HAL
open science

An integrative modeling framework reveals plasticity of TGF-β signaling.

Geoffroy Andrieux, Michel Le Borgne, Nathalie Théret

To cite this version:

Geoffroy Andrieux, Michel Le Borgne, Nathalie Théret. An integrative modeling framework reveals plasticity of TGF-β signaling.. BMC Systems Biology, 2014, 8 (1), pp.30. 10.1186/1752-0509-8-30 . inserm-00978313

HAL Id: inserm-00978313

https://inserm.hal.science/inserm-00978313

Submitted on 14 Apr 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Andrieux et al. BMC Systems Biology 2014, 8:30
http://www.biomedcentral.com/1752-0509/8/30

BMC Systems Biology

RESEARCH ARTICLE Open Access

An integrative modeling framework reveals plasticity of TGF-β signaling

Geoffroy Andrieux[1], Michel Le Borgne[2] and Nathalie Théret[1*]

Abstract

Background: The TGF-β transforming growth factor is the most pleiotropic cytokine controlling a broad range of cellular responses that include proliferation, differentiation and apoptosis. The context-dependent multifunctional nature of TGF-β is associated with complex signaling pathways. Differential models describe the dynamics of the TGF-β canonical pathway, but modeling the non-canonical networks constitutes a major challenge. Here, we propose a qualitative approach to explore all TGF-β-dependent signaling pathways.

Results: Using a new formalism, CADBIOM, which is based on guarded transitions and includes temporal parameters, we have built the first discrete model of TGF-β signaling networks by automatically integrating the 137 human signaling maps from the Pathway Interaction Database into a single unified dynamic model. Temporal property-checking analyses of 15934 trajectories that regulate 145 TGF-β target genes reveal the association of specific pathways with distinct biological processes. We identify 31 different combinations of TGF-β with other extracellular stimuli involved in non-canonical TGF-β pathways that regulate specific gene networks. Extensive analysis of gene expression data further demonstrates that genes sharing CADBIOM trajectories tend to be co-regulated.

Conclusions: As applied here to TGF-β signaling, CADBIOM allows, for the first time, a full integration of highly complex signaling pathways into dynamic models that permit to explore cell responses to complex microenvironment stimuli.

Keywords: TGF-β, Discrete dynamic model, Signaling pathways, Guarded transition

Background

Complex signaling by the polypeptide transforming growth factor TGF-β is one of the most intriguing networks that governs multifunctional processes and plays a pivotal role in tissue homeostasis and morphogenesis [1]. Whereas TGF-β is ubiquitously expressed in all cell types and tissues, its effects differ according to cellular type and microenvironment. For example, TGF-β inhibits epithelial cell growth but can promote stromal cell proliferation. The confounding pleiotropic effects of TGF-β derive from the complex nature of its signaling, whose full characterization requires modeling approaches that are still lacking. TGF-β signals through a heteromeric complex of two types of transmembrane serine/threonine kinases, the type I (TβRI) and type II (TβRII) receptors. TGF-β binding to TβRII induces the recruitment and phosphorylation of TβRI, which transduces signals to downstream intracellular substrates, the first of which are the R-Smad proteins. Once phosphorylated, R-Smad proteins hetero-dimerize with a common partner, CoSmad, and heterodimeric complexes move into the nucleus where they regulate the transcription of TGF-β target genes. Alternatively, non-Smad pathways involved in TGF-β signaling include several MAP kinase pathways, the Rho-like GTPase signaling pathways, and phosphatidylinositol-3-kinase/AKT pathways [2,3]. Hence, combinations of Smad and non-Smad pathways contribute to the high heterogeneity of cell responses to TGF-β. Moreover, the molecular actors of these pathways are part of other cell signaling networks activated by additional extracellular stimuli, leading to complex and extensive crosstalk that has been difficult to model.

Modeling the dynamics of the TGF-β canonical signaling pathway has followed classical differential approaches that either couple signaling together with receptor trafficking [4] or focus on Smad phosphorylation [5], Smad

* Correspondence: nathalie.theret@univ-rennes1.fr
[1]INSERM U1085, IRSET, Université de Rennes 1, 2 avenue Pr Léon Bernard, 35043 Rennes, France
Full list of author information is available at the end of the article

BioMed Central

© 2014 Andrieux et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

nucleocytoplasmic shuttling [6,7] and Smad oligodimerization [8]. In addition, integrative models have coupled receptor trafficking to Smad pathways [9-11]. We have taken advantage of these existing quantitative approaches to develop our own models, which we have applied to the study of the role of the ADAM12 tumor biomarker [12] and of the TIF1γ tumor suppressor [13]. These advances suggested that such small differential models could be developed into useful tools to investigate the role of new regulatory components of the canonical TGF-β signaling pathway. A major obstacle, however, is that differential equation-based models remain limited to a small number of reactions [14]. Because the explosion in the number of variables in complex networks makes parameterization intractable at the cellular scale [15], integration of the TGF-β signaling networks clearly requires other methods.

Alternative qualitative modeling approaches are based on discrete-event systems where each entity (genes or proteins) is represented by a finite-state variable and rules encode the possible states and interactions of biomolecules [16]. The two-state Boolean logic is the simplest discrete formalism and has been successfully applied for modeling specific signaling pathways such as those for the Epithelial Growth Factor Receptor [17] and insulin [18]. While Boolean approaches are based on switching functions and consider networks as logical circuits, other discrete formalisms such as rule-based methods allow for the biochemical and biophysical description of multi-state components [19]. For instance, agent-based modeling approaches have been employed to describe the behavior of TGF-β and EGF crosstalk in non-small-cell lung cancer models [20] and epithelial restitution [21]. Similarly, the role of TGF-β in epidermal wound healing has been deciphered using computational agent-based models [22]. More recently, hybrid models for tumor-stromal environment centered on TGF-β and EGF canonical pathways have described interactions between the extracellular matrix and growth-factor effects [23] and a related hybrid discrete-element cellular automata model has been proposed to understand how TGF-β modulates tumor-stroma interactions [24]. While informative, these studies remain partial and fail to account for the full complexity of TGF-β dependent signaling networks. A major challenge, then, is to integrate all available information within a dynamic model that fully addresses how TGF-β modulates heterogeneous cell responses.

The last decades have seen the accumulation of a rich trove of information about the molecular actors of signaling (extracellular stimuli, membrane receptors, transduction signal proteins, transcriptional factors) and their associated biochemical reactions (receptor activation, protein phosphorylation, cytoplasmic-nuclear shuttling, transcription). Considerable effort has gone into mining the literature to build signaling databases and provide a view of cell signaling pathways through descriptive graph-based representations (KEGG [25], Ingenuity Pathway [26], Biocarta, Reactome [27], Pathway Interaction Database (PID) [28]). While all of these approaches improve graph-based analyses of signaling pathways, the translation schemes used to translate biological data into discrete variables and the parameterization of reaction times relative to each other to obtain dynamic models raise many difficulties. Within a given database, distributed knowledge blends various biological concepts such as biochemical reactions and functional processes and includes heterogeneous levels of details that do not permit automatic translation into discrete models. More recent databases such as Reactome [27] and the Pathway Interaction Database (PID) [28] use a homogeneous concept of biological reactions based on mechanistic information to facilitate further formal interpretations and development of large-scale systems, and a Boolean framework was recently proposed for modeling cellular signaling from the whole Reactome database. In this case, however, parameterization of time remains to be achieved [29] as discrete models do not easily lend themselves to the handling of logical time based on partial ordering of events. In synchronous time models, all components change simultaneously while only one component of the state is allowed to change at each step of model evolution in asynchronous time models. While mixed-rule dating has been proposed to obtain a more realistic view of biological events [30,31], these approaches still fall short of successfully modeling signaling reactions.

To address these daunting difficulties, we have developed a new non-ambiguous formal interpretation of signaling pathways as discrete dynamic models. This interpretation differs from previous approaches in that it models the signal rather than the molecular biochemical network that transmits the signal. The resulting language, Computer-Aided Design for BIOlogical Models (CADBIOM), is based on a simplified version of guarded transitions [32] in which we introduced temporal parameters for each transition. This leads to high degree of flexibility in signal distribution and allows, for the first time, the construction of a fully integrative discrete dynamic model of TGF-β signaling pathways. To do so, we integrated the whole PID database into a single unified model containing 9177 biomolecules that include all TGF-β related signaling networks. Based on model-checking methods implemented in the CADBIOM application, we further analyzed all trajectories involved in the regulation of 145 TGF-β target genes included in the model. Importantly, clustering analyses of signaling trajectories successfully discriminate between Smad versus non-Smad-dependent genes and identify for the first time new combinations of extracellular stimuli involved in the regulation of TGF-β target genes. In addition,

trajectory analyses are predictive for gene co-regulation as assessed by experimental gene expression data.

