Explaining response to drugs using Pathway Logic *

Carolyn Talcott and Merrill Knapp[1]

SRI International, Menlo Park, CA 94025, USA
firstname.lastname@sri.com

Abstract. Pathway Logic (PL) is a general system for modeling signal transduction and other cellular processes with the objective of understanding how cells work. Each specific model system builds on a knowledge base of rules formalizing local process steps such as post translational modification. The Pathway Logic Assistant (PLA) is a collection of visualization and reasoning tools that allow users to derive specific executable models by specifying of an initial state. The resulting network of rule instances describes possible behaviors of the modelled system. Subnets and pathways can then be computed (they are not hard wired) by specifying states to reach and/or to avoid. The STM knowledge base is a curated collection of signal transduction rules supported by experimental evidence. In this paper we describe methods for using the PL STM knowledge base and the PLA tools to explain observed perturbations of signaling pathways when cells are treated with drugs targeting specific activities or protein states. We also explore ideas for conjecturing targets of unknown drugs. We illustrate the methods on phosphoproteomics data (RPPA) from SKMEL133 melanoma cancer cells treated with different drugs targeting components of cancer signaling pathways. Existing curated knowledge allowed to us explain many of the responses. Conflicts between the STM model predictions and the data suggest missing requirements for rules to apply.

1 Introduction

Understanding how cells work is a fundamental question in Biology. It is important for basic science, as well as for practical applications including understanding disease, drug discovery, and synthetic biology. There are many aspects, including the different processes within a cell (metabolism, signaling, transcription/translation, ...), how these processes interact, what are the normal states, and what happens in response to some perturbation.

Executable mechanistic models [7] play an important role in understanding cellular processes, as they support in silico experiments, hypothesis generation, and feedback between laboratory experiments and model development. In the

* The work was partially supported by funding from the DARPA Big Mechansim program. The authors would like to thank the PL team for their many contributions, and the anonymous reviewers for helpful criticisms.

case of drug discovery such models help to determine details of the mechanism of action (MOA) and dually, drugs with a known MOA are used to learn details about how cells work.

The work reported here was done as part of a DARPA Big Mechanism project. The challenge was to use our Pathway Logic Signal Transduction model (STM) to explain how drugs with a known mechanism of action caused the changes in protein expression and/or phosphorylation measured by Reverse Phase Protein Array (RPPA) using data from [10].

The contributions of this paper are

- methods to explain effects of drugs on exponentially growing cells as measured by high throughput phosphoproteomics assays.
- a method to build a model of exponentially growing cells from a knowledge base of rules describing cellular events.
- methods to derive the mechanism (network of events) underlying response to treatment by drugs with known specific targets
- methods to hypothesize targets of unknown drugs, i.e. perturbations of the network that could explain measured responses.

Using these methods we were able to explain many of the observed changes in expression and phosphorylation in SKMEL133 cells when treated with drugs with known targets, and to make some conjectures regarding possible targets of two of the unknown drugs.

The SKMEL133 model is available at `pl.csl.sri.com/online.html` as part of the Pathway Logic suite of models. The accompanying guided tour is available as a link from the Online launcher, or directly from `pl.csl.sri.com/` along with a techreport version of this paper.

Plan. We provide a brief introduction to Pathway Logic and describe the general method for explaining drug study data in section 2. In section 3 we describe the data set and how it was processed in order to map the data to a PL model. The model of exponentially growing SKMEL133 cells is presented in section 4. In section 5 we use the model to explain the data for drugs with known, experimentally validated, targets. In section 6 we analyze the data for two of the unknown drugs, with consistent results in one case and many mysteries in the other case. Some related work is discussed in section 7, and we conclude with a summary and discussion of future work in section 8.

2 Pathway Logic models and their use to analyze data

The objective of Pathway Logic (PL) is to understand how cells work. A recent overview of PL can be found in [16]. The PL collection of models, knowledge bases, software, documentation, papers, and tutorials are available from the PL website [13]. The PL model collection includes models of metabolism, protease signaling in bacteria, protein glycosylation, and fragments of the human immune system. The most highly developed model is STM (Signal Transduction Model). This will be our starting point for modeling response to drugs.

2.1 PL concepts and reasoning tools

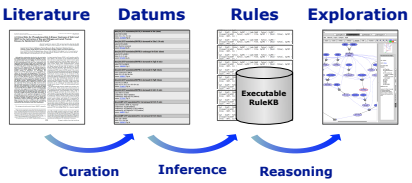


Fig. 1. From data to models in PL.

As shown in Figure 1, the STM Pathway Logic models are founded on two formal knowledge bases: a curated datum knowledge base (DKB), and a rules knowledge base (RKB), that share a controlled vocabulary formalized in Maude [4].

A datum formalizes an experimental observation of the state or location of protein or other biomolecule (RNA, Lipids, ...) either in some well-defined experimental condition, or a change in response to some signal or perturbation [12].

Signaling events are formalized as rewrite rules. They are generally inferred from datums, although rule sets can also be curated from review articles and text books, or simply hypothesized. A rule contains terms representing the change (before and after state) as well as terms representing the biological context required for the change to take place. A rule may be parametric, containing variables that can be instantiated in multiple ways to give different rule instances usable in different contexts. Rules in PL do not have rates.

The RKB can be thought of as a global model. Executable models of specific situations are generated by specifying initial conditions and constraints, formalized using a notion of *dish* (as in Petri dish). A dish is a term representing the initial state of the modeled system. It can be thought of as representing an experimental setup: cell type, growth conditions, and treatments or other perturbations. The cell type and growth conditions are represented by specifying which proteins and other biomolecules are present, their location, and their modification and/or activity state. The PL STM consists of rules concerning response to over 35 different stimuli (including Egf, IL1, Ngf, Tnf, Tgfb ...) as well as *common rules* that formalize local changes independent of a particular stimulus.

