted to find the cause of discrepancies.

- The curated models are assigned a curation identifier (of the form "BIOMD" followed by a ten digit number), which reflects when the model was processed; for example, "BIOMD0000000488" denotes the 488<sup>th</sup> successfully curated model. As with submission identifiers, the curation identifiers are unique and stable and can be used for search, retrieval, and referencing the model in BioModels. It should be noted that noncurated models can be accessed by their submission identifier; while curated models can additionally be identified through their curation identifier. A third identifier type is associated with models from the Path2Models project (see "Path2Models" below).
- About 55% of the literature models are in the noncurated branch, of which 38% are quantitative kinetic models, and the remaining 17% are nonkinetic models (highlighted with a network icon in the display interface). These include models of genome-scale metabolic networks, flux balance





**Figure 2** Model production pipeline of BioModels. (A) The diagram schematically represents the model generation pipeline of the two branches ((a) models published in literature and (b) Path2Models) of BioModels. The literature models undergo a sequence of steps from submission until publication in BioModels. Models imported from pathway resources are converted to SBML with additional information (see B for details), and are submitted to BioModels. (B) Detailed representation of the workflow of models generated from different pathway resources. Mathematical features, such as kinetic rate equations and flux bounds are added during the process, along with a graphical description.

analysis, logical models, and Petri-net. Due to their intrinsic lack of kinetic information, these models cannot be validated with respect to their output and reside in the noncurated branch. Some kinetic models remain in this branch for other reasons: (1) Models that can be curated, but remain temporarily in this branch until curators are able to process them. (2) Models do not reproduce the results published in the paper; this may be due to errors in the implementation of the model or due to missing information in the publication. In such cases, authors are contacted for clarification, and models which can be corrected are migrated to the curated branch. (3) Models may be described in a publication, but contain some components that cannot be encoded in SBML.

- Models are given a meaningful name, in accordance with MIRIAM guidelines, using the biological process they address. The convention in BioModels follows the generic scheme “FirstAuthorLastNameYear – Topic.” For example, “Bianconi2012 – EGFR and IGF1R pathway in lung cancer” (BIOMD0000000427),<sup>19</sup> and “Proctor2013 – Effect of A $\beta$  immunization in Alzheimer’s disease” (BIOMD0000000488).<sup>20</sup>

### Annotation phase

In addition to the requirements listed in “curation phase,” the MIRIAM guidelines<sup>14</sup> also require that a model is provided with the contact information of the author(s), the date and time of model creation and last modification, a precise statement about model terms of distribution, as well as cross-referencing model components to external resources (reference databases and ontological records). This additional information is added to the model description at the first stage of the annotation process.

Different model encoders will routinely use different acronyms to represent a given entity, resulting in proliferation of acronyms for the same biological component. For instance, “Glucose” is often labeled “Glu,” “Glc,” “Glucose,” or even “G.” Such an ambiguous nomenclature must be resolved to permit the identification of equivalent entities across models and beyond. The cross-referencing (annotation) of model elements to records from external database resources precisely relates them to the corresponding biological processes or physical entities they represent. This facilitates efficient search and retrieval of the models from the database, and also helps in model comparison, merging and expansion with novel information.

Models in the curated branch are always annotated at the model level and at the level of variables. In some cases, annotations are also added to the parameters and mathematical expressions. Models in the noncurated branch are always annotated at the model level. They may additionally carry other annotations (at the level of physical entities, reactions, parameters, etc.) provided by the authors with the submitted models. These additional annotations may be unchecked by the curators.

Annotations are made using resolvable Identifiers.org Uniform Resource Identifiers.<sup>2</sup> Over 50 external resources and ontologies are used for model annotation purposes. Some of the predominantly used resources include Gene

Ontology (GO),<sup>21</sup> Chemical Entities of Biological Interest (ChEBI),<sup>22</sup> Human Disease Ontology,<sup>23</sup> NCBI Taxonomy, Brenda Tissue Ontology,<sup>24</sup> and UniProt.<sup>25</sup> Any resource that is listed in the Identifiers.org Registry can potentially be used to annotate models.

Additionally, SBML controlled annotations use qualifiers (<http://co.mbine.org/standards/qualifiers>). These qualifiers define the relation between the external resource and the model component. More details on this annotation scheme can be found in section 6 of the SBML specifications (<http://sbml.org/Documents/Specifications>). In addition to the SBML supporting tools offering access to annotations, curators have access to a purpose-built annotation facility, within the BioModels Web interface.

Annotations for models from the Path2Models portion of BioModels (**Figure 2A(b)**) are handled differently. This is described below in the “Path2Models” section, together with the process that generates these models (**Figure 2B**).

### Model publication

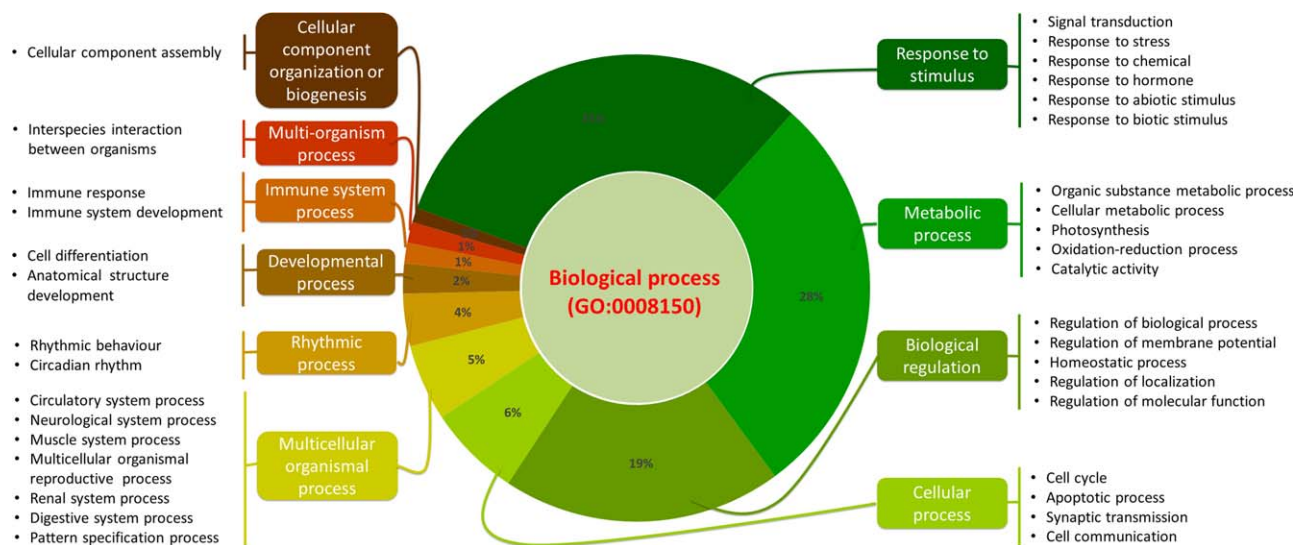
Following the curation and annotation phases, models are ready to be published and a notification is sent to the submitter. Models then become publicly accessible through BioModels. All models available in BioModels are provided under the terms of the Creative Commons CC0, Public Domain Dedication. This means that the encoded models are available freely for use, modification, and distribution, to all users.