Results and discussion

Building a TGF-β CADBIOM model

To translate and integrate any kind of biological reaction involved in TGF-β signaling pathways, we developed a new language, CADBIOM, based on guarded-transition formalism (see Methods and Additional file 1: Figures S1 and S2). Briefly, a biological reaction is formalized as a transition from input biomolecules to output biomolecules under some conditions or guards. Conditions involve biomolecules as activators or inhibitors and the biological reaction takes place only if the inputs are present and the condition is true. We then consider that the occurrence of a biological reaction induces information propagation from each input to each output. Next, all the biological reactions are integrated by connecting the biomolecules: the target of a transition becomes the origin of the next one. The resulting network is turned into a dynamic system according to the choice of transitions that are fired at each step. To overcome the limitations of synchronous or asynchronous models, we introduce events, discrete signals that guide or restrain the choice of fireable transitions. The potential firing of reactions is dependent on the presence or absence of events that is formalized within the guard. By default, events are initially attributed to each transition, except for the formation or dissociation of complexes that share the same event since both components are simultaneously present.

Taking advantage of the mechanistic view of the PID database, we extracted TGF-β signaling-pathway knowledge by developing a program that automatically translates XML files from PID into CADBIOM language. Three TGF-β signaling related pathways have been reported in PID: these include "TGF-β receptor signaling" (PID Ref: tgfbrpathway), "regulation of cytoplasmic and nuclear Smad2/3 signaling" (PID Ref: Smad2-3pathway) and "regulation of nuclear Smad2/3 signaling" (PID Ref: Smad2- 3nuclearpathway). According to the integrative rules described in Methods, we built the union of the three files by automatically removing redundancies (Figure 1). This model integrates 435 biomolecules and 141 reactions from PID, but only from the canonical Smad-dependent pathway. To extract all signaling information related to the non-Smad TGF-β signaling pathways, we had to incorporate information related to MAPK, Rho-like GTPase and phosphatidylinositol-3-kinase/AKT. A difficulty immediately arises in that these components are part of many other PID pathways; accordingly, all biomolecules that influence TGF-β signaling pathways need to be integrated. To overcome this obstacle and identify all components related to TGF-β, we first built a single CADBIOM model for the complete PID database. As illustrated in Figure 2, all 9248 reactions from 137 PID pathways were integrated into 9264 CADBIOM transitions. Note that transitions and reactions are not equivalent since a reaction for the formation of a complex between two biomolecules must be translated into two transitions. In contrast, two different reactions described in PID and sharing the same inputs and outputs can be translated into one or several CADBIOM transitions. The efficiency of integration is demonstrated by the reduction of the 27876 biomolecules that are part of the 137 PID pathways to 9177 non-redundant places in the CADBIOM model. Indeed, one PID biomolecule can be implicated in several reactions and/or pathways, while a CADBIOM biomolecule is represented by a unique place and all of its reactions are either incoming or outgoing transitions. It is important to note that the translation and integration of the 137 PID pathways into a single CADBIOM model does not alter the distribution of the ontology terms associated with biomolecules, ruling out loss of information resulting from integration (Additional file 1: Figure S3).

Based on this complete signaling model, we next sought to delineate parts of the network that influence TGF-β signaling pathways. We did this by creating an activation graph from all PID information using biomolecules as nodes and the dependencies between biomolecules as edges (Figure 3A). The dependency relationships include transitions and the influence of conditions on the output of transitions. The resulting influence graph illustrates all the relations between the biomolecules of the database, not only in terms of information propagation, but also in terms of regulation of reactions. Quite spectacularly, all the information from PID is highly connected with a major component containing 8986 nodes (98% of the total PID content) included TGF-β, indicating that integration of TGF-β signaling amounts to compiling the entire database. Most importantly, this integration is achieved without discarding non-canonical pathways that are not described as TGF-β-related in PID. Based on the distribution of the connectivity and modularity analysis, we showed that the network exhibits scale-free behavior (Figure 3B) and clustering of connected components identifies a TGF-β related associated module (Figure 3C, Additional file 2: Table S1). Taken together, our data provide the first dynamic discrete model of the TGF-β signaling network, including both Smad and non-Smad-dependent pathways, which integrates all networks that influence the cellular response to TGF-β. Because topological analyses of the activation graph could not be explain the heterogeneity of TGF-β signaling pathways, we further explored this highly complex model using model-checking approaches as described below.

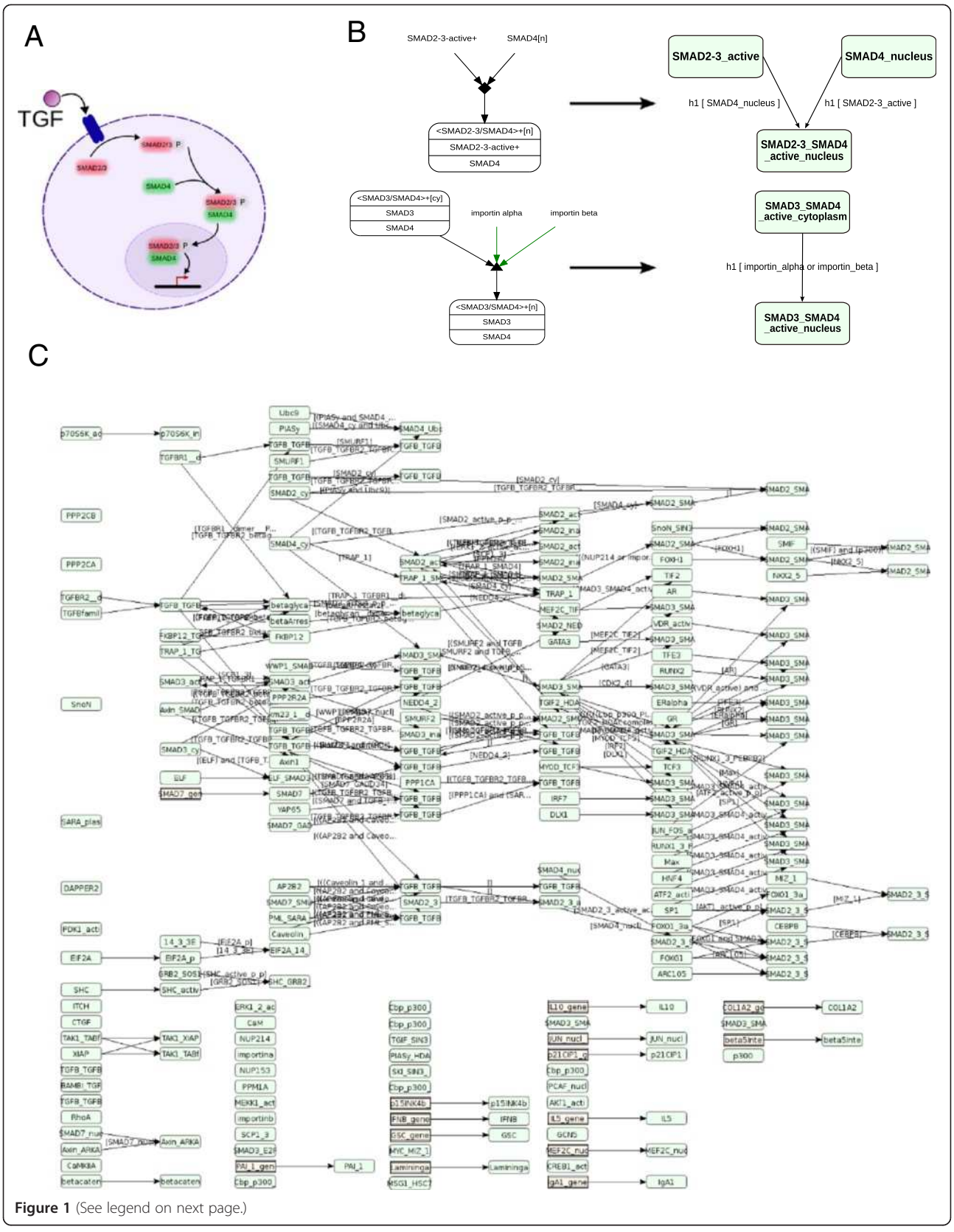

Figure 1 (See legend on next page.)

(See figure on previous page.)
Figure 1 CADBIOM model for the canonical TGF-β pathway. (A) Schematic representation of the canonical pathway. Binding of TGF-β to the receptor induces recruitment and phosphorylation of Smad2/3. Activated Smad2/3 heterodimerizes with Smad4 and the complex migrates to nucleus where it regulates gene transcription. **(B)** Two examples of translations from the PID database (left) into the CADBIOM representation (right). Top: interaction of Smad2/3 with Smad4. The CADBIOM representation uses two transitions from a component toward the complex, conditioned by the presence of the other component and leading to the synchronization of the transition. The same event h1 is associated with the two transitions that occur simultaneously. Bottom: migration of the Smad complex to the nucleus. The CADBIOM representation uses a transition between the two places "SMAD3_SMAD4_active_cytoplasm" and "SMAD3_SMAD4_active_nucleus" conditioned by the presence of importins within the guard. The event h1 indicates the timing of the transition. **(C)** CADBIOM illustration of the canonical pathway after integration of the three PID pathways: "TGF-β receptor signaling" (PID Ref: tgfbrpathway), "regulation of cytoplasmic and nuclear Smad2/3 signaling" (PID Ref: Smad2-3pathway) and "regulation of nuclear Smad2/3 signaling" (PID Ref: Smad2-3nuclearpathway).