In PL, model elements and state are represented using a controlled vocabulary that is specified as a functional module in Maude. There is a core vocabulary shared by all PL knowledge bases/models and a model specific vocabulary that declares specific model elements (proteins, chemicals, modifications, locations, ...). The PL controlled vocabulary has several roles: organizing concepts via a sort/type hierarchy; determining legal/well-formed/meaningful terms by specifying constants and typed term constructors, and giving meaning to constants by providing metadata linking constant symbols to external references (Uniprot, HMDB, ...).

A PL executable model state is multi-set of *occurrences* of entities (proteins, chemicals, genes, ...). An occurrence specifies an entity, its modifications and/or activity state, and its location. For example `Braf-act@CLc` is an occurrence of active Braf in the cytoplasm (`CLc`), `PIP3@CLm` is an occurrence of the lipid `PIP3` in the cell membrane (`CLm`), `S6k1-phos!T412@CLc` is an occurrence in the cytoplasm of `S6k1` phosphorylated on threonine 412.

The STM model uses the term *family* for groups of proteins that cannot be differentiated by antibodies. For example, the anti-Akt antibody (CST#4691) used in [10] detects Akt1, Akt2, and Akt3. We cannot determine whether the increase in the level of protein expression is due to one and/or two and/or three of the Akts so we use the constant `Akts` to refer to some or all members of this family. Similarly, the antibody used to detect `Akt1-phos!S473` (CST#9271) also recognizes `Akt2-phos!S474` and `Akt3-phos!S472`. We use a site code (symbolic name) to represent the corresponding residues in all three proteins. The families and site codes used in the current work are shown in the table below.

Site Code	Refers to	and/or	and/or
Akts-phos!FSY	Akt1-phos!S473	Akt2-phos!S474	Akt3-phos!S472
Akts-phos!KTF	Akt1-phos!T308	Akt2-phos!T309	Akt3-phos!S307
Gsk3s-phos!SFAE	Gsk3a-phos!S21	Gsk3b-phos!S9	
Mek12s-phos!SMANS	Mek1-phos!S218-phos!S222	Mek2-phos!S222-phos!S226	
Erks-phos!TEY	Erk1-phos!T202-phos!Y204	Erk2-phos!T185-phos!Y187	

An important part of the PL system is the Pathway Logic Assistant (PLA), which is a tool to generate, visualize, browse, and analyse executable PL models. Given a dish and an RKB, PLA uses a symbolic reasoning and abstraction technique called *forward collection* to infer a minimal set of rule instances that cover all situations reachable from the initial state. The resulting concrete rule set naturally forms a network, linking rules by shared output/input elements. The initial state together with the collected rules forms an executable model. A theory transformation is used to convert the model to a Petri Net to be able to use reasoning tools for Petri Nets. PLA can now be used to specify goals and/or knockouts, derive the subnet of all pathways satisfying the goals (omitting the knockouts), invoke a model checker [15] to find specific pathways, and export nets as images or data structures for use by other tools.[1] Within a subnet one can ask for all the execution pathways leading to the goal, using an inference algorithm described in [6]. Knowing all the pathways one can compute properties such as single and double knockout occurrences or essential rules. If a single knockout occurrence is removed from the model, the goal will no longer be reachable. Similarly for double knockouts and essential rules.

2.2 Use of PL to explain data: generating a model

The first step in explaining experimental results is to define a model of the unperturbed cell system being studied. For the drug studies we want a snapshot of an exponentially growing cell system that is perturbed by addition of one or more drugs. Ideally, a model is built by defining an initial state (using expert knowledge, literature, the datum KB, and the COSMIC database (for mutations). Then, using PLA, we do a forward collection from this initial state, to collect all reachable rules in the STM RKB.

[1] One can knockout an occurrence, either from the initial state or a potentially reachable occurrence, or a rule. Each choice corresponds to a different experimental perturbation.

However, the world is not ideal, and the above steps may not work without some refinement. One problem is finding information about protein expression levels of a given cell line under different growth conditions, and the other is that, a priori, the rules in the RKB may capture different levels of detail (say Yphos vs phos!Y123) due to different experimental methods, and the rules may be more specific than necessary, or a rule may represents a set of more specific rules, for example by referring to a family of proteins rather than specific members.

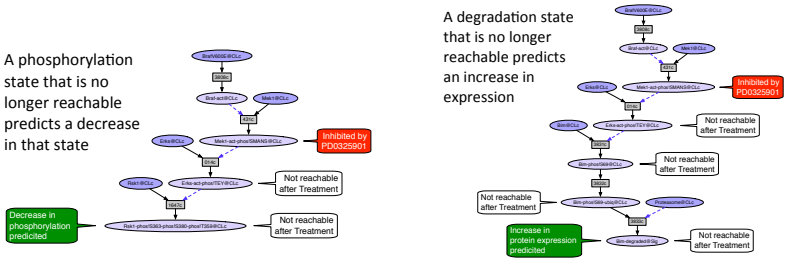
To address the first problem, we only attempt to include in the model the measured entities and any relevant up/down stream entities. We do this by a combination of "fuzzy" backward and forward collection (currently implemented by hand). The idea is (i) identify rules that would cause the changes seen in the data; (ii) identify rules that would meet the requirements of the first set of rules; and (iii) iterate until there are no more requirements to be met. Now we prepare an initial state: for each entity in the collected rules, determine the locations and modifications that cannot be produced by any rules. Modify the result using any available information about mutations and deletions for the cell line being studied. The unperturbed network is generated from the rule set and the resulting initial state using 'fuzzy' forward collection. The idea here is that some rules may need to be generalized in order to apply to generated states. For example a rule may require `Mek1-act@CLc` but the state may contain `Mek1-act-phos!SMANS@CLi`. Adding a variable to the modification set of the occurrences of `Mek1` in the rule solves the problem. After these adaptations, the PLA forward collection process can be used to generate a model of the unperturbed system.