### Classification of models based on GO terms

The annotations present in the models facilitate their efficient search and retrieval. All models are annotated with GO terms at the model level that represent, as accurately as possible, the biological processes that the model describes. In order to provide a broader level of classification of models based on these precise GO terms, three increasingly generic ancestor terms are chosen. In all cases, the third (broadest) term is chosen to be one of the GO root term “Biological Process” (GO:0008150), which applies to all models. This type of hierarchical classification (from generic to most specific terms, through four GO terms) allows users to narrow down their search to specific biological processes of interest. The three generic category terms are not stored in the model itself, but rather stored in the database. **Figure 3** shows the GO classification of the models in both the curated and noncurated branch. The top three GO terms used for the classification are shown, whereas the fourth term is the specific GO term used in the models. For example, for BIOMD0000000488, the corresponding hierarchical classification begins with the most general GO term “biological processes” (GO:0008150), followed by “cellular component organization and biogenesis” (GO:0071840), “cellular component assembly” (GO:0022607), and ends with the most specific term “inclusion body assembly” (GO:00470841) that is used in the annotation of the model. This allows categorized searching of models by following the GO term classification chart, and this facility is available through the BioModels website.

### Record of model evolution

Wherever appropriate, literature models are cross-linked with the parent model(s) from which the structure and/or equations are derived or adapted. This information is stored



**Figure 3** GO classification of models. All models are annotated with at least one GO term at the model level to represent the biological process it describes. In order to provide a broader level of classification of models based on these precise GO terms, three increasingly generic ancestor terms were chosen, the third term being one of the GO root term “Biological Process” (GO:0008150), which applies to all models. This type of hierarchical classification (from generic to most specific terms, through four GO terms) allows users to narrow down the search to specific biological process of interest. The top three GO terms used for the classification are shown, whereas the fourth term is the specific GO term used in the models. This allows categorized searching of models by following the GO term classification chart, and this facility is available through the BioModels website.

in the SBML file using the qualifier “bqmodel:isDerivedFrom” and the identifiers from BioModels. Where BioModels identifiers are not available, for instance if the parent model has not yet been added to the database, the identifier for the publication (PubMed identifier or digital object identifier) in which the model is described is used.

This model lineage derivation shows currently over 150 clusters encompassing 819 models in total, with individual clusters ranging from 2 to over 90 models. A total of 561 of the models are already present in BioModels. **Figure 4** shows the largest network of signaling pathway models. This network presents models describing multiple signaling pathways downstream a diverse range of receptor types, including ErbB, FGFR, NGFR, Insulin receptor, NMDAR, and AMPAR. This shows the importance of sharing and reuse of existing models to generate more elaborate models. For example, models “Bidkhor2012 - normal EGFR signaling” (BIOMD0000000452) and “Bidkhor2012 - EGFR signaling in NSCLC” (BIOMD0000000453), describe the cross-talk mechanism between Ras/ERK, PI3K/AKT, and JAK/STAT pathways in normal and non-small-cell lung cancer (NSCLC) cells,<sup>26</sup> respectively, and are both derived from several other models.

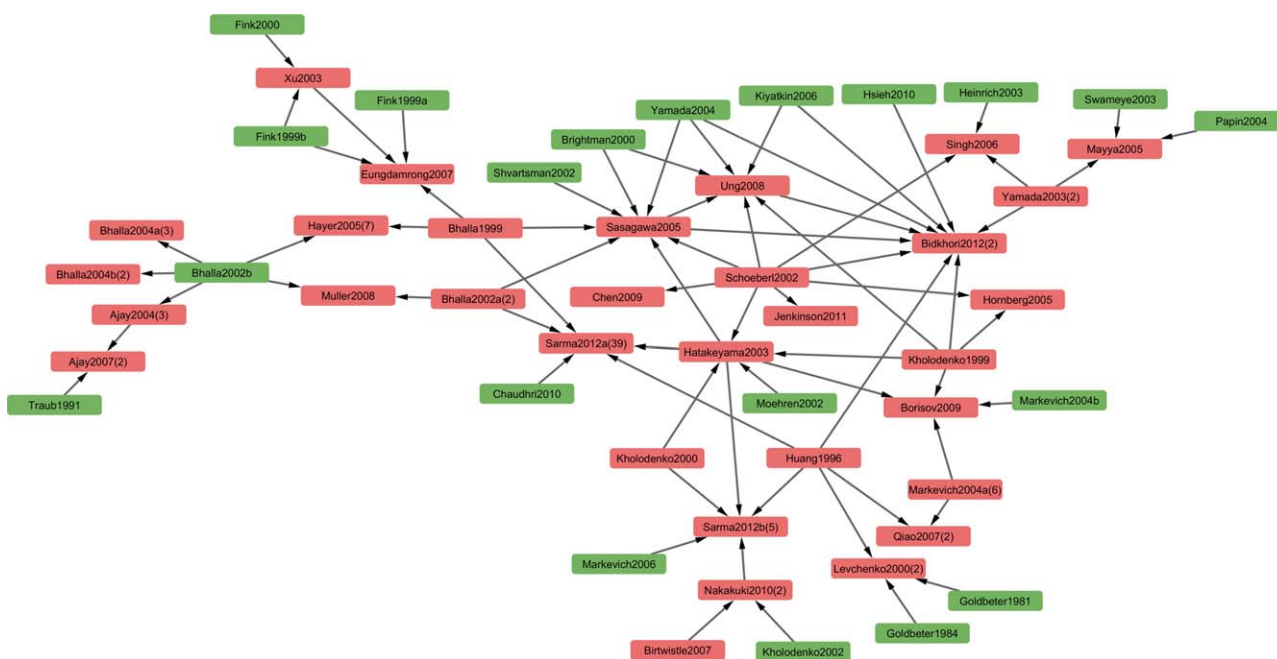
**PATH2MODELS (MODELS DERIVED FROM PATHWAY RESOURCES)**

The Path2Models<sup>5,6</sup> section in BioModels provides computational models derived from pathway resources (**Figure 2A(b)**). The pathway information from data resources such as the Kyoto Encyclopedia of Genes and Genomics,<sup>27</sup> BioCarta/Nature Pathway Interaction Database,<sup>28</sup> MetaCyc,<sup>29</sup> and SABIO-RK<sup>30</sup> are automatically converted into standard-

compliant computational models. Multiple biological processes, described in 2,616 organisms, were converted into computational models encoded in the SBML format, some of which are additionally available in SBGN-ML format.<sup>31</sup> Depending upon the source data, three types of models have been generated: (i) Genome-scale metabolic reconstructions: these models are generated by extracting pathway data from the Kyoto Encyclopedia of Genes and Genomics and MetaCyc databases and subjected to flux balance analysis; (ii) quantitative, kinetic models of metabolic pathways: these models are generated based on the metabolic pathways distributed by the Kyoto Encyclopedia of Genes and Genomics and described as processes, in combination with experimentally determined rate laws and parameter values from the SABIO-RK database; (iii) qualitative, logical models of nonmetabolic (primarily signaling) pathways: these models are generated based on the non-metabolic pathways distributed by the Kyoto Encyclopedia of Genes and Genomics and Pathway Interaction Database, with additional information from BioCarta pathways. **Figure 2B** illustrates the workflow used for the construction of the above mentioned three types of models from pathway resources.