Application to the regulation of TGF-β dependent genes

Understanding how TGF-β induces heterogeneous biological responses is a critical question, which we addressed next by exploring the complexity of pathways that govern the regulation of TGF-β-dependent genes. Looking for signaling pathways that regulate a gene G coding for a protein P amounts to searching for all scenarios that verify the reachability of a property P (see Methods and Additional file 1: Figure S4). A scenario is a list of biomolecules activated at the initialization of the model. However, due to the size of the model and the wide set of potential solutions, we focused on minimal scenarios that take place in 10 steps, the representative size of solutions in our CADBIOM model. For each minimal scenario, we analyzed all possible trajectories to reach the property P, a trajectory being expressed as the list of biomolecules activated during signal propagation. Among the 679 genes described in PID, 145 have minimal scenarios containing the term TGF-β and correspond to extracellular stimulation by TGF-β.

To decipher the signaling regulatory network of TGF-β, we searched for all the signaling trajectories played according to these scenarios. We identified 15934 such trajectories. Based on their content in biomolecules, clustering analysis of trajectories showed that the presence or absence of the Smad intracellular signaling proteins defined specific groups (Figure 4 and Additional file 3: Table S2). As shown in Table 1, 18 TGF-β-dependent genes were reachable only if Smad proteins were present in their trajectories. An additional set of 5 genes was also found both with Smad- and non-Smad-dependent trajectories and clustered within a unique group, suggesting similar regulation. Four of these genes, DAPK1, COX_2/PTGS2, JUN and NOS2, are indeed functionally associated within the "cancer pathways" from the KEGG pathway database. CADBIOM also identified a much larger number of TGF-β-dependent genes that were reachable in the absence of Smad proteins, resulting in a set of 122 genes with non-Smad-dependent trajectories (See Additional file 4: Table S3). Importantly, all trajectories containing Smad proteins are associated with genes regulated by TGF-β as their unique extracellular stimulus, while trajectories induced by TGF-β associated with other extracellular stimuli never contained Smad proteins (Additional file 5: Table S4).

To assess the functional association of target genes within each group, we performed gene ontology annotation using the Database for Annotation, Visualization and Integrated Discovery, DAVID (Figure 4C). Remarkably, genes regulated by the Smad-dependent TGF-β pathway were found to be specifically associated with the biological processes that include the GO terms "regulation of cell proliferation", "negative regulation of gene expression" and "regulation of programmed cell death" (blue color). In contrast, genes regulated by non-Smad-dependent pathways (grey or red) were always associated with multiple biological processes according to the absence or presence of other extracellular stimuli. The non-Smad pathways (colored in red) containing only TGF-β were found to regulate genes with GO terms "immune response and response to external stimulus", "positive regulation of cell proliferation" and "developmental process, behavior and regulation of programmed cell death". In cases where non-Smad pathways linked TGF-β to other stimuli in their trajectories, we observed gene expression responses that were part of Smad-dependent TGF-β trajectories (*e.g.* "programmed cell death" GO) or of non-Smad-dependent trajectories (*e.g.* "immune response" GO), as well as a broader range of responses belonging to GOs such as "metabolism", "development" and "homeostasis".

Altogether, we identified 31 combinations where TGF-β was linked to other extracellular stimuli, illustrating the high degree of plasticity of TGF-β gene regulation (Additional file 5: Table S4 and Figure 5). Among these combinations, 18 associate TGF-β with IL12 and are involved in the regulation of 9 genes: CCR5, GADD45A and GADD45B, MIP1A and MIP1B, Granzyme A and Granzyme B and IL17F and IL1RA. Interestingly all these genes are functionally linked to viral infection/inflammation and stress response, suggesting that different combinations of stimuli can lead to a similar biological function. Indeed CCR5 is a beta-chemokine receptor that binds HIV, and Mip1A and B are major HIV-suppressive factors that bind CCR5. Additionally, Granzyme A and B are serine proteases that mediate apoptosis of virus-infected

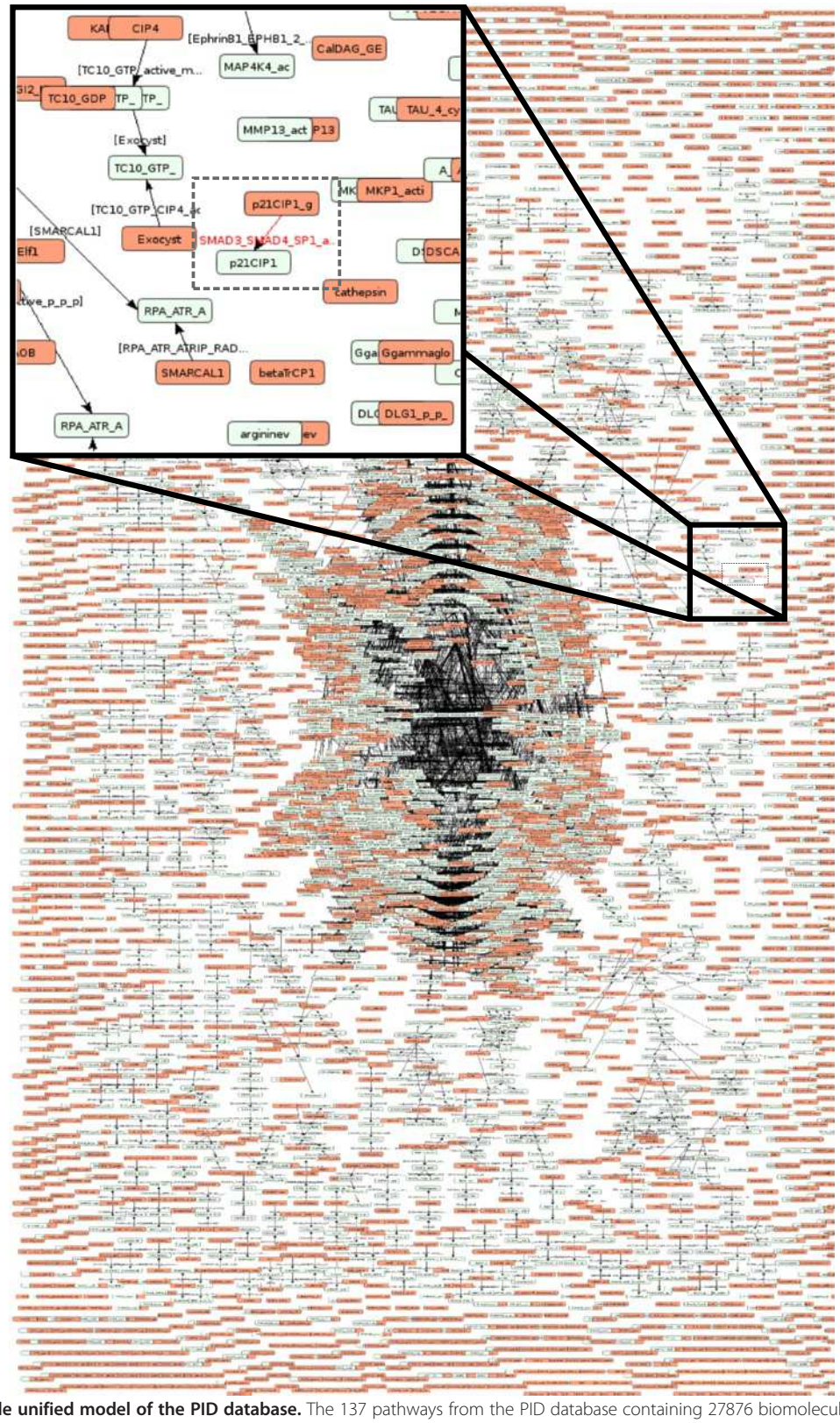

Figure 2 A single unified model of the PID database. The 137 pathways from the PID database containing 27876 biomolecules are integrated into a CADBIOM dynamic model containing 9177 places. Places without input are colored in orange. The insert illustrates an example of regulation by Smad3/Smad 4 and shows the transition of the p21CIP1 gene to p21CIP1 protein, which represents activation of the p21 gene. The condition for the transition contains an extensive logical formula that is not detailed here.