2.3 Use of PL to explain data: using the model

In PL, explanations for measured changes in response to treatment of a cell system with a given drug can be found in several ways. One way is to knock out the drug target and use model checking to see if increases/decreases observed in the data agree with reachability results. We can also find all the paths (in the network model) to different observed significant changes and combine this information to suggest targets if the drug or its mechanism of action is unknown.

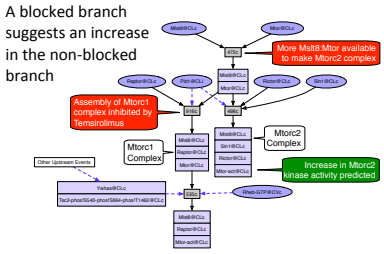
Here we focus on direct comparison of models of untreated and treated systems. Given a drug that is known to inhibit some occurrence in the model, we generate a model of the treated system by removing that occurrence from the network and use PLA to do a forwards collection to determine the remaining reachable subnet. Now we can compare the unperturbed (untreated) and perturbed (treated) model networks to obtain a qualitative prediction of increase/decrease in levels of some of the network occurrences. Three principles for inferring expected change are illustrated in figure 2.

Note that some of the drugs inhibit activity by direct allosteric inhibition. The conformational change caused by the drug should not be interpreted as the inhibition or enhancement of an upstream kinase. Some of the changes cannot be explained by a PL model because they are caused by things other than signal transduction. Some of the changes are due to proteins that are only expressed



(a) Blocked Modification

(b) Blocked degradation



(c) Blocked branch

Fig. 2. Three principles

during certain phases of the cell cycle. If a drug causes a cell cycle arrest, the proportion of cells in that phase are increased and the proteins only seen during that phase will be increased over those in cycling cells.

3 The experiment and data

To correctly interpret data, it is important to understand how it is generated and the criteria for interpreting measurements.

Primer on Interpreting the Results of Cell Based Assays

- An *experiment* starts with seeding cells into the containers (petri dish, flask, test tube) where they will be treated.
- The number of *biological replicate*s is the number of containers used for each treatment. This detects differences in results caused by the seeding and treatment procedures.

- The number of *technical replicate*s is the number of measurements made for each biological replicate. This gives you the probability that your detection method will give you the same value for the same sample.
- The number of *experimental replicate*s is the number of times the procedure is performed from different cell seedings. This gives you the probability that the change observed will occur in another experiment.
- The convention for publication in a cell biology data paper is to perform at least three independent experiments using three biological replicates for each treatment and control.
- The number of technical replicates required depends on the detection method used. The noisier the detection method, the more technical replicates required.

For the data set to be analyzed here, exponentially growing SKMEL133 cells were treated with 12 drugs at two concentrations. Change in protein expression/phosphorylation was measured for 138 entities at 24 hours using Reverse Phase PhosphoProteomics Analysis (RPPA) [3].

The data to be explained was available in two formats: (i) fold-change measurements using 3 biological replicates from one experiment based on an unreported number of technical replicates; (ii) relative concentration values for each of the 3 biological replicates from one experiment and from 1 to 4 technical replicates. Variance analysis showed that the noise from the provided technical replicates was larger than that of the biological replicates. This tells us that one technical replicate is not sufficient for realistic quantitation. Without quantitative information we resorted to using the fold-change measurements with a cutoff of 1.2 fold change (up or down) based on the number of changes that we would expect to see in response to what is known about the mechanism of action of the drugs.

Only the highest drug concentration was considered. Changes in the phosphorylation of a protein were normalized to the total expression of that protein. If the total expression was not measured, the phosphorylation change could not be reliably determined, so we didn't attempt to explain those results. The one exception is the change in the Erks TEY site because the protein concentration of Erks rarely changes over 24 hour perturbations.

To map the data onto a PL model it is necessary to determine what each antibody actually detects and map this to PL terms. The antibodies used in the RPPA analysis were obtained from commercial suppliers and validated by the MD Anderson Cancer Center RPPA Core Facility. Information about the validation status and source of the antibodies was obtained from the Standard Antibody List downloaded from [2]. We determined the antibody targets by mapping the antibody name reported in the data set to the Official Antibody Name used in the Standard Antibody List. Specificity and site information was obtained from the supplier. The protein or family names of the target proteins were converted into Pathway Logic names and the sites were adjusted to agree with the canonical sequence of each protein in UniProt. In the case of protein families, letter codes were used to match all members, as described in section 2.

To explain the response to a drug treatment it is useful to know what the drug is, i.e. its chemical structure, to have clear experimental evidence of the target and its action on the target, and to know whether there are off-target effects. We were able to identify (find a PubChem identifier for) 8 of the 12 drugs used in the experiment. Subsequent literature search revealed solid evidence for proposed mechanisms of action for 5 of the 8. This is summarized in section 5 as part of the explanation of the data.

4 Inferring the SKMEL133 model

As discussed in section 2 our idea is to build the minimal model needed to explain the data, rather than attempting a full model of SKMEL133 cells. Thus we include as a minimum the proteins such that the change in protein expression or phosphorylation passed the 1.2 fold cutoff. We carried out (by hand) the fuzzy backwards collection starting from the changed occurrences, adding occurrences with a degradation modification to represent a possible cause of change in protein expression. For example rule `3823c`

```
rl[3823c.Irs1.degraded]:
Irs1-ubiq-phos!S270-phos!S307-phos!S636-phos!S1101@CLc
=>
Irs1-degraded@Sig
if Cul7@CLc
```

is collected to account for changes in `Irs1` expression level. This also introduces the protein `Cul7` into the model. Here we use informal rule notation where following the *if* are the controls (the required biological context) of the reaction.

```
rl[109c.Akts.by.Pdpk1]: Akts@CLc => Akts-phos!KTF@CLc if Pdpk1-act@CLc
```

Rule `109c` is collected to produce `Akts-phos!KTF`, which then introduces a requirement for `Pdpk1-act`. This can be satisfied by rule `3818c`

```
rl[3818c.Pdpk1.by.PIP3]: Pdpk1@CLc => Pdpk1-act@CLc if PIP3@CLm
```

which leads to collecting rule `3820c`

```
rl[3820c.PIP3.from.PIP2]: PIP2@CLm => PIP3@CLm if  Pi3k@CLi
```

to produce *PIP3*. This chain stops here, as `PIP2` is a common component and there are no rules producing the protein `Pi3k` so we assume it is expressed by SKMEL133 cells normally.