All of these models are semantically annotated according to the MIRIAM guidelines: the model components (metabolites, genes, proteins, enzymes, reactions, etc.) are cross-linked with external data resources and controlled vocabularies, akin to the process for the literature based model, although assignments are made automatically based on the information present in the source data resources. A third type of BioModels identifier has been introduced to distinguish these automatically generated models from the literature models of the form “BMID”





**Figure 4** Importance of sharing and re-use of models. Clusters of models created based on the qualifier “bqmodel:isDerivedFrom” (see text for more information). The largest clusters are formed by models of signaling pathways. This diagram shows how signaling models are shared and re-used by different groups to develop more comprehensive models. The network is formed by the cross-talks of a range of signaling pathways activated downstream of ErbB, FGFR, NGFR, Insulin receptor, NMDAR, and AMPAR. The arrows (→) are used to link models that are derived/adopted from one or more parent models. Models available from BioModels are highlighted in salmon, otherwise in green. The models are named with the first author’s surname followed by the year of publication. For some models, Author’sLastNameYear is followed by a number in brackets denoting the number of models in BioModels coming from the same publication. For example, “Bidkhori2012(2)” means that two variants of the model are described in the paper and both are available in BioModels. The two models “Bidkhori2012 - normal EGFR signalling” (BIOMD0000000452) and “Bidkhori2012 - EGFR signaling in NSCLC” (Bidkhori *et al.*, 2012, BIOMD0000000453), describe the cross-talk mechanism between Ras/ERK, PI3K/AKT, and JAK/STAT pathways in normal and NSCLC non-small-cell lung cancer (NSCLC) cells, respectively. As can be seen from the figure, the model is itself derived from several other models.

followed by 12 digits. The models in this branch are provided as ready-made models that can be used as initial starting models, where the kinetic data from experiments can be introduced for further simulation and analysis.

### Comprehensive guide to BioModels’ web interface

Through its rich graphical user interface, BioModels provides a variety of methods to locate and access specific models, and to download them in a variety of formats. Model information is displayed in various ways, such as maps and textual reports. One can perform several analyses, for instance simulating the model and extract model components as submodels. Moreover, since each model is enriched with reference information drawn from diverse biologically relevant data resources, BioModels can be regarded as a Web portal for the modeling domain.

### Browse content

The BioModels front page provides the details of the models currently available. A generic browse feature allows to list models in each of the BioModels branches (curated, noncurated, and Path2Models).

Curated and noncurated models can also be browsed using the GO terms that are associated with individual models. This information is stored in the SBML file itself, and can be viewed in the interface from the “Model tab”

(see “Model Display page” below). An interactive chart of GO terms gives access to increasingly more precise terms used in the annotations of models (see *Classification of models based on GO terms* for details). An alternative way is to search models using GO terms, where the information is displayed as a tree. The expandable branches highlight the total number of models available in each category as a parenthesized value. This feature operates using a condensed list of GO terms predetermined to be relevant to BioModels.

Models in the Path2Models branch are listed in categories, according to the type of model (metabolic, nonmetabolic, or whole genome metabolic models). Whole genome metabolic models are listed by species, while metabolic and nonmetabolic models are associated with particular GO terms, and subsequently divided into lists, ordered by species.

### SEARCH AND RETRIEVAL OF MODELS

Models can be retrieved using their assigned identifiers. However, since a user may not know or recall this identifier, BioModels also incorporates a search engine for locating specific models. A simple search can be launched from any

page, with an advanced search facility also available. In the execution of a search, three main information sources are utilized. Used sequentially: (i) metadata about the model (high level model information, including authors and publication information); (ii) information contained in the model itself, e.g., present in the “notes” element of the SBML files; and (iii) cross-reference information present in the models (whenever possible the search query is expanded using additional information fetched from external data resources).

Postprocessing of search results is used to improve the quality of information provided to the user. For instance, taxonomical search queries are expanded to take into consideration the relationships between taxons; a search for “mammalia” will additionally retrieve models annotated with “Homo sapiens” and “Mus musculus,” since they are descendants of the original query term.

In the next sections, which demonstrate the various features of the resource, we use the example of a user who has launched a query for “Alzheimer’s disease” and decided to further explore one of the models listed in the search results: “BIOMD000000488 (Proctor2013 - Effect of  $A\beta$  immunization in Alzheimer’s disease).”<sup>20</sup>

## MODEL DISPLAY PAGE

Once a model has been selected, either through browsing, or identified through search, a model page is displayed. Model components of manually curated models are organized, interlinked, and presented using tabs (**Figure 5**).

### Model tab

The default view when accessing a model, the *model tab* displays information about the model as a whole. It provides core information, such as the model name, publication information, link to the original file (original submission, either in SBML or an alternative format that is used as the basis to write the SBML), names of the encoder(s) and submitter, submission and modification dates, etc. In addition, this view also provides the most relevant GO terms (in this case “Inclusion body assembly”), reflecting the biological processes represented by the model, together with the taxonomic range (here “Mammalia”) to which the model may be relevant. The human readable notes provided by the submitters and curators are also displayed. In this case, the model is also annotated with specific disease information (Alzheimer’s disease).

Upon further examination, it is apparent that this model (BIOMD000000488)<sup>20</sup> is derived from earlier versions BIOMD000000286<sup>32</sup> and BIOMD000000462.<sup>33</sup> The annotation states this relationship using the qualifier “bqmodel:isDerivedFrom.”

### Overview tab

This summary tab lists all the components of the model, whether they are physical entities, individual parameters, mathematical relationships between entities, or defined events. Each component is linked to a detailed view in the corresponding tab listed below. Selection of a component in the “Overview” acts as a shortcut, redirecting the user to a

specific tab, where the selection made is highlighted, for convenience.

A user may wish to explore only part of a model, or use this part to create a submodel. Users can, therefore, select the components of interest. The other tabs subsequently only display the relevant parts of the model. For instance, if a molecular species is selected, only the reactions in which this species participates are displayed in the *math tab*. The user can then click the “Create a submodel with selected elements” link. This generates an additional tab in the viewing area, entitled “Submodel  $n$ ,” where  $n$  denotes the  $n^{\text{th}}$  number of submodel that is created. The resulting submodel is the minimal valid SBML model that encompassed all the selected components and the relevant mathematical relationships. This model can then be downloaded.