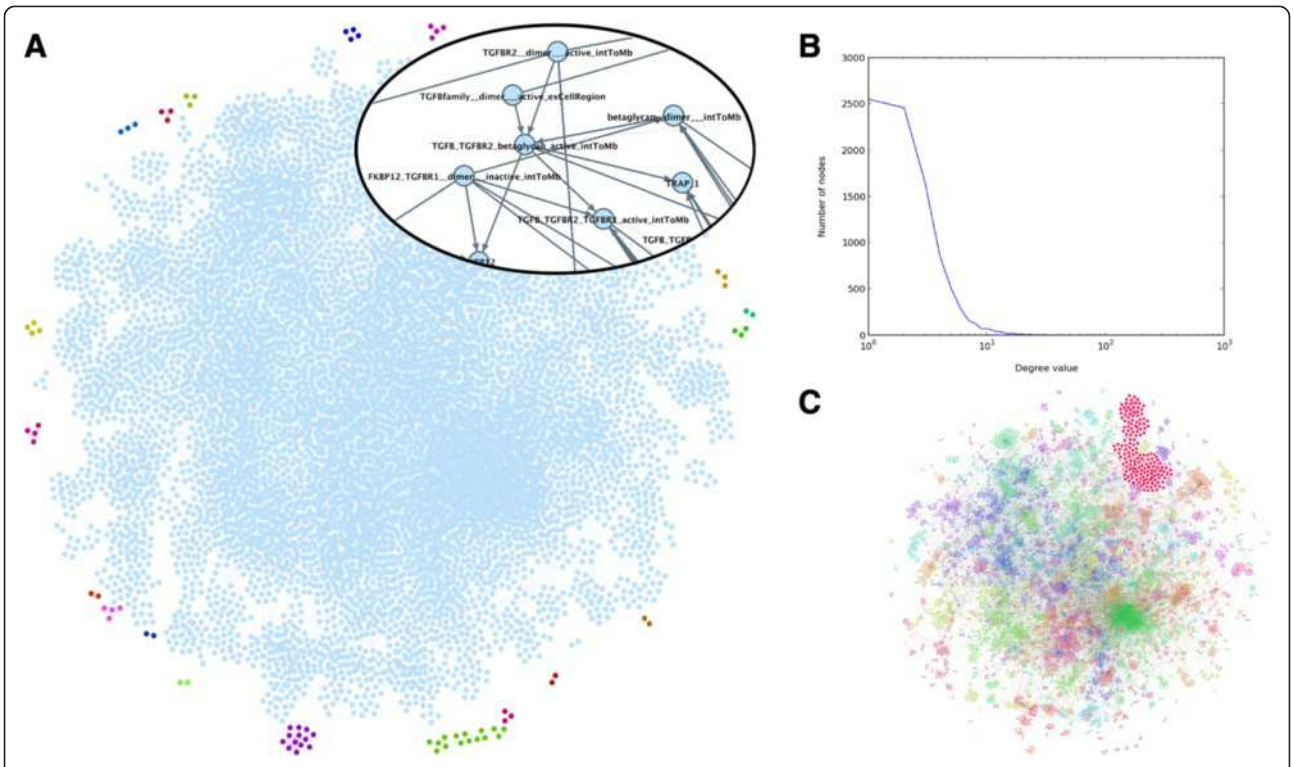

Figure 3 Activation graph of the PID database. (A) The activation graph is built using places from the CADBIOM model as nodes and the dependencies between places as edges. The dependency relationships include the transitions and the influence of conditions on the output of transition. The resulting activation graph contains 9077 nodes and 15499 edges. TGF-β belongs to the major connected component containing 8986 nodes (98% of nodes from the activation graph) demonstrating that all information from PID influences TGF-β signaling. Colors distinct from light blue denote non-connectivity of components. In the zoom view (insert), thin directed edges denote condition dependencies and thick directed edges transition dependencies. **(B)** Representation of degree distribution. **(C)** Modularity based hierarchical agglomerative clustering. Colors show the different clusters that are further detailed in Additional file 2: Table S1. A major specific module containing 168 places that include TGF-β, Smad proteins, TGF-β receptors and related molecules such as BMP (members of the TGF-β family) is shown as enlarged red nodes.

cells and IL1RA and IL17F act as proinflammatory cytokines. Finally, GADD45A/B are transcriptional factors that mediate global response to environmental stress. These functional links revealed by CADBIOM analysis have not been previously reported using other modeling approaches. Taken together, these data are in accordance with and strengthen the known concept of Smad- and non-Smad-dependent TGF-β pathways and provide for the first time trajectories for regulatory ligands. Of note also is the identification, through CADBIOM analyses of trajectories for gene regulation, of the 31 combinations that associate TGF-β with other extracellular stimuli to drive gene regulation through Smad-independent pathways. In addition, CADBIOM identifies the TGF-β/IL12 association as a novel basic module for signaling networks involved in the regulation of 9 genes implicated in the response to viral infection. Because such complex associations cannot be evaluated by classical experimental approaches using *in vitro* cell stimulation, we investigated the biological relevance of these combinations by analyzing co-expression of genes sharing extracellular stimuli.

CADBIOM trajectories are associated with co-expression of genes

The cellular microenvironment induces numerous simultaneous stimuli and cells respond according to the integration of all of these signals. For a given set of stimuli, signaling pathways activate gene expression profiles that can be identified using transcriptomic methods. Based on this supposition, we expected that two genes regulated by TGF-β together with other extracellular stimuli might be found to be co-regulated in biological samples. To test this hypothesis, we analyzed gene expression of each pair of genes sharing extracellular stimuli by using experimental data from Gemma, a database for the meta-analysis of gene expression profiles [33]. Based on the available set of 2110 Gemma human profiling expression studies, we determined whether two genes sharing a combination of TGF-β and another stimulus are co-expressed or not. In this case, 19 out of identified 31 combinations were found to regulate at least two genes. As shown in Figure 6, gene pairs activated by similar combinations were found to be

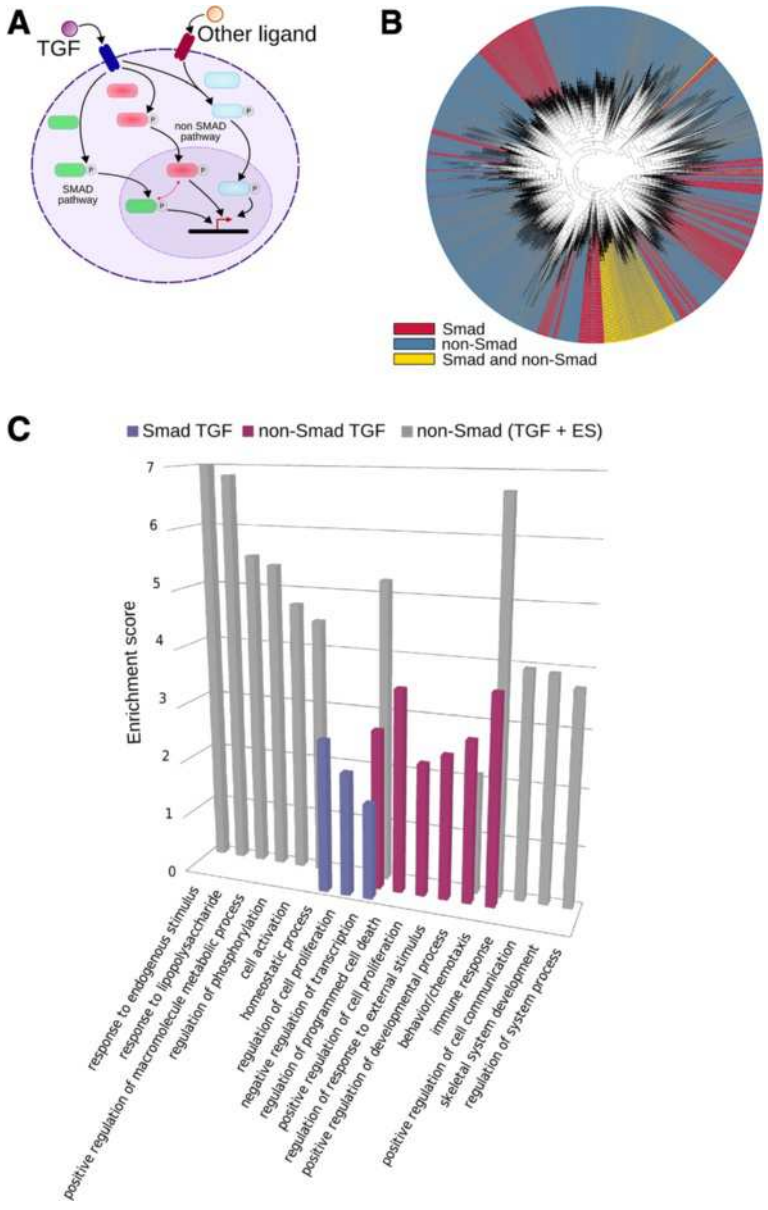

Figure 4 Analysis of the 15934 signaling trajectories containing TGF-β. (A) Schematic representation of the Smad and non-Smad TGF-β cell signaling pathways. **(B)** Trajectories are clustered according to their content in biomolecules using Euclidian distance metrics. Trajectories that contain Smad protein are colored in red, those that do not in blue and those containing both Smad and non-Smad proteins are in yellow. **(C)** Functional annotations of TGF-β dependent gene targets using the Database for Annotation, Visualization and Integrated Discovery (DAVID) tool. GOs are colored in grey, blue and red as indicated in the text. ES: extracellular stimuli.

significantly co-expressed, validating the biological relevance of trajectory analyses. As illustrated in Figure 7 for CCR5 and MIP1B, CADBIOM analysis can lead to the identification of genes that are co-expressed and share combinations, thus uncovering new and highly complex common trajectory networks.

To generalize this key observation to any genes from the model, we randomly generated 649 pairs of genes present in both the CADBIOM model and the Gemma database (Additional file 6: Table S5). Extracting all gene pairs that share signaling trajectories from the CADBIOM model and the corresponding co-expression data from Gemma reveals that 81% of co-expressed gene pairs in Gemma share regulatory signaling trajectories in the CADBIOM model. In contrast, 82% of non co-expressed gene pairs did not share regulatory pathways. These results demonstrate that trajectories identified by CADBIOM are highly predictive for gene co-expression in biological samples.