Collecting the occurrences that can not be produced by a rule we have a preliminary version of the initial state. SKMEL133 cells contain the constitutively active mutation `BrafV600E` so we replaced wild-type `Braf` with `BrafV600E`. They also have a homozygous deletion of `Pten`, so we eliminated `Pten`. The result, called the *SKMEL133dish*, contains 31 occurrences (listed in Appendix 1 of the techreport version).

As discussed in section 2 some iteration is required to achieve a connected set of rules because the curated rules reflect what experiments measured and may have different levels of detail, or need generalization. Also, the following rule was added to model the `BrafV600E` activity.

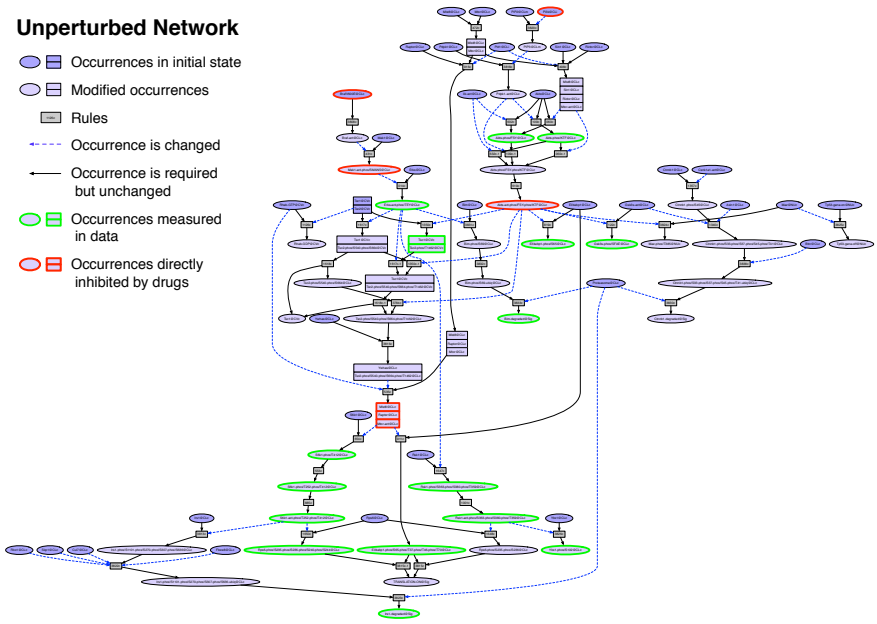


Fig. 3. The unperturbed SKMEL133 model.

```
rl[3808c.BrafV600E.act]: BrafV600E@CLc => Braf@act@CLc
```

This rule reflects the observation that the mutated form of Braf behaves like the active form of wild type Braf. This is a simplification which is adequate in the context of the current model, although it would fail if there were rules to deactivate Braf, since the mutated form can not be deactivated. After adding the above rule and generalizing some rules by hand, PLA is used to assemble the executable model, called the SKMEL133dishnet, shown in Figure 3. [2]

5 Explaining response to known drugs

As discussed in section 3, we selected 5 drugs for which we could determine a well-defined chemical id (PUBCHEM), and for which there is reasonable evidence for the proposed mechanism of action (determined by literature search): AktI12, PD0325901, PLX4720, Temsirolimus, and ZSTK474 (described in more detail below). For each of these drugs we determined occurrences that changed significantly using the fold change table from [10] and a fold change cutoff of 1.2 for increase and 0.8 for decrease as described in section 3. A table summarizing these changes is included in Appendix 1. Using the methods described in section

[2] Although in printed form the node labels are not readable, zooming in with a pdf reader reveals all the details.

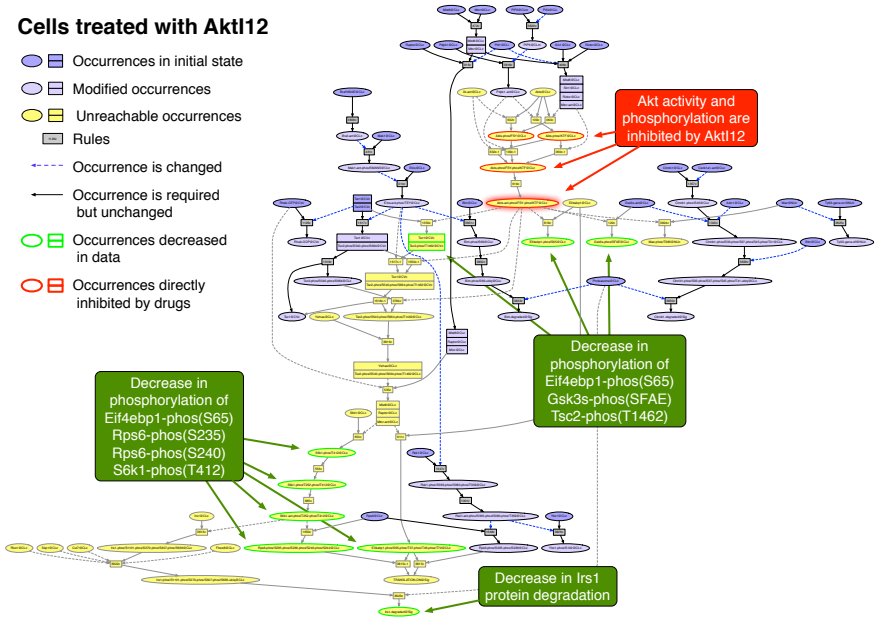


Fig. 4. The SKMEL133 model treated with AktI12.