### Math tab

The *math tab* lists the mathematical relationships linking model components. These mathematical constructs cover reactions, events, and explicit mathematical formulae (SBML rules). Each construct may be provided with further details, including any compartment restrictions, parameter value where applicable, a rendered mathematical equation, and optional cross-references to relevant Systems Biology Ontology (SBO) terms<sup>3</sup> or GO terms.

### Physical entities tab

The *Physical entities tab* lists the physical or conceptual entities described in the model. Entities consist of the compartments and various entity pools (protein, chemical, etc.) that occupy specific compartments (cell, mitochondria, etc.). Thus a chemical species present in several compartments is considered as a set of different entities. Each entity may be listed with its initial size and cross-references to external databases, specifying the identity of the entity, and other annotations to specify the nature of the entity in the context of the model (protein complex, ligand, etc.).

### Parameters tab

The *Parameter tab* provides the list of every parameter used in mathematical relations. Parameters are grouped according to their scope (global or reaction specific), and where applicable, linked to the appropriate section of the Math tab. Parameter values for few models are also provided, potentially annotated with ontology terms to describe the type of parameter (dissociation rate constant, forward rate constant, etc.).

### Curation tab

The curation tab describes the result of the reproduction step of the curation process. It provides a representative plot generated by the curators reproducing one or more results present in the original paper. In addition to the typical graph, there may be some clarifications or amendments to the published procedure, which were found necessary to obtain correct simulation results. If changes are made to the model in order to reproduce the simulation figures, the changes are described and relevant additional files are provided. For an example, see the curation tab of the model “Lemaire2004 – Role of RANK-RANKL-OPG pathway in bone remodeling process” (BIOMD000000278).<sup>34</sup>



The screenshot shows the BioModels Database interface for model BIOMD000000488. At the top, there are navigation links for Services, Research, Training, and About us. The main header includes the BioModels Database logo and a search bar. Below the header, there are tabs for Model, Overview, Math, Physical entities, Parameters, and Curation. The main content area is divided into several sections: Reference Publication, Model, and Notes. The Model section contains a table with columns for set number, property, and value. The Reference Publication section includes the title, authors, and a link to the full text. The Notes section contains an abstract and a description of the model.

set #	Property	Value
set #1	bioModel isDerivedFrom	BioModels Database Proctor2010 - a link between GSK3 and p53 in Alzheimer's disease BioModels Database Proctor2012 - Role of Amyloid-beta dimers in aggregation formation
set #2	qbciol isVersionOf	Gene Ontology DNA damage response, signal transduction by p53 class mediator Gene Ontology inclusion body assembly
set #3	qbciol isPartOf	Pathway Ontology PW:000015
set #4	qbciol hasTaxon	Taxonomy Mammalia
set #5	qbciol hasProperty	Mathematical Modelling Ontology MAMO_0000046
set #6	qbciol hasVersion	Human Disease Ontology DOID:10652

**Figure 5** Screen image of a “Model” page. This displays the landing page of “Proctor2013 – Effect of A $\beta$  immunization in Alzheimer’s disease” (BIOMD000000488). The page displays model download formats (Download SBML, Other formats (autogenerated)), links to features such as Online simulations (Actions), distribution of model components that span various tabs (Model, Overview, Math, Physical entities, Parameters, Curation), and model elements cross-linked (annotated) with various external resources such as Gene Ontology terms and Human Disease Ontology.

### Model download

The models in BioModels are stored internally as SBML files that can be downloaded from each model page. Since the SBML standard has evolved over time, it is possible that the original submission file may be in a form that has been superseded. Since sequential SBML formats are largely backwards-compatible, it is possible to automatically convert a model between these alternatives. Users are, therefore, also able to download curated models in many alternative SBML Levels and Versions. The file used to generate the curation figure is highlighted.

Users can also download models in other formats. These alternative formats are produced from the SBML using converters developed for this purpose. This includes tool specific formats such as XPP,<sup>35</sup> Octave (MatLab m-file; <http://www.gnu.org/software/octave/>), SciLab (<http://www.scilab.org/>), and VCML.<sup>36</sup> In addition, there are converters to generate other standard formats, such as BioPAX (<http://www.biopax.org/>). We also provide a PDF report, providing human-readable summaries of model components and structure.<sup>37</sup>

In addition to individual model files, a bulk download of repository content is also available. Model archives are generated for each release and are provided *via* the EBI FTP server (<http://ftp.ebi.ac.uk/pub/databases/biomodels/releases/>). These provide access to only the SBML files of the literature models, all the model files (all formats in addition to the SBML files) of the literature models, all the RDF/XML files for the curated models, and all models from Path2Models (in three archives for the metabolic, nonmetabolic, and whole genome metabolism models). In addition to the full database releases, some of the archives are automatically generated on a weekly basis.

### Actions button

Each model page permits execution of a number of actions, for example displaying various views on the model, or the capacity to instantiate a simulation.

The reaction graph of the model network are displayed in SVG and PNG. This view is generally created automatically, directly from the SBML file. In some cases, the graph drawn automatically are improved manually.

Users can also launch simulations directly for all curated models on BioModels' infrastructure. This opens a popup window that permits the selection of the SBML species to be displayed in the results, the simulation time, and the time steps. Simulation jobs are sent to the EBI compute cluster, with a link given to retrieve the results. Both the numerical results from the simulation, and the graphical output are available to download. For a subset of curated models, simulation *via* JWS Online<sup>10</sup> is available.

### User feedback

BioModels encourages feedback from its users, and provides a number of avenues for this purpose. A "Send feedback" link is present on each model display page to post problems with this specific model. Furthermore, the "Contact us" button in the main panel bar provides links for:

- Email support related to the services provided (Web services, user interface)
- Email support related to repository content (model error reporting, update requests)
- Discussion list (general modeling community discussions)

Archives of the messages are hosted on SourceForge for service and discussion lists. In addition, SourceForge also hosts a ticket system (<http://sourceforge.net/projects/biomodels>), where bug reports can be provided, and new features requested.

### MODEL OF THE MONTH

"Models of the Month" (<http://www.ebi.ac.uk/biomodels-main/modelmonth>) are short articles showcasing selected models from the curated branch of the repository. They include some introductory information on the subject of the model, placing it into biological and theoretical contexts, and discuss the results and significance of the model simulations presented. Such articles can be used for teaching purposes, and to make models more easily accessible for newcomers in the field of bio-mathematical modeling.

### PROGRAMMATIC ACCESS TO MODELS

#### Web services

An alternative way to access BioModels information is through Web services.<sup>38</sup> Details on the available services can be found at <http://www.ebi.ac.uk/biomodels-main/webservices>.