Table 1 Genes with SMAD-dependent trajectories identified by CADBIOM

PID ID	HUGO ID	Description
COX_2*	PTGS2	Prostaglandin-endoperoxide synthase 2
DAPK1 *	DAPK1	Death-associated protein kinase 1
GSC	GSC	Goosecoid homeobox
ID1	ID1	Inhibitor of DNA binding 1
IFNB	IFNB1	Interferon, beta 1
IgA1	IGHA1	Immunoglobulin heavy constant alpha 1
IL10	IL10	Interleukin 10
IL5*	IL5	Interleukin 5 (colony-stimulating factor, eosinophil)
JUN*	JUN	Jun oncogene
KLK2	KLK2	Kallikrein-related peptidase 2
Laminin gamma1	LAMC1	Laminin, gamma 1
MEF2C	MEF2C	Myocyte enhancer factor 2C
NOS2*	NOS2	Nitric oxide synthase 2, inducible
p15INK4b	CDKN2B	Cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)
p21CIP1	CDKN1A	Cyclin-dependent kinase inhibitor 1A (p21, Cip1)
RAB7	RAB7A	RAB7A, member RAS oncogene family
SMAD7	SMAD7	SMAD family member 7
SNAI1	SNAI1	Snail homolog 1 (Drosophila)
SOCS3	SOCS3	Suppressor of cytokine signaling 3
TCF3	TCF3	Transcription factor 3
TLX2	TLX2	T-cell leukemia homeobox 2
TRPC6	TRPC6	Transient receptor potential cation channel, subfamily C, member 6
TRPV1	TRPV1	Transient receptor potential cation channel, subfamily V, member 1

The Table lists genes with SMAD-dependent trajectories. Gene names (ID) are given according to their nomenclature in the PID and HUGO databases, respectively, and are followed by their description. Asterisks denote genes that are also found with non-SMAD-dependent trajectories (see Text and Additional file 4: Table S3).

In summary, CADBIOM constitutes a new formalism for modeling signaling networks and provides the first discrete dynamic model for TGF-β signaling pathways. Based on model-checking methods, we describe the highly complex signaling trajectories that regulate TGF-β-dependent gene targets and we demonstrate that combinations of TGF-β with other extracellular stimuli lead to non-Smad-dependent pathways. Finally, we show that CADBIOM modeling allows for the identification of complex trajectories induced by multiple extracellular stimuli, successfully predicting gene co-expression patterns and biological functions that cannot be characterized by direct experimental approaches. More generally, we expect that CADBIOM will be of broad use to help unravel other highly complex pathways that have proven intractable using classical modeling approaches. To provide for detailed explorations of signaling models, CADBIOM design and analysis, all required tools have been implemented into a user-friendly open-source application package that is readily available to researchers (cadbiom.genouest.org).

Conclusions

Signaling pathway networks orchestrate cell life through a uniquely connected complex molecular circuitry for propagation of information. Because of the huge number of biological reactions that are implicated, modeling cell signaling to predict cell responses to this huge flow of information is a major challenge that requires new approaches if all information is to be dynamically integrated. In the present work, we have developed a new formalism, CADBIOM, based on guarded transitions and combined discrete abstraction to propose, for the first time, a fully integrated model for TGF-β signaling pathways.

Major issues in modeling biological large-scale phenomena are the collection of information from the literature. While cell signaling pathways are described in numerous databases, a recent report demonstrated a high degree of inconsistencies when different databases are compared [34]. Based on the Jaccard similarity coefficient, the authors compared four well understood pathways that involve the cytokines EGF (Epidermal growth factor), TGF-β (Transforming growth factor), TNFα (Tumor necrosis factor) and the signaling protein WNT (wingless-type) described in six databases, including GeneGo (www.genego.com), KEGG [25], NCI-PID [28], NetPath [35], PANTHER [36] and Reactome [27]. Only 10% similarity was found, suggesting that the description of each pathway is database- or even curator-specific. To circumvent such potential pitfalls, we extracted instead all relevant information from TGF-β pathways using a single database, PID, integrating them into a single unified model. The choice of PID is fully justified by the observation that, unlike other databases, PID formalizes signal propagation instead of describing biochemical reactions. This means that an interaction is described as a biological event that includes its participating molecules and conditions, an interaction which consumes its inputs and produces its outputs. This overall description is close to that used for guarded-transition systems [32], a state/event formalism that can represent both flow circulation with transitions and natural composition rules, and remote influences with transition guards. The direct translation of the complete XML-formatted database content into CADBIOM formalism then allows for the integration of information and automatically creates the first dynamic model of the TGF-β signaling network.

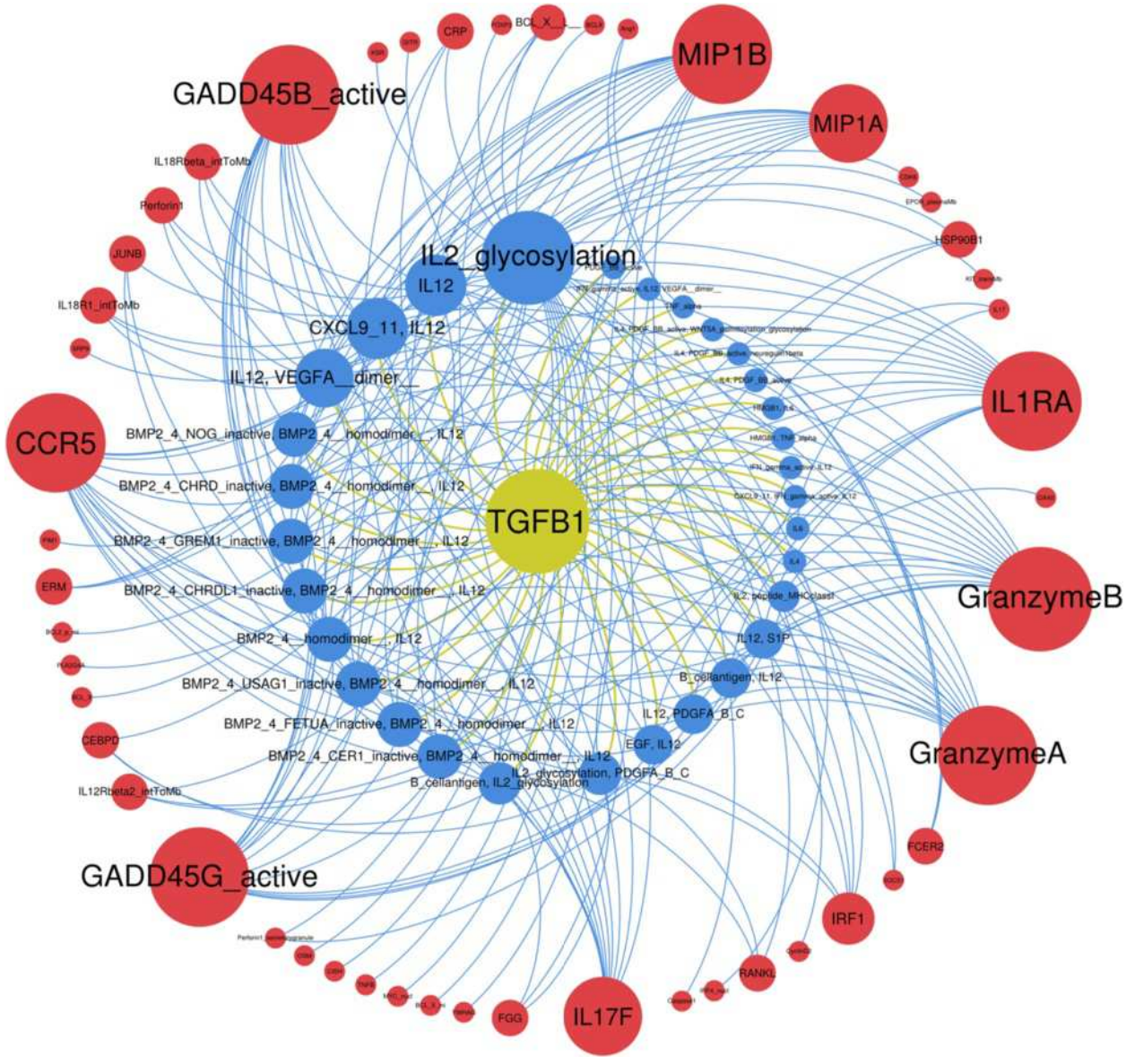

Figure 5 Plasticity of gene responses to combinations of TGF-β with other extracellular stimuli. 49 genes (red nodes) are regulated through signaling trajectories containing TGF-β (yellow node) and at least one other extracellular stimulus (blue nodes). The size of the nodes is relative to the number of edges. CCR5, GADD45A/B, MIP1A/B, GranzymeA/B, IL1RA and IL17F are the most represented genes and share the largest number of combinations in their trajectories.