2 we could explain 42 out of 107 changes in response to the 5 drugs. Many of the unexplained changes are in protein expression levels, which was generally not the focus of our curation efforts in the past. In the following we illustrate the analysis for AktI12 and Temsirolimus in some detail, and briefly summarize the results for the other three drugs. Recall that the SKMEL133 model and a guided tour allowing the user to reproduce these results and carry out other gedanken experiments are available for download or in the Online collection at [13].

5.1 Effects of AktI12

AktI12 (PubChemCID 10196499) is a reversible allosteric inhibitor of Akt1 and Akt2 which prevents the conformational change that permits phosphorylation and activation [11]. To model the effect of AktI12 we use PLA to block (avoid) the occurrence `Akts-act-phos!FSY-phos!KTF@CLc` in the SKMEL133 dishnet. Recall, this occurrence can be interpreted as Akt1 phosphorylated at S473/T308 and/or Akt2 phosphorylated at S474/T309 in the cytoplasm. Now we compute the resulting reachable network, and compare it to the untreated model to determine what has become unreachable.

Figure 4 shows the explanation as an annotated version of network produced by PLA in the context of the unperturbed model. It shows how drug perturbations interrupt the path between the initial state and the measured goals. The key in the figure describes the color coding in detail. Yellow coloring highlights the unreachable part of the SKMEL133 dishnet. Occurrences outlined in red

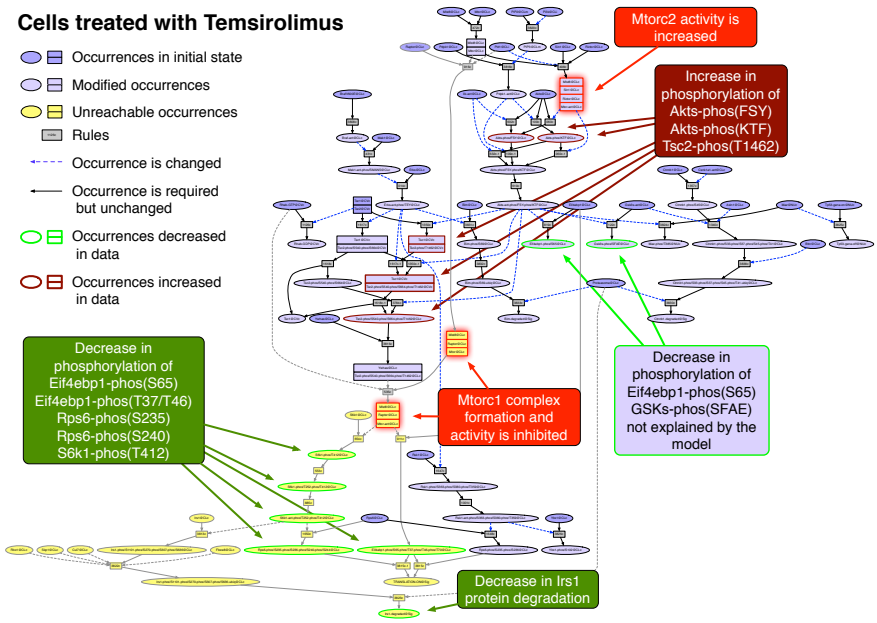


Fig. 5. The SKMEL133 model treated with Temsirolimus.

are directly inhibited by the drug. Occurrences outlined in green decrease in response to the drug. In particular the measured decrease in `Eif4ebp1-phos!S65`, `Eif4ebp1-phos!T37`, `Gsk3b-phos!S9`, `Gsk3s-phos!SFAE`, `Rps6-phos!S235`, `Rps6-phos!S240`, `S6k1-phos!T412`, and `Tsc2-phos!T1462` in response to AktI12 is explained by the unreachability of the corresponding occurrences. The increase in Irs1 protein expression is explained by the inhibition of the degradation of Irs1 by ubiquitination and degradation in the proteasome. The remaining changes are increases in protein expression of Cav1, Fn1, Pai1, and Tp53 and a decrease in Cox2 and CyclinB1, which are not represented in our model.

5.2 Effects of Temsirolimus

Temsirolimus (PubChemCID 23724530) inhibits Mtorc1 activity (a complex of Mtor, Mlst8, and Raptor) but enhances Mtorc2 activity (a complex of Mtor, Mlst8, Sin1, and Rictor) [5]. Figure 5 shows the annotated model of Temsirolimus response.

The model explains measured decrease in events downstream of `Mtorc1`: `Eif4ebp1-phos!T37`, `Rps6-phos!S235`, `Rps6-phos!S240`, `S6k1-phos!T412`, and `Irs1-degradation`. It also explains measured increase in events that are downstream of `Mtorc2`: `Akts-phos!FSY`, `Akts-phos!KTF`.

The model also predicts increases in `Eif4ebp1-phos!S65@CLc` (the data shows a decrease) and `Gsk3s-phos!SFAE@CLc` (the data shows no change). What might cause this discrepancy? A common cause of such discrepancy is a missing

control on the phosphorylation rule, either because there are no published experiments giving evidence, or because they have not yet been curated. It is also possible that there are alternative activities of the Akts. Note that the RPPA experiments do not measure activity directly. Unraveling this mystery is a topic of ongoing/future work.