Using these Web Services, it is possible to retrieve extensive information from the model repository. The method "getModelSBMLById" retrieves the SBML file of a given model using its unique identifier. This can be used, for instance, to get the "Alzheimer's disease" model using getModelSBMLById("BIOMD000000488"). This method accepts both the submission ("MODEL") and publication ("BIOMD") forms of identifiers. If neither identifier is known to the user, it is also possible to use the methods "getModelByIdByName," "getModelByIdByPerson," and "getModelByIdByPublication" to retrieve a list of model identifiers when the model name, one of the authors, or a publication identifier (PubMed identifier or digital object identifier) are

known. Furthermore, to assist users in the retrieval of models specific to a particular subject area, it is also possible to use the Web services to list models associated with a particular Gene Ontology term: the methods "getModelldByGO" and "getModelldByGOld" can be used to perform a query using a Gene Ontology term such as getModelldByGO ("inclusion body assembly"), or the GO term identifier, getModelldByGOld("GO:0070841"). Similar methods are available for other types of resources. These services can also be used to generate a sub-model, given a model identifier, and specific SBML model component identifiers.

### SPARQL endpoint

BioModels SPARQL endpoint (<http://www.ebi.ac.uk/rdf/services/biomodels/sparql>) provides access to the content of BioModels as a linked dataset. This consists of an RDF representation of all curated models from the literature, comprising 174,156,735 triples, and 34,311,872 cross-references pointing to 2,781,565 different biological concepts (as of September 2014). It is part of a wider EBI effort to provide access to data using semantic Web technologies: the EBI RDF platform.<sup>39</sup>

Bioinformatics resources cross-reference each other and have direct links to ontological terms. These mappings can be used to construct queries which retrieve and integrate information from different linked data resources. For example, one can query for ChEMBL protein targets present in a model such as BIOMD000000488 and obtain the drug compound that interacts with it, while no direct reference to ChEMBL is recorded in this model. This is achieved by writing one query which integrates data from BioModels and ChEMBL *via* the common cross references to UniProt proteins (see "One Example Use Case" and Supplementary material SI-2 for more details).

### MODEL FORMATS CONVERTERS

The Systems Biology Format Converter framework (SBFC; <http://sourceforge.net/projects/sbfc/>) provides a generic framework into which many format manipulation tools can be incorporated. The objective is to allow conversion of models between the different file formats used in Systems Biology, many of which are tool-specific. Many such converters have been created historically by different investigators. SBFC, implemented in Java and available as a standalone program, provides interfaces where new and pre-existing converters can be minimally modified, and plugged in. It also allows sequential "chaining" of individual converters to allow conversion between input/output combinations that were not previously possible. BioModels relies on SBFC to generate exports in alternative formats.

### CONTENT USAGE

#### Building blocks for developing new models

*One example use case.* Models distributed by BioModels can serve multiple purposes, either as a body of knowledge on existing processes or as building blocks for further

development. As new mechanistic details and experimental data become available, existing models can be extended/modified to incorporate these new hypotheses. The following use case describes components of various models that have been extended progressively to describe biological problems in a different context. Where the referenced model is available in BioModels, its accession number is provided along with the publication from which it was derived.

The shaded region in **Figure 6** illustrates the development history of the model BIOMD000000488<sup>20</sup> through various intermediate models from the same research group. This model (Proctor2013) includes the key processes involved in passive and active immunization against Amyloid-beta ( $A\beta$ ) and demonstrates the immunological effect on  $A\beta$  and tau-protein aggregation, the major landmarks of Alzheimer's disease. The lineage of this model (**Figure 6**, shaded region) begins with BIOMD000000091<sup>40</sup> (Proctor2005), which describes the fate of misfolded proteins under stress and age-related conditions. It describes the role of Heat Shock Protein 90 and its autoregulation, while the degradation and aggregation processes are simplified in a single step. Molecular details on ubiquitin-mediated proteolysis and aggregation processes were incorporated in BIOMD000000105<sup>41</sup> (Proctor2007). Subsequent experimental studies suggested improvements to this model; glycogen synthase kinase 3 was shown to play a major role in Alzheimer's disease,<sup>42</sup> with an additional potential role for p53.<sup>43</sup> This link was explored in BIOMD000000286<sup>32</sup> (Proctor2010a), using components from BIOMD000000105 (Proctor2007) and BIOMD000000188<sup>44</sup> (Proctor2008).

Predictions from this model (Proctor2010a) suggested, among other things, that stress mediated positive feedback between glycogen synthase kinase  $3\beta$  and p53 led to an increase in  $A\beta$  production. Integrating this model with one that described the role of  $A\beta$  dimer levels in the initiation of the aggregation process BIOMD000000462<sup>33</sup> (Proctor2012), generated the model BIOMD000000488<sup>20</sup> (Proctor2013), which describes the effect of  $A\beta$  immunization in Alzheimer's disease. This model introduces a variable to represent  $A\beta$  immunization, and explores the effect of this intervention on soluble  $A\beta$ , plaques, etc.

Besides their contribution towards the evolution of BIOMD000000488 (Proctor2013), components of BIOMD000000091 (Proctor2005), BIOMD000000105 (Proctor2007) and BIOMD000000188/9 (Proctor2008(2)) are used as building blocks for developing various other models<sup>45–48</sup> (unshaded region in **Figure 6**).

*Possible extension using biomodels SPARQL endpoint search.* There is an increasing number of research articles where the mechanism underlying the treatment (side-effect, resistance to drug compounds, etc.), are investigated using Systems Pharmacology modeling approaches ("Demin2013 – PKPD behavior - 5-Lipoxygenase inhibitors" (BIOMD000000490),<sup>49</sup> "Faratian2009 – Role of PTEN in resistance to trastuzumab" (BIOMD000000424),<sup>50</sup> Palmer *et al.*<sup>51</sup>), which provide new directions in the design of better therapies. A recent multiscale systems model-based analysis of interleukin-6 mediated immune regulation in

Crohn's disease suggested that targeting dual components of the pathway is more effective than targeting a single component.<sup>52</sup>