Three TGF-β signaling pathways are described in PID: TGF-β receptor signaling, regulation of cytoplasmic and nuclear Smad2/3 signaling and regulation of nuclear Smad2/3 signaling (see Figure 1) and summarize the canonical TGF-β pathway. We compiled these into CADBIOM and, importantly, also included non-Smad TGF-β pathways and all their influencing components (see Figure 2). The resulting single unified CADBIOM model is a highly connected graph containing 9177 places where 98% of nodes directly or indirectly influence TGF-β-dependent signals. Note that this high connected component contained also other receptor/factors. An important point to consider in automatic translation from a database is the continuous upgrading of the generated model. PID upgrading was recently stopped and closing of the PID project, a collaborative effort between the National Cancer Institute (NCI, Besthesda) and the Nature Publishing Group, has been announced. Fortunately, the Reactome project from the EBI consortium shares information with PID, which already imported biological data from Reactome's BioPAX2. While some limitations in the specifications of Reactome's BioPAX2 were previously reported by PID's authors [28], the

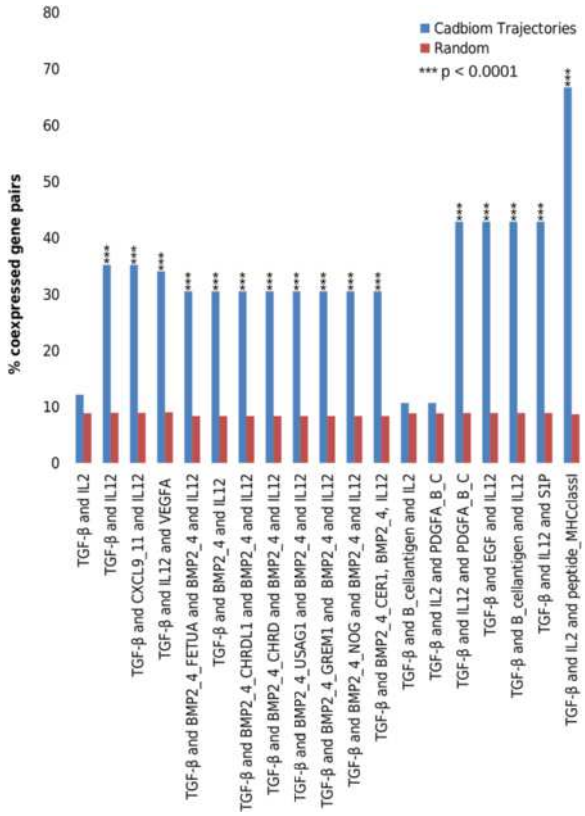

Figure 6 Association between extracellular stimuli involving TGF-β and gene co-regulation. For each gene pairs sharing combinations of extracellular stimuli (blue), we extracted the experimental data of gene co-expression from the Gemma database. Results are expressed as the ratio of co-expressed genes to possible gene pairs. Controls (red) are random gene pairs (n = 1000) from Gemma.

recent upgrade of BioPAX 2 to BioPAX3 [37] should facilitate data extraction from Reactome into CADBIOM.

Besides the straightforward structural analyses of large integrated graphs, the critical point in building signaling models is to describe the dynamics of signal propagation. We have accommodated this key requirement by using a state/event formalism based on guarded transitions that differ from traditional ones through the use of event algebra. Formalisms used in UML (Unified Modeling Language) or in state-charts [38] include an event in the transition guard but lack an operation for combining events into new events. In contrast, using CADBIOM formalism, the timing of model evolution is directly linked to the biological reactions by associating an event to each reaction. Consequently and although the transition systems use a Boolean expression for the guard, CADBIOM formalism differ from classical Boolean approaches [29,39-42] because the temporal order of the reaction is not forced (synchronous or asynchronous) but included in the guard. CADBIOM modeling permits simultaneous changes to several reactions and the exploration of all potential behaviors. The resulting modeling framework is much richer since simultaneous reactions are neither excluded nor imposed.

Biological models mainly focus on the search for steady states [43]. However, while such approaches make sense for gene regulatory or metabolic networks, they are not appropriate for modeling signal propagation. CADBIOM overcomes this limitation by implementing tools based on model-checking methods to investigate the reachability of properties instead [44,45]. These analyses usually require the computation of highly complex state-transition diagrams, which contain approximately 2^{9000} (10^{2709}) states in our model. To avoid graph calculus, we use propositional logic and SAT solver-based approaches that have been demonstrated to be efficient for characterizing biochemical networks [46]. We have used these approaches to analyze scenarios and trajectories involved in the activation of the 145 TGF-β-dependent gene targets identified by our unified model. We demonstrate that scenarios can indeed discriminate between Smad and non-Smad-dependent pathways and that TGF-β associates with other extracellular stimuli to regulate non-Smad-dependent genes that are functionally related to shared biological processes. In these cases, the complexity of functions regulated by non-Smad pathways is mainly due to the association of TGF-β with other extracellular stimuli. We note that such a context-dependent role for external stimuli has been previously suggested by a global analysis of the effects of pair-wise ligand combinations [47]. However, these experimental approaches strongly limit the number of combinations that can be studied and a broad analysis of signaling trajectories is impossible. Using discrete models and model-checking approaches, we reasoned in the opposite way and, instead of performing TGF-β stimulation to identify signaling toward gene targets, we investigated trajectories starting from genes, thereby allowing identification of complex extracellular stimuli that include TGF-β.

Classical modeling approaches perform simulations using extracellular stimuli or activation of receptors at the initialization step to mimic experimental conditions in cell culture. In contrast, CADBIOM uses reachability properties to explore all signaling pathways without *a priori* on external stimuli that may be combined with TGF-β. This constitutes a radically different approach that is closer to the biological reality of cells, which live in a complex environment where TGF-β is always a part of stimuli. As a case in point, our study identifies a novel TGF-β-dependent network containing 9 viral response-related genes that are regulated by combinations that chiefly involve IL12, a cytokine with diverse functional effects [48]. The anti-inflammatory effect of TGF-β has been proposed to modulate the functions of IL12 in the

Figure 7 Common-trajectory network for two co-regulated genes, MIP1B and CCR5. Solutions for reachability of MIP1B and CCR5 genes were calculated using the CADBIOM application and trajectories for each solution were compiled in a graph representation. Blue arrows symbolize transitions and red arrows conditions. Yellow nodes are specific to CCR5 trajectories, red nodes are specific to MIP1B trajectories and green nodes are common to both CCR5 and MIP1B trajectories.

immune response [49,50], but pathological contexts involving changes in environmental stimuli lead to a pro-inflammatory role for TGF-β [51,52]. These observations illustrate the need to investigate complex TGF-β models that include all potential influences and take into account all physio-pathological contexts. CADBIOM provides for such integration by creating a new state-event formalism to integrate biological reactions into a dynamic model for signal propagation. Finally, the creation of the first single unified model of the TGF-β signaling network, which integrates both Smad- and non-Smad-dependent pathways, constitutes an important landmark and provides a unique and powerful tool for the full exploration of TGF-β-dependent functions. More generally, the new computational approach provided by CADBIOM for modeling signaling networks should improve our overall understanding of cellular responses to complex stimuli, as illustrated here by the extraordinarily complex example of TGF-β signaling.

Methods

CADBIOM

CADBIOM software is covered by a GNU-public license that permits the conception, simulation and questioning of the discrete dynamic models describe in this article. All these features can be reached using the graphical user interface or through its application programming interface (API). Automatic conception of models is made possible through the PID database translation scheme (XML format). Resulting models can be stored in Cadlang, a text representation of CADBIOM models. CADBIOM is freely available at http://cadbiom.genouest.org/.

Model formalism

CADBIOM formalism is based on guarded transitions and the introduction of discrete signals called events. A guarded transition is graphically represented by: $A \xrightarrow{h[Cond]} B$. A and B are respectively the origin and target places of the transition. The guard of the transition is composed by an event h and a condition Cond, a logical formula with places as variables and ∨, ∧ and ¬ as logical operators.

Places are state variables that take Boolean values and represent biomolecules such as proteins or complexes. An event is the mathematical concept that denotes an occasional occurrence. An event has a name h and a singleton domain to denote the occurrence of the event. At each step, an event can occur (denoted by the symbol T) or not (denoted by the symbol ⊥). A realization of a finite set of events $(h_i)_{i\in I}$ is a sequence of elements of $\{T, \perp\}^I \setminus \{\perp^I\}$. To handle integrated models containing large amount of data, we combine events and states. Operations on events are well known in computer science. The two basic operators are a merge that corresponds to multiplexing and a selection of occurrences corresponds to under-sampling. CADBIOM borrows the default operator and the when operator from the Signal language [53]. The default operator merges the two events h_1 and h_2. The event $h = (h_1$ default $h_2)$ is present when either the event h_1 or the event h_2 is present, h is absent otherwise. The when operator is an operator between events and logical combinations of state variables and selects occurrences of an event when the propositional formula evaluates to True on its right-hand side. The event $h = (h_1$ when B) is present when h_1 is present and B is True, it is absent otherwise.

Model simulation

The system evolves according to transition firings that change the value of places, which is either *True* or *False* at any step. Given a guarded transition $A \xrightarrow{h[C]} B$, we define the transition event as $h_{tr} = h$ when (A ∧C). The transition event h_{tr} is present if and only if h is present and (A ∧ C) is *True*, that is, when the input place is *True* and the condition C is verified. When a transition is fired, the source is inactivated and the target is activated. The evolution function of a state A relies on the current value A_{k+1} at step k to the next value A_{k+1} at step k + 1.

The state of a place changes if either an in-transition or an out-transition is fired. In the case of simultaneous firing, we postulate that activation prevails over inactivation.