5.3 Effects of PD0325901, PLX4720 and ZSTK474

PD0325901 (PubChemCID 9826528) is an allosteric inhibitor of Mek1 and Mek2 kinase activity [14]. To represent the effects of PD0325901, the SKMEL133 model can be blocked at the occurrence `Mek1-act-phos!SMANS@CLc` which can be interpreted as Mek1 phosphorylated at S218 and S222. Although the antibody used in generating the data identifies both phospho-Mek1 and phospho-Mek2, the STM DKB lacks sufficient datums to include Mek2 in the rules. The resulting unreachable set explains decreases in `Erks-phos!TEY`, `Rps6-phos!S235`, `Rps6-phos!S240`, `Rsk1-phos!T359`, `S6k1-phos!T412`, and `Ybx1-phos!S102`. Using the decrease in `Bim-degraded@Sig`, it also explains the increase in Bim protein expression.

PLX4720 (PubChemCID 24180719) binds to the ATP binding site of active Braf and Raf1. It is 10 times more effective towards BrafV600E than wild-type Braf or Raf1. At the concentration used to produce the dataset (120 nM) it should be more effective on BrafV600E than Raf1 [17]. As expected, the perturbation profile PLX4720 is almost identical to that of PD0325901, since `Braf` is responsible for phosphorylation of `Meks`.

ZSTK474 (PubChemCID 11647372) inhibits all four isoforms of the catalytic subunit of Pi3k [5]. This then inhibits `Akts-phos!FSY-phos!KTF@CLc` via decrease in the activity of the upstream kinase Pdpk1. The perturbation profile is the same as that for AktI12 except that the decrease in `Akts-phos(FSY)` and `Akts-phos(KTF)` are caused by a decrease in the activity of the upstream kinase `Pdpk1`.

6 Conjecturing mechanisms of unknown drugs

We looked at two of the drugs that were not identifiable: (1) a drug referred to as SR with claimed target Src (although the data shows no significant effect on measured Src), and (2) a drug referred to as RY, with claimed target CDK4 although no form of CDK4 was measured. Our approach to analyzing the data for these unknown drugs consisted of the following steps.

1. Identify changed occurrences in the model (for protein expression we use change of opposite sign in degradation of the protein as a representative).
2. Form the subnet containing all the pathways to these occurrences
3. For each occurrence with negative change, compute the subnet of pathways leading to that occurrence and use the pathway analysis tool to list the rules and occurrence that are single knock outs (i.e., if removed from the network the goal occurrence is no longer reachable).

4. Make a table with columns corresponding to the negatively changed occurrences and rows labeled by the knockouts. The entry in a cell is 1 if the knockout labeling the row is in knockout list of the occurrence labeling the column and 0 otherwise.

Now we want minimal subsets of rows that add to 1 for each column. Then inhibiting each of the row labels in such a subset will explain all the negative changes. Of these minimal sets, we prefer those that are furtherest down stream, since otherwise there are likely to be off-target effects.

Given a candidate drug target list, we need to check if this predicts changes consistent with the data. This can be done as for the drugs with known action. Namely, starting with the unperturbed model (the SKMEL133 dishnet), knock out the hypothesized drug target(s), compute the subnet, compare to the unperturbed net to see what is missing. Clearly, the set of occurrences used to generate the knockout lists will be unreachable and thus consistent with the hypothesized targets. Are the other unreachables plausible? We also need to look for explanations for occurrences that increased, such as blocked or diverted branches. As for the case of drugs with known targets we use the 1.2/.8 fold cutoff to determine the list of changed occurrences, and require phosphorylation change relative to protein expression change to meet the cut off criteria. In the following we discuss the for SR. The results for RY can be found in the techreport version of the paper.

6.1 Analysis of the effects of SR

From the data for the drug SR we determined 2 instances of increase in protein expression (1 is in the model), 3 instances of decrease in protein expression (none in the model), 2 instances of increase in phosphorylation (none in the model) and 8 instances of decrease in phosphorylation (6 in the model). Converting the one increase in protein expression to a decrease in degradation, the decreases represented in the model to consider are: `Bim-degraded@Sig`, `Eif4ebp1-phos!S65-phos!T37-phos!T46-phos!T70@CLc`, `Eif4ebp1-phos!S65@CLc`, `Erks-phos!TEY@CLc`, `Gsk3s-phos!SFAE@CLc`, `Rps6-phos!S235@CLc`, and `S6k1-phos!T412CLc`.

After computing the subnet containing these changed occurrences and computing the knockouts for each of these occurrences, we find that no single knockout can explain the observed decreases. There are many double knockouts that can explain the decreases. They all involve blocking Mek1 activity and Akts activity, either directly or by an upstream effect. Thus the minimal pair is

```
[Akts-phos!FSY-phos!KTF@CLc, Mek1-act-phos!SMANS@CLc]
```

Although these occurrences are not decreased in response to SR, it is quite possible that the drug blocks their action and hence causes the observed downstream effects. Choosing targets upstream of this pair, say [`Braf-act@CLc`, `Pi3k@CLi`] would be inconsistent with the observed data as in this case one should observe a decrease in the phosphorylation of Akts and Mek1.

Now we check whether blocking this pair of occurrences is consistent with the measured response to SR. We start with the unperturbed model, knockout (avoid) the conjectured pair of occurrences, compute the resulting reachable subnet, and the unreachable set. The following occurrences that are predicted by the model to decrease are measured:

- `Irs1-degraded@Sig`: protein expression did not change.
- Occurrences involving `Rsk1-phos!T359`: neither `Rsk1` protein expression or `Rsk1-phos!T359` changed. Note that the antibody for `Rsk1` is labeled "use with caution" and the antibody for `Rsk1-phos!T359` is not validated.
- `Ybx1-phos!S102@CLc`: This decreased, which is consistent. The total protein for `Ybx1` was not measured, so it was not included in the list of changes to explain.

6.2 Analysis of the effects of RY

From the data for the drug referred to as RY we determined 14 instances of increase in protein expression (1 is in the model), 9 instances of decrease in protein expression (none in the model), 8 instances of increase in phosphorylation (3 in the model) and 5 instances of decrease in phosphorylation (4 in the model). Just from the numbers it seems this drug has a rather different effect on SKMEL133 cells than SR.