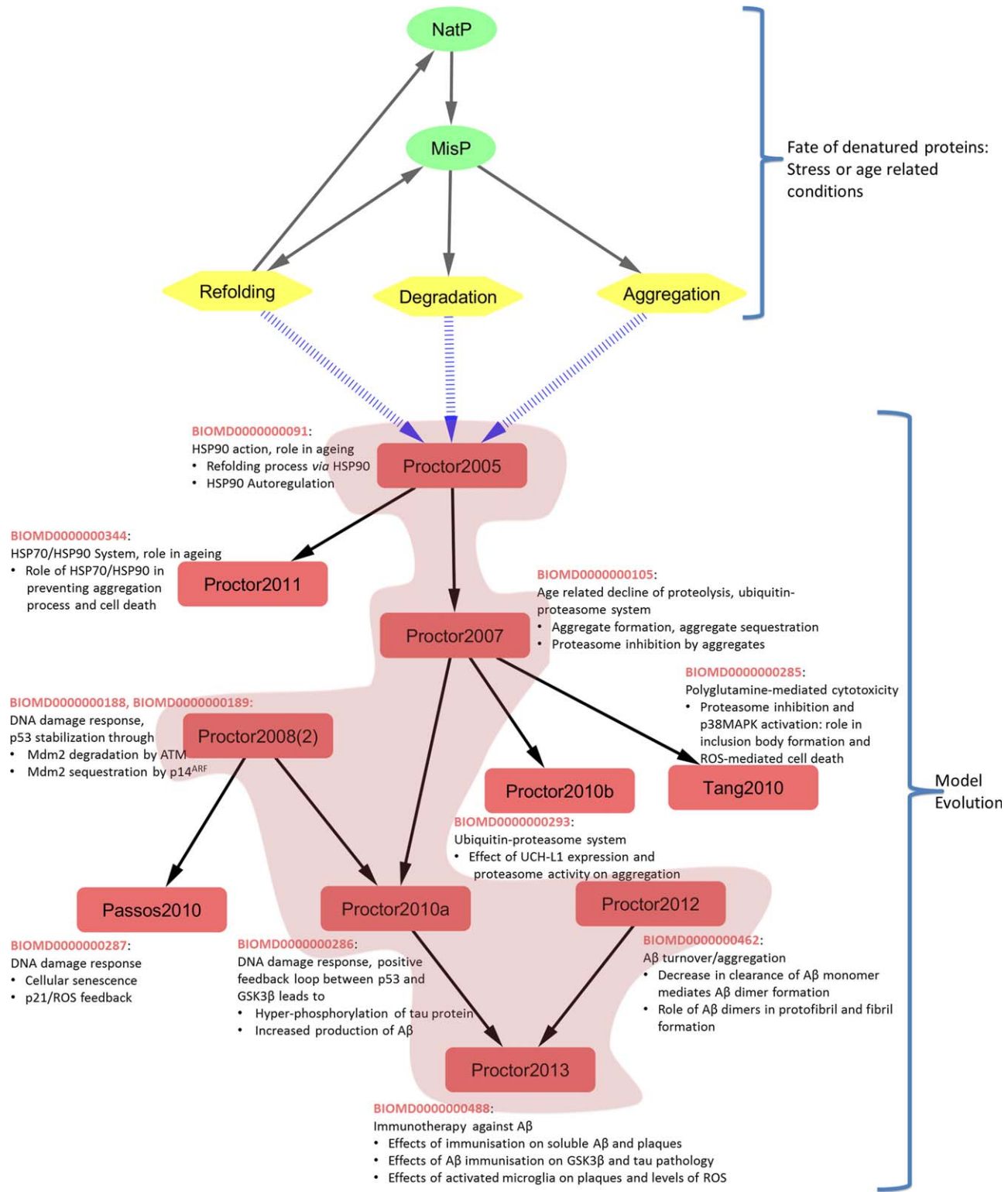
One possible direction to understand the process of aggregation and the effect of treatment could follow these lines. BIOMD000000488 predicted that the antibody against  $A\beta$  reduces the aggregation of both  $A\beta$  and tau, and that immunotherapy may not be effective unless it is performed at very early stages of disease progression, or combined with other therapies. Extrapolating these ideas, one way of extending this model would be to obtain the drugs for the protein molecules that are used in the model and study the effects of each drug or combinations of drugs on the aggregation process (disease progression). ChEMBL<sup>53</sup> is a large-scale bioactivity database, where one can use the bioactivity data to find associations between targets and drugs (which have been tested in the literature). BioModels SPARQL endpoint can be used to fetch the available drug compounds for the protein molecules used in the model. The query result may be filtered based on the half-maximal response concentration/potency/affinity and the current clinical development phase of the molecule (see Supplementary Materials S1 and S2, which are available online, for model details). Querying with varying combinations of these two parameters provides the potential target molecules in the model and their interacting compounds from ChEMBL.

#### Advanced use cases

Besides its established, historical function as a central repository for mathematical models, BioModels is itself an input source for a variety of novel works. For example, models have been used as the input for workflows: qualitative models can, through a series of sequential operations involving Internet-accessible resources, be annotated to external resources, parameterized using experimental data, and processed to optimize assigned parameter values and generate results in a format-agnostic manner. Additionally, the semantic annotations can be used to identify components across which different models can be merged.<sup>54</sup> Furthermore, it is also possible to analyze the individual model components, based on their semantic annotations, across the set of all models. This allows the clustering of related models (by shared components), and can be used to further assist in efficient search strategies.<sup>55</sup> Semantic information can also be used to generate human-readable and queryable forms of curated models.<sup>56</sup> BioModels content is also used to facilitate the standardized benchmarking of tools used for model simulations.

In conclusion, mechanistic models are becoming one of the cornerstones of computational life sciences. They allow us to understand the mechanisms underlying biological observations, make predictions, and test hypotheses. For example, modern drug discovery and design increasingly rely on mechanistic models. It offers insight into the nature of the complex PK and PD processes and offers a number of additional advantages, such as reduced extent of animal testing, and potentially increased accuracy of efficacy prediction and toxic effects. With the growing number and





**Figure 6** Evolution of BIOMD0000000488: Amyloid-beta (Aβ) immunization (Proctor *et al.*, 2013). Under stress or age-related conditions, a correctly folded native protein (NatP) can become misfolded (MisP). There are three possible outcomes for misfolded proteins; the protein may be refolded into its native state with the help of chaperones such as HSP90, may be degraded and removed from the system, or may aggregate together. Such aggregates are the main characteristics of most neurodegenerative diseases, such as Alzheimer's disease and Parkinson's disease. The models in this figure describe to varying extents these three mechanisms. The arrows (→) are used to link models that are derived/adopted from one or more parent models. The models are named as described previously (Author'sLastNameYear) and numbers in parentheses denote the number of models in BioModels which result from the same publication. The mechanism described in each model is recorded adjacent to the model name along with its BioModels Identifier. The shaded portion in the figure depicts the model lineage for "Proctor2013" (BIOMD0000000488), which was built using components of previous models. Models in the unshaded regions do not contribute to the lineage of Proctor2013, but may be relevant for the lineage of other models. See text for further information. GSK3, glycogen synthase kinase 3; HSP, heat shock protein; ROS, reactive oxygen species.

complexity of such models published in the scientific literature, it becomes necessary for these models to be deposited and preserved in a central repository in standard formats, and BioModels serves this purpose.

Models describing a wide range of biological processes at different biological scales are available through BioModels. Besides the usual curation tasks, efforts to collate, encode, and curate mechanistic models focusing on specific disease areas, are being undertaken. This has, for instance, resulted in the targeted curation and recent review on diabetes models.<sup>57</sup>

BioModels, with its vast selection of physiologically and pharmaceutically relevant mechanistic models, could play a prominent role in clinical domain. However, unlike models in Systems Biology (which are typically concerned with individual experimental data), for the drug discovery process, clinical trial results often manifest as population data which requires statistical modeling. One of the standard methodologies involves the use of nonlinear mixed effects modeling. Despite its statistical nature, at the very core of most nonlinear mixed effects models lies a deterministic prediction model, which in this context is called the “structural” model. The new XML based language, PharmML (Pharmacometrics Markup Language; <http://www.ddmore.eu/pharmml>), which is under active development by the DDMoRe consortium, provides the facility to encode the structural model with SBML. This will allow models from BioModels to be used for clinical data analysis.