We define two events associated with a place A by:

$$h_{in} = \text{default } t_r \in T_{in}\, h_{tr}$$

$$h_{out} = \text{default } t_r \in T_{out}\, h_{tr}$$

The h_{in} event is present if at least one of the in-transitions is fired. The h_{out} event is present if at least one of the out-transitions is fired. The mathematical formalization of the rules is given by:

$$A' = (h_{in} \text{ default } \neg h_{out}) \text{ default } A$$

where A' is the value of A at next step.

The initial condition for simulation analyses requires the design of activated places (called scenario) and a specific timing. CADBIOM proposes either a manual selection or an automatic selection of "frontier" places. Frontier is the set of places which cannot be activated from inside the model. In guarded transition models, places without input transitions belong to the frontier. The script for events timing is provided during reachability analyses as a txt file that can be loaded in simulation tool

Property search

The temporal properties of models are explored using methods based on model checking. The aim of these methods is to determine an allocation of state variables and event timing that verify a property, that is, any logical formula composed of model places and logical operators ∨, ∧ and ¬. To investigate the reachability of biological

properties such as "how to express a given gene?", we search for the allocation of states variables and event timing that lead to the activation of the place symbolizing the gene. We focus on frontier places because signaling networks are usually polarized from the extracellular environment to the nucleus. A frontier is the set of places without any in-transitions that cannot be activated from inside the model. All other places are initialized in an inactivated state.

We first generated the formulas that describe the evolution of the system. Using the evolution rules above, dynamic models are translated into propositional logic formulas using the Tseitin translation. The resulting formulas are under conjunction normal form (CNF) and qualify the dynamics between steps i and $i+1$. We then unfold the trajectory from step 0 to n. The whole formula is then given to a SAT solver to find an allocation of state variable and event timing that satisfies the formula.

We next define a scenario as the (F; T) pair, where F is a set of places which are activated at the initialization step and T is a sequence of sets of events h. This reachability property is then used to search for minimal scenarios such that the property is not achieved as soon as one component is removed in the initialization places or as soon as a component of event timing is disabled. A scenario is said to be minimal if, for any place $A \in$ F, (F \ {A}, T) is not an activation condition and for any $i < n$ and any $h \in H_i$, (F; ($H_{1,\ldots,}H_i$ \{h}, …,H_{n-1})) is not an activation condition. Any scenarios can be simulated to retrieve all the activated places that lead to reachability of the property. These activated places compose a trajectory.

Compilers

CADBIOM software includes several compilers sharing the same back-end. The front-ends compile PID XML files, CADBIOM XML files and Cadlang files into an intermediate representation of guarded-transition models. The common back-end generates logical constraints in propositional clause form. During this step, constant propagation and common sub-expression elimination optimizations are performed. Combined with many peephole optimizations, these techniques allow the back-end to generate reduced sets of constraints that facilitate the work of the SAT solver.

Resources

The CADBIOM model is a translation of the PID database content (http://pid.nci.nih.gov/). See Additional file 1: Figure S1 for the complete translation scheme. Model-checking analysis is performed using the cryptominisat SAT solver (http://www.msoos.org/cryptominisat2/). Graph representations of solutions are displayed using Gephi, an open-source platform for graph visualization (https://gephi.org/).

Additional files

Additional Material. The following additional data are available with the online version of this paper. Additional data file 1 is a figure describing the scheme for translating biological reactions from PID database into CADBIOM formalism. Additional data file 2 is a figure illustrating translation schemes for the TGF-β model. Additional data file 3 is a figure showing the conservation of ontology during the translation process. Additional data file 4 is a figure detailing the procedure for calculating reachability. Additional data file 5 is a text detailing mathematicals semantics for CADBIOM formalism.

Additional file 1: Figure S1. Translation scheme for biological reactions into guarded transitions. **Figure S2.** Illustration of translation schemes for the TGF-b model. **Figure S3.** Conservation of biological annotations. **Figure S4.** Reachability of a property P. **Supplementary Methods.** mathematical semantics.

Additional file 2: Table S1. Modularity classes in activation graph. The activation graph has a significant modularity score of 0,819 that reveals a non-random distribution of edges and a high connectivity between nodes inside classes. Most of TGF-β signaling-related components, including TGFβ, TGFβR and SMAD proteins belong to modularity class 10 that contains also TGF-β superfamily members like BMP.

Additional file 3: Table S2. Vector matrix of trajectories that regulate BCL_X_L, JUNB, CRP, FGG, IRF1, CEBPD, RANKL genes. The occurrence of components is expressed as 0 (absent) and 1 (present) in trajectories.

Additional file 4: Table S3. Genes with non_Smad-independent trajectories identified by CADBIOM. The Table lists genes with non-Smad-dependent trajectories. Gene names (ID) are given according to their nomenclature in the HUGO and PID databases, respectively, and are followed by their description.

Additional file 5: Table S4. Target genes regulated by combinations of extracellular stimuli identified by CADBIOM. Proteins that serve as extracellular stimuli and their combinations thereof are listed together with their target genes, listed according to the PID nomenclature.

Additional file 6: Table S5. List of the 649 pairs of genes randomly chosen to evaluate the association between co-expression and trajectories.

Competing interests

The authors declare that they have no conflict of interest.

Authors' contributions

GA conceived the study, developed the software and drafted the manuscript. MLB conceived the study, developed the software and drafted the manuscript. NT conceived the study and drafted the manuscript. All authors have read and approved the final manuscript.

Acknowledgements

This work was supported by the Institut National de la Santé et de la Recherche Médicale (INSERM), the Ligue Nationale Contre le Cancer and the National Research Agency (ANR). The authors would like to thank Dr. Emmanuel Käs (CNRS UMR 5099, Toulouse, France) for help in writing this manuscript and the GenOuest Bioinformatics Platform for hosting the CADBIOM website.

Author details

[1]INSERM U1085, IRSET, Université de Rennes 1, 2 avenue Pr Léon Bernard, 35043 Rennes, France. [2]Université de Rennes 1, IRISA, 263 avenue du général Leclerc, 35042 Rennes, France.