Converting the one increase in protein expression to a decrease in degradation, the decreases represented in the model to consider are: `Bim-degraded@Sig`, `Eif4ebp1-phos!S65-phos!T37-phos!T46-phos!T70@CLc`, `Eif4ebp1-phos!S65@CLc`, `Gsk3s-phos!SFAE@CLc`, and `Irs1-phos!S1101-phos!S270-phos!S307-phos!S636@CLc`.

Looking at the knockouts for these occurrences we see that blocking Akts activity explains everything but the increase in `Bim` expression. If we knockout Akts activity, the following measured phosphorylations become unreachable: `Rps6-phos(S235)`, `Rps6-phos(S240)`, `S6k1-phos!T412`, and `Tsc2-phos!T1462`. The data shows no significant change in these entities. The trouble with this explanation is that `Akts` protein expression decreases substantially, while the levels of the phosphorylated forms increases relative to the total `Akts` protein. It is possible that the drug inhibits the activity of the phosphorylated form.

In our model `Bim` degradation is controlled by activity of `Erks`, which is controlled by activity of `Mek1`, which is controlled by `Braf`. Although activity of these proteins is not measured, it is generally believed that phosphorylation is required and the data shows no change in the relevant phosphorylation levels. Thus our model does not provide an explanation for the increase in `Bim` expression consistent with other changes.

`Ctnnb1-phos!S33` increases, which in the model leads to an increase in `Ctnnb1-degraded` and hence we should observe a decrease in `Ctnnb1`. The measured level is .82 which is consistent with our cutoff. According to rule `1340c`, an increase in `Gsk3s-act` would explain the measured increase in `Ctnnb1-phos!S33`. The decrease in `Gsk3s-phos!SFAE` (a consequence of hypothesized decrease in

Akts activity) could explain an increase in `Gsk3s-act`, since it is not being used up.

One final observation about RY. In addition to the case of `Akts` protein expression decreasing while the relative phosphorylation levels increase, this happens for `Accs`, `Atr`, and `P38s`. It would be interesting to know of other drugs that exhibit this pattern.

7 Related work

We focus on the use of RPPA data to analyze cellular systems. Existing work generally focuses on inferring network models that fit the data in order to identify interactions and possible causal relations among responding proteins and/or to use the resulting models to predict response to new perturbations. To the best of our knowledge our approach of using an existing curated model to explain the mechanisms underlying cellular response to drugs, and consequently validate or find gaps or problems with the parts of the model, or to hypothesize alternative actions of a drug is unique.

The work presented in [10] is the source of the data explained in the present paper. The work was motivated by the problem of drug resistance, particularly in cancers. The paper describes a combined experimental/ computational perturbation biology method to look for anti-resistant target combinations. The experiment was described in section 3, with cells being treated by pair-wise combinations of drugs as well as the single drug treatments. A space of executable ODE models corresponding to influence network topologies with weighted edges are derived from the data using belief propagation techniques. The process is seeded with a prior network extracted from Pathway Commons using the PERA tool [1]. The 4000 best models were selected to make predictions of phenotypic effects of thousands of combinations of perturbations. As a result they propose cMyc as a co-target of Mek or Braf.

The results of the HPN-DREAM network inference challenge are summarized in [9]. This challenge focused on learning causal influences in signaling networks. The objective here was to train models capable of predicting context-specific phosphoprotein time courses, in contrast to the Big Mechanism objective to provide mechanistic explanations for the effects of perturbations. Participants were provided with RPPA phosphoprotein data from four breast cancer cell lines under eight ligand stimulus conditions combined with three kinase inhibitors and a vehicle control (dimethyl sulfoxide). Data for each biological context (cell line, stimulus combination) comprised time courses for approximately 45 phosphoproteins. Models were assessed using context-specific test data that were obtained under a different intervention (inhibition of the kinase mTOR). The best-scoring method for the experimental data task, Prophetic Granger with heat diffusion prior, used a prior network created by averaging similarity matrices. The matrices were obtained via simulated heat diffusion applied to links derived from the Pathway Commons database. The best AUROC score was just under .8 while most methods scored between .5 and .6. While some of the models succeeded

in reasonable predictive power, more work is needed to obtain more detailed mechanistic explanations.

Reverse Phase Protein Arrays (RPPAs or RPLAs) were used in [8] to profile signaling proteins in 56 breast cancers and matched normal tissue as a method to discover phosphorylation-mediated signal transduction patterns in human tumor samples. The paper discusses the process of validating antibodies (100 antibodies validated of 400 screened), and methods for quantitation of data in some detail. Unsupervised hierarchical clustering was used as a first step in discovering patterns of co-regulation. The hierarchy was cut to yield twelve clusters, which were mapped onto pathways derived from Gene Network Central Pro. This revealed a cluster involving increased abundance of the Axl receptor tyrosine kinase (RTK) and the cMet RTK pathway. Structured Bayesian inference was then used to further analyze this cluster to find the interaction network topology with good generalization properties and that best classified cancer vs non-cancer data. The results suggested two cancerous categories: 1) where MET is highly phosphorylated and cRAF is always highly phosphorylated and 2) where MET phosphorylation is low and cRAF phosphorylation is low at sites consistent with cRaf inactivation.

8 Conclusions and Future Directions

We have shown how the Pathway Logic STM model, capturing what we know about intracellular signal transduction, can be used to explain experimental results. The rules used in the model are derived from experimental results, so if the model were complete we should be able to use the network derived from exponentially growing cultured cells to trace the paths from a known perturbation to the measured effects. In some of the cases, we were successful. Our successes were predominantly in the phosphorylation cascades and protein degradation events used in growing cells. We were less effective in explaining the decreases in expression of proteins due to inhibition of translation or transduction, or changes in the cell cycle. There is still a lot of experimental evidence in the literature to collect and make into rules. There are still a lot of experiments that need to be performed and published. Work is in progress to automate this fuzzy backwards and forwards collection carried out by hand to generate the SKMEL133 model. We are also investigating representation of executable models, network perturbations, and experimental observations as constraints and using abductive reasoning to generate potential explantations. This would unify the treatment of various aspects and help automatic the end to end reasoning process.