So far, BioModels has mainly focused on models encoded in the SBML format, and the current software infrastructure is limited to SBML models. However, with the evolution of the modeling landscape, new challenges have arisen including the creation of new model formats to support specific domains. Therefore, following several queries from authors about submitting models in non-SBML formats, we are working to facilitate the hosting and distribution of non-SBML models from BioModels. Wherever possible, users are encouraged to submit models in COMBINE archive format (<http://co.mbine.org/documents/archive>). Moreover, an improved software infrastructure, JUMMP, being developed will allow the resource to widen its scope to contend with this issue in an effective manner. JUMMP is being developed collaboratively and is hosted on Bitbucket (<https://bitbucket.org/jummp/jummp/>).

Hosting models encoded in standard formats (such as SBML), the resource has grown substantially since its introduction in 2005. As of April 2014, it has expanded approximately 20-fold in content, with average model complexity rising fivefold over the same period ([http://www.ebi.ac.uk/biomodels-main/static-pages.do?page=release\\_20140411](http://www.ebi.ac.uk/biomodels-main/static-pages.do?page=release_20140411)).

The increasing number of journals describing models submitted to the resource (almost 200 journals) also indicates the growing awareness of BioModels by modelers in life sciences. The resource is accessed predominantly from Europe, North America, and Asia, serving over 65,000 unique visitors, with almost a million page views per year. In the past two years, content download from our FTP server has risen 10-fold.

BioModels, with its vast collection of biologically and biomedically relevant mechanistic models, serves as an

invaluable resource for academic research, and for pharmaceutical and biotechnological industries.

**Acknowledgments.** This work received support from the Biotechnology and Biological Sciences Research Council (BioModels, BB/J019305/1; Identifiers.org, BB/K016946/1), the Innovative Medicines Initiative Joint Undertaking (DDMoRe, 115156), the European Commission (ISBE, 312455; AgedBrainSYSBIO, 305299), the National Institute of General Medical Sciences (SBML, R01 GM070923), the Babraham Institute and the European Molecular Biology Laboratory. We thank Mark Davies in EMBL-EBI ChEMBL group for helping us to query our models against ChEMBL using the EBI RDF platform.

**Author Contributions.** N.J. and R.A. contributed equally to this work (N.J.: curation and writing of the manuscript; R.A.: software development and technical support). R.A., M.G., S.K., N.R., S.M.W. and C.L. contributed to the tools, software infrastructure and technical support. V.C. contributed to curation and performed model analysis. NJ, MJS, SMW, SK, CL and VC wrote the manuscript, with HH and NLN contributing suggestions as well as overall project guidance.

**Conflict of Interest.** The authors declared no conflict of interest.

- Hucka, M. *et al.* The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531 (2003).
- Juty, N., Le Novère, N. & Laibe, C. Identifiers.org and MIRIAM registry: community resources to provide persistent identification. *Nucleic Acids Res.* **40**, D580–D586 (2012).
- Courtot, M. *et al.* Controlled vocabularies and semantics in systems biology. *Mol. Syst. Biol.* **7**, 543 (2011).
- Chelliah, V. *et al.* BioModels: ten-year anniversary. *Nucleic Acids Res.* **43**, D542–D548 (2015).
- Büchel, F. *et al.* Path2Models: large-scale generation of computational models from biochemical pathway maps. *BMC Syst. Biol.* **7**, 116 (2013).
- Büchel, F. *et al.* Qualitative translation of relations from BioPAX to SBML qual. *Bioinformatics* **28**, 2648–2653 (2012).
- Keating, S.M. & Le Novère, N. Encoding neuronal models in SBML. In *Computational Systems Neurobiology* (ed. Le Novère, N.) 459–488 (Springer, New York, 2012).
- Hucka, M., Bergmann, F.T., Keating, S.M. & Smith, L.P. A profile of today's SBML-compatible software. IEEE Seventh International Conference on e-Science Workshops, Stockholm, Sweden, 5–8 December, 2011.
- Lloyd, C.M., Halstead, M.D. & Nielsen, P.F. CellML: its future, present and past. *Prog. Biophys. Mol. Biol.* **85**, 433–450 (2004).
- Olivier, B.G. & Snoep, J.L. Web-based kinetic modelling using JWS Online. *Bioinformatics* **20**, 2143–2144 (2004).
- Sivakumaran, S., Hariharaputran, S., Mishra, J. & Bhalla, U.S. The database of quantitative cellular signaling: management and analysis of chemical kinetic models of signaling networks. *Bioinformatics* **19**, 408–415 (2003).
- Lloyd, C.M., Lawson, J.R., Hunter, P.J. & Nielsen, P.F. The CellML model repository. *Bioinformatics* **24**, 2122–2123 (2008).
- Bornstein, B.J., Keating, S.M., Jouraku, A. & Hucka, M. LibSBML: an API library for SBML. *Bioinformatics* **24**, 880–881 (2008).
- Le Novère, N. *et al.* Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat. Biotechnol.* **23**, 1509–1515 (2005).
- Hoops, S. *et al.* COPASI—a Complex Pathway Simulator. *Bioinformatics* **22**, 3067–3074 (2006).
- Machne, R., Finney, A., Muller, S., Lu, J., Widder, S. & Flamm, C. The SBML ODE Solver Library: a native API for symbolic and fast numerical analysis of reaction networks. *Bioinformatics* **22**, 1406–1407 (2006).
- Keller, R. *et al.* The systems biology simulation core algorithm. *BMC Syst. Biol.* **7**, 55 (2013).
- Sauro, H.M. *et al.* Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration. *OMICS* **7**, 355–372 (2003).
- Bianconi, F., Baldelli, E., Ludovini, V., Crino, L., Flacco, A. & Valigi, P. Computational model of EGFR and IGF1R pathways in lung cancer: a systems biology approach for translational oncology. *Biotechnol. Adv.* **30**, 142–153 (2012).