Received: 26 November 2013 Accepted: 3 March 2014
Published: 12 March 2014

References

1. Massague J: **TGFbeta signalling in context.** *Nat Rev Mol Cell Biol* 2012, **13:**616–630.
2. Moustakas A, Heldin CH: **Non-Smad TGF-beta signals.** *J Cell Sci* 2005, **118:**3573–3584.
3. Mu Y, Gudey SK, Landstrom M: **Non-Smad signaling pathways.** *Cell Tissue Res* 2012, **347:**11–20.
4. Vilar JM, Jansen R, Sander C: **Signal processing in the TGF-beta superfamily ligand-receptor network.** *PLoS Comput Biol* 2006, **2:**e3.
5. Clarke DC, Betterton MD, Liu X: **Systems theory of Smad signalling.** *Syst Biol Stevenage* 2006, **153:**412–424.
6. Melke P, Jonsson H, Pardali E, ten Dijke P, Peterson C: **A rate equation approach to elucidate the kinetics and robustness of the TGF-beta pathway.** *Biophys J* 2006, **91:**4368–4380.
7. Schmierer B, Tournier AL, Bates PA, Hill CS: **Mathematical modeling identifies Smad nucleocytoplasmic shuttling as a dynamic signal-interpreting system.** *Proc Natl Acad Sci U S A* 2008, **105:**6608–6613.
8. Nakabayashi J, Sasaki A: **A mathematical model of the stoichiometric control of Smad complex formation in TGF-beta signal transduction pathway.** *J Theor Biol* 2009, **259:**389–403.
9. Chung SW, Miles FL, Sikes RA, Cooper CR, Farach-Carson MC, Ogunnaike BA: **Quantitative modeling and analysis of the transforming growth factor beta signaling pathway.** *Biophys J* 2009, **96:**1733–1750.
10. Zi Z, Klipp E: **Constraint-based modeling and kinetic analysis of the Smad dependent TGF-beta signaling pathway.** *PLoS One* 2007, **2:**e936.
11. Zi Z, Feng Z, Chapnick DA, Dahl M, Deng D, Klipp E, Moustakas A, Liu X: **Quantitative analysis of transient and sustained transforming growth factor-b signaling dynamics.** *Mol Syst Biol* 2011, **7:**492.
12. Gruel J, Leborgne M, LeMeur N, Theret N: **In silico investigation of ADAM12 effect on TGF-beta receptors trafficking.** *BMC Res Notes* 2009, **2:**193.
13. Andrieux G, Fattet L, Le Borgne M, Rimokh R, Théret N: **Dynamic regulation of TGF-β signaling by Tif1g: a computational approach.** *PLoS One* 2012, **7:**e33761.
14. Chelliah V, Laibe C, Le Novere N: **BioModels Database: a repository of mathematical models of biological processes.** *Methods Mol Biol* 2013, **1021:**189–199.
15. Aldridge BB, Burke JM, Lauffenburger DA, Sorger PK: **Physicochemical modeling of cell signaling pathways.** *Nat Cell Biol* 2006, **8:**1195–1203.
16. Saadatpour A, Albert R: **Discrete dynamic modeling of signal transduction networks.** *Methods Mol Biol* 2012, **880:**255–272.
17. Samaga R, Saez-Rodriguez J, Alexopoulos LG, Sorger PK, Klamt S: **The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data.** *PLoS Comput Biol* 2009, **5:**e1000438.
18. Wu M, Yang X, Chan C: **A dynamic analysis of IRS-PKR signaling in liver cells: a discrete modeling approach.** *PLoS One* 2009, **4:**e8040.
19. Sekar JA, Faeder JR: **Rule-based modeling of signal transduction: a primer.** *Methods Mol Biol* 2012, **880:**139–218.
20. Wang Z, Birch CM, Sagotsky J, Deisboeck TS: **Cross-scale, cross-pathway evaluation using an agentbased non-small cell lung cancer model.** *Bioinformatics* 2009, **25:**2389–2396.
21. Stern JR, Christley S, Zaborina O, Alverdy JC, An G: **Integration of TGF-b- and EGFR-based signaling pathways using an agent-based model of epithelial restitution.** *Wound Repair Regen* 2012, **20:**862–871.
22. Sun T, Adra S, Smallwood R, Holcombe M, MacNeil S: **Exploring Hypotheses of the Actions of TGFb1 in Epidermal Wound Healing Using a 3D Computational Multiscale Model of the Human Epidermis.** *PLoS One* 2009, **4:**e8515.
23. Kim Y, Othmer HG: **A Hybrid Model of Tumor-Stromal Interactions in Breast Cancer.** *Bull Math Biol* 2013, **75**(8)**:**1304–1350.
24. Basanta D, Strand DW, Lukner RB, Franco OE, Cliffel DE, Ayala GE, Hayward SW, Anderson AR: **The role of transforming growth factor-beta-mediated tumor-stroma interactions in prostate cancer progression: an integrative approach.** *Cancer Res* 2009, **69:**7111–7120.
25. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res* 2000, **28:**27–30.
26. Guerra C: **Ingenuity Pathways Analysis: software for discovering and modelling pathways and networks in your systems data.** *Comp Biochem Physiol A Mol Integr Physiol* 2008, **150:**S50–S50.
27. Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, Caudy M, Garapati P, Gopinath G, Jassal B, Jupe S, Kalatskaya I, Mahajan S, May B, Ndegwa N, Schmidt E, Shamovsky V, Yung C, Birney E, Hermjakob H, D'Eustachio P, Stein L: **Reactome: a database of reactions, pathways and biological processes.** *Nucleic Acids Res* 2011, **39:**D691–D697.
28. Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, Buetow KH: **PID: the Pathway Interaction Database.** *Nucleic Acids Res* 2009, **37:**D674–D679.
29. Fearnley LG, Nielsen LK: **PATHLOGIC-S: a scalable Boolean framework for modelling cellular signalling.** *PLoS One* 2012, **7:**e41977.
30. Faure A, Naldi A, Chaouiya C, Thieffry D: **Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle.** *Bioinformatics* 2006, **22:**e124–e131.
31. Garg A, Di Cara A, Xenarios I, Mendoza L, De Micheli G: **Synchronous versus asynchronous modeling of gene regulatory networks.** *Bioinformatics* 2008, **24:**1917–1925.
32. Rauzy A: **Guarded Transition Systems: a new States/Events Formalism for Reliability Studies.** *J Risk Reliability* 2008, **222:**495–505.
33. Zoubarev A, Hamer KM, Keshav KD, McCarthy EL, Santos JR, Van Rossum T, McDonald C, Hall A, Wan X, Lim R, Gillis J, Pavlidis P: **Gemma: a resource for the reuse, sharing and meta-analysis of expression profiling data.** *Bioinformatics* 2012, **28:**2272–2273.
34. Kirouac DC, Saez-Rodriguez J, Swantek J, Burke JM, Lauffenburger DA, Sorger PK: **Creating and analyzing pathway and protein interaction compendia for modelling signal transduction networks.** *BMC Syst Biol* 2012, **6:**29.
35. Kandasamy K, Mohan SS, Raju R, Keerthikumar S, Kumar GS, Venugopal AK, Telikicherla D, Navarro JD, Mathivanan S, Pecquet C, Gollapudi SK, Tattikota SG, Mohan S, Padhukasahasram H, Subbannayya Y, Goel R, Jacob HK, Zhong J, Sekhar R, Nanjappa V, Balakrishnan L, Subbaiah R, Ramachandra YL, Rahiman BA, Prasad TS, Lin JX, Houtman JC, Desiderio S, Renauld JC, Constantinescu SN, *et al*: **NetPath: a public resource of curated signal transduction pathways.** *Genome Biol* 2010, **11:**R3–R14.
36. Mi H, Lazareva-Ulitsky B, Loo R, Kejariwal A, Vandergriff J, Rabkin S, Guo N, Muruganujan A, Doremieux O, Campbell MJ, Kitano H, Thomas PD: **The PANTHER database of protein families, subfamilies, functions and pathways.** *Nucleic Acids Res* 2005, **33:**D284–D288.
37. Demir E, Cary MP, Paley S, Fukuda K, Lemer C, Vastrik I, Wu G, D'Eustachio P, Schaefer C, Luciano J, Schacherer F, Martinez-Flores I, Hu Z, Jimenez-Jacinto V, Joshi-Tope G, Kandasamy K, Lopez-Fuentes AC, Mi H, Pichler E, Rodchenkov I, Splendiani A, Tkachev S, Zucker J, Gopinath G, Rajasimha H, Ramakrishnan R, Shah I, Syed M, Anwar N, Babur O, *et al*: **The BioPAX community standard for pathway data sharing.** *Nat Biotechnol* 2010, **28:**935–942.
38. Harel D: **Statecharts: A visual formalism for complex systems.** *Sci Comput Program* 1987, **8:**231–274.
39. De Jong H: **Modeling and simulation of genetic regulatory systems: a literature review.** *J Comput Biol* 2002, **9:**67–103.
40. Handorf T, Klipp E: **Modeling mechanistic biological networks: an advanced Boolean approach.** *Bioinformatics* 2012, **28:**557–563.
41. Saez-Rodriguez J, Alexopoulos LG, Epperlein J, Samaga R, Lauffenburger DA, Klamt S, Sorger PK: **Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction.** *Mol Syst Biol* 2009, **5:**331.
42. Helikar T, Rogers JA: **ChemChains: a platform for simulation and analysis of biochemical networks aimed to laboratory scientists.** *BMC Syst Biol* 2009, **3:**58.
43. Naldi A, Berenguier D, Faure A, Lopez F, Thieffry D, Chaouiya C: **Logical modelling of regulatory networks with GINsim 2.3.** *BioSystems* 2009, **97:**134–139.
44. Clarke E, Faeder J, Langmead C, Harris L, Jha S, Legay A: **Statistical Model Checking in BioLab: Applications to the Automated Analysis of T-Cell Receptor Signaling Pathway**. In *Computational Methods in Systems Biology*, volume 5307 of Lecture Notes in Computer Science. Edited by Heiner M, Uhrmacher A. Springer Berlin Heidelberg, ISBN; 2008:231–250. 978-3-540-88561-0.
45. Monteiro PT, Ropers D, Mateescu R, Freitas AT, de Jong H: **Temporal logic patterns for querying dynamic models of cellular interaction networks.** *Bioinformatics* 2008, **24:**i227–i233.
46. Carrillo M, Gongora PA, Rosenblueth DA: **An overview of existing modeling tools making use of model checking in the analysis of biochemical networks.** *Front Plant Sci* 2012, **3:**155.

47. Natarajan M, Lin KM, Hsueh RC, Sternweis PC, Ranganathan R: **A global analysis of cross-talk in a mammalian cellular signalling network.** *Nat Cell Biol* 2006, **8:**571–580.
48. Vignali DA, Kuchroo VK: **IL-12 family cytokines: immunological playmakers.** *Nat Immunol* 2012, **13:**722–728.
49. Musumeci M, Malaguarnera L, Simpore J, Messina A, Musumeci S: **Modulation of immune response in Plasmodium falciparum malaria: role of IL-12, IL-18 and TGF-beta.** *Cytokine* 2003, **21:**172–178.
50. Otano I, Suarez L, Dotor J, Gonzalez-Aparicio M, Crettaz J, Olague C, Vales A, Riezu JI, Larrea E, Borras F, Benito A, Hernandez-Alcoceba R, Menne S, Prieto J, Gonzalez-Aseguinolaza G: **Modulation of regulatory T-cell activity in combination with interleukin-12 increases hepatic tolerogenicity in woodchucks with chronic hepatitis B.** *Hepatology* 2012, **56:**474–483.
51. Yoshimura A, Muto G: **TGF-β function in immune suppression.** *Curr Top Microbiol Immunol* 2011, **350:**127–147.
52. Han G, Li F, Singh TP, Wolf P, Wang XJ: **The pro-inflammatory role of TGFb1: a paradox?** *Int J Biol Sci* 2012, **8:**228–235.
53. Le Guernic P, Gautier T, Le Borgne M, Le Maire C: **Programming Real-Time Applications with Signal.** *Proc IEEE* 1991, **79:**1321–1336.

doi:10.1186/1752-0509-8-30
Cite this article as: Andrieux *et al.*: An integrative modeling framework reveals plasticity of TGF-β signaling. *BMC Systems Biology* 2014 **8**:30.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central