20. Proctor, C.J., Boche, D., Gray, D.A. & Nicoll, J.A. Investigating interventions in Alzheimer's disease with computer simulation models. *PLoS One* **8**, e73631 (2013).
21. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* **25**, 25–29 (2000).
22. Hastings, J. *et al.* The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Res.* **41**, D456–D463 (2013).
23. Schriml, L.M. *et al.* Disease ontology: a backbone for disease semantic integration. *Nucleic Acids Res.* **40**, D940–D946 (2012).
24. Gremse, M. *et al.* The BRENDA Tissue Ontology (BTO): the first all-integrating ontology of all organisms for enzyme sources. *Nucleic Acids Res.* **39**, D507–D513 (2011).
25. Apweiler, R. *et al.* UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* **32**, D115–D119 (2004).
26. Bidkhori, G., Moeini, A. & Masoudi-Nejad, A. Modeling of tumor progression in NSCLC and intrinsic resistance to TKI in loss of PTEN expression. *PLoS One* **7**, e48004 (2012).
27. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
28. Schaefer, C.F. *et al.* PID: the Pathway Interaction Database. *Nucleic Acids Res.* **37**, D674–D679 (2009).
29. Karp, P.D., Riley, M., Saier, M., Paulsen, I.T., Paley, S.M. & Pellegrini-Toole, A. The EcoCyc and MetaCyc databases. *Nucleic Acids Res.* **28**, 56–59 (2000).
30. Wittig, U. *et al.* SABIO-RK–database for biochemical reaction kinetics. *Nucleic Acids Res.* **40**, D790–D796 (2012).
31. van Iersel, M.P. *et al.* Software support for SBGN maps: SBGN-ML and LibSBGN. *Bioinformatics* **28**, 2016–2021 (2012).
32. Proctor, C.J. & Gray, D.A. GSK3 and p53 - is there a link in Alzheimer's disease? *Mol. Neurodegener.* **5**, 7 (2010).
33. Proctor, C.J., Pienaar, I.S., Elson, J.L. & Kirkwood, T.B. Aggregation, impaired degradation and immunization targeting of amyloid-beta dimers in Alzheimer's disease: a stochastic modelling approach. *Mol Neurodegener.* **7**, 32 (2012).
34. Lemaire, V., Tobin, F.L., Greller, L.D., Cho, C.R. & Suva, L.J. *Modeling the interactions between osteoblast and osteoclast activities in bone remodeling. J. Theor. Biol.* **229**, 293–309 (2004).
35. Ermentrout, B. XPPAUT. In *Computational Systems Neurobiology* (ed. Le Novère, N.) 519–531 (Springer, New York, 2012).
36. Moraru, I.I. *et al.* Virtual cell modelling and simulation software environment. *IET Syst. Biol.* **2**, 352–362 (2008).
37. Dräger, A. *et al.* SBML2L(A)T(E)X: conversion of SBML files into human-readable reports. *Bioinformatics* **25**, 1455–1456 (2009).
38. Li, C., Courtot, M., Le Novère, N. & Laibe, C. BioModels.net Web Services, a free and integrated toolkit for computational modelling software. *Brief. Bioinform.* **11**, 270–277 (2010).
39. Jupp, S. *et al.* The EBI RDF platform: linked open data for the life sciences. *Bioinformatics* **30**, 1338–1339 (2014).
40. Proctor, C.J. *et al.* Modelling the actions of chaperones and their role in ageing. *Mech Ageing Dev.* **126**, 119–131 (2005).
41. Proctor, C.J., Tsigotis, M. & Gray, D.A. An in silico model of the ubiquitin-proteasome system that incorporates normal homeostasis and age-related decline. *BMC Syst. Biol.* **1**, 17 (2007).
42. Hooper, C., Killick, R. & Lovestone, S. The GSK3 hypothesis of Alzheimer's disease. *J. Neurochem.* **104**, 1433–1439 (2008).
43. Hooper, C., Meimaridou, E., Tavassoli, M., Melino, G., Lovestone, S. & Killick, R. p53 is upregulated in Alzheimer's disease and induces tau phosphorylation in HEK293a cells. *Neurosci. Lett.* **418**, 34–37 (2007).
44. Proctor, C.J. & Gray, D.A. Explaining oscillations and variability in the p53-Mdm2 system. *BMC Syst. Biol.* **2**, 75 (2008).
45. Passos, J.F. *et al.* Feedback between p21 and reactive oxygen production is necessary for cell senescence. *Mol. Syst. Biol.* **6**, 347 (2010).
46. Proctor, C.J. & Lorimer, I.A. Modelling the role of the Hsp70/Hsp90 system in the maintenance of protein homeostasis. *PLoS One* **6**, e22038 (2011).
47. Proctor, C.J., Tangeman, P.J. & Ardley, H.C. Modelling the role of UCH-L1 on protein aggregation in age-related neurodegeneration. *PLoS One* **5**, e13175 (2010).
48. Tang, M.Y., Proctor, C.J., Wouffe, J. & Gray, D.A. Experimental and computational analysis of polyglutamine-mediated cytotoxicity. *PLoS Comput. Biol.* **6**, pii: e1000944 (2010).
49. Demin, O. *et al.* Systems pharmacology models can be used to understand complex pharmacokinetic-pharmacodynamic behavior: an example using 5-lipoxygenase inhibitors. *CPT Pharmacometrics Syst. Pharmacol.* **2**, e74 (2013).
50. Faratian, D. *et al.* Systems biology reveals new strategies for personalizing cancer medicine and confirms the role of PTEN in resistance to trastuzumab. *Cancer Res.* **69**, 6713–6720 (2009).
51. Palmer, R., Nyman, E., Penney, M., Marley, A., Cedersund, G. & Agoram, B. Effects of IL-1beta-blocking therapies in type 2 diabetes mellitus: a quantitative systems pharmacology modeling approach to explore underlying mechanisms. *CPT Pharmacometrics Syst. Pharmacol.* **3**, e118 (2014).
52. Dwivedi, G. *et al.* A multiscale model of interleukin-6-mediated immune regulation in Crohn's disease and its application in drug discovery and development. *CPT Pharmacometrics Syst. Pharmacol.* **3**, e89 (2014).
53. Bento, A.P. *et al.* The ChEMBL bioactivity database: an update. *Nucleic Acids Res.* **42**, D1083–D1090 (2014).
54. Krause, F., Uhlendorf, J., Lubitz, T., Schulz, M., Klipp, E. & Liebermeister, W. Annotation and merging of SBML models with semanticSBML. *Bioinformatics* **26**, 421–422 (2010).
55. Schulz, M., Krause, F., Le Novère, N., Klipp, E. & Liebermeister, W. Retrieval, alignment, and clustering of computational models based on semantic annotations. *Molecular Syst. Biol.* **7**, 512 (2011).
56. Neal, M.L., Galdzicki, M., Gallimore, J.T. & Sauro, H.M. A C library for retrieving specific reactions from the BioModels database. *Bioinformatics* **30**, 129–130 (2014).
57. Ajmera, I., Swat, M., Laibe, C., Le Novère, N. & Chelliah, V. The impact of mathematical modeling on the understanding of diabetes and related complications. *CPT Pharmacometrics Syst. Pharmacol.* **2**, e54 (2013).

© 2015 The Authors CPT: Pharmacometrics & Systems Pharmacology published by Wiley Periodicals, Inc. on behalf of American Society for Clinical Pharmacology and Therapeutics. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Supplementary information accompanies this paper on the *CPT: Pharmacometrics & Systems Pharmacology* website (<http://www.wileyonlinelibrary.com/psp4>)