One caveat, not all of the unexplained results are due to an incomplete model. Only one experiment was performed so the probability that the results could be reproduced cannot be measured. Although 3 biological replicates were used - no information about the variance were provided. In addition, we obtained the mechanism of action of the drugs from a small sampling of the literature. Any of the drugs could have additional effects that we did not find.

Learning about how a cell works is still a work in progress. The Pathway Logic STM model is a tool designed to help. Hopefully it does.

Appendix 1.

Drug	Change	Target	explained by model?
AktI12	-	Akts-phos(FSY)	no
AktI12	-	Akts-phos(KTF)	no
AktI12	+	Cav1-prot-exp	no
AktI12	-	Cox2-prot-exp	no
AktI12	-	CyclinB1-prot-exp	no
AktI12	-	Eif4ebp1-phos(S65)	yes
AktI12	-	Eif4ebp1-phos(T37/T46)	yes
AktI12	-	Erks-phos(TEY)	no
AktI12	+	Fn1-prot-exp	no
AktI12	-	Gsk3b-phos(S9)	yes
AktI12	-	Gsk3s-phos(SFAE)	yes
AktI12	+	Igfbp2-prot-exp	no
AktI12	+	Irs1-prot-exp	yes
AktI12	+	Pai1-prot-exp	no
AktI12	+	Pax2-prot-exp	no
AktI12	-	Plk1-prot-exp	no
AktI12	-	Rps6-phos(S235)	yes
AktI12	-	Rps6-phos(S240)	yes
AktI12	-	S6k1-phos(T412)	yes
AktI12	+	Tp53-prot-exp	no
AktI12	-	Tsc2-phos(T1462)	yes
PD0325901	-	Akts-phos(FSY)	no
PD0325901	-	Akts-phos(KTF)	no
PD0325901	+	Ampkas-phos(LRtSC)	no
PD0325901	+	Bim-prot-exp	yes
PD0325901	+	Cdkn1b-prot	no
PD0325901	-	Cox2-prot-exp	no
PD0325901	+	Ctnnb1-prot-exp	no
PD0325901	-	CyclinB1-prot-exp	no
PD0325901	-	CyclinD1-prot-exp	no
PD0325901	-	Erks-phos(TEY)	yes
PD0325901	+	Foxo3-prot-exp	no
PD0325901	-	Mek12s-phos(SMANS)	no
PD0325901	-	Myc-prot-exp	no
PD0325901	-	Pai1-prot-exp	no
PD0325901	-	Plk1-prot-exp	no
PD0325901	-	Rb1-phos(S608/S807/S811)	no
PD0325901	-	Rps6-phos(S235)	yes
PD0325901	-	Rps6-phos(S240)	yes
PD0325901	-	Rsk1-phos(T359)	yes
PD0325901	-	S6k1-phos(T412)	yes
PD0325901	+	Stat5a-prot-exp	no
PD0325901	+	Tp53-prot-exp	no
PD0325901	-	Ybx1-phos(S102)	yes
PLX4720	+	Bim-prot	yes
PLX4720	+	Cdkn1b-prot	no
PLX4720	-	Cox2-prot-exp	no
PLX4720	+	Ctnnb1-prot-exp	no
PLX4720	-	CyclinB1-prot-exp	no
PLX4720	-	CyclinD1-prot-exp	no
PLX4720	-	Erks-phos(TEY)	yes
PLX4720	-	Mek12s-phos(SMANS)	yes
PLX4720	-	Myc-prot-exp	no
PLX4720	-	Rb1-phos(S608/S807/S811)	no
PLX4720	-	Rps6-phos(S235)	yes
PLX4720	-	Rsk1-phos(T359)	yes
PLX4720	-	S6k1-phos(T412)	yes
PLX4720	+	Tp53-prot-exp	no
PLX4720	-	Ybx1-phos(S102)	yes
Temsirolimus	-	Akts-prot-exp	no
Temsirolimus	+	Akts-phos(FSY)	yes
Temsirolimus	+	Akts-phos(KTF)	yes
Temsirolimus	-	Ampkas-phos(LRtSC)	no
Temsirolimus	+	Col6a1-prot-exp	no
Temsirolimus	+	Cox2-prot-exp	no
Temsirolimus	+	Cav1-prot-exp	no
Temsirolimus	-	CyclinB1-prot-exp	no
Temsirolimus	+	CyclinE1-prot-exp	no
Temsirolimus	-	Eif4ebp1-phos(S65)	no
Temsirolimus	-	Eif4ebp1-phos(T37/T46)	yes
Temsirolimus	-	Eif4ebp1-phos(T70)	yes
Temsirolimus	+	Fn1-prot-exp	no
Temsirolimus	+	Irs1-prot-exp	yes
Temsirolimus	-	Mek12s-phos(SMANS)	no
Temsirolimus	+	Pai1-prot-exp	no
Temsirolimus	-	Plk1-prot-exp	no
Temsirolimus	-	Rps6-prot-exp	no
Temsirolimus	-	Rps6-phos(S235)	yes
Temsirolimus	-	Rps6-phos(S240)	yes
Temsirolimus	-	S6k1-phos(T412)	yes
Temsirolimus	+	Tp53-prot-exp	no
Temsirolimus	+	Tsc2-phos(T1462)	yes
ZSTK474	-	Acc1-prot-exp	no
ZSTK474	+	Akts-prot-exp	no
ZSTK474	-	Akts-phos(FSY)	yes
ZSTK474	-	Akts-phos(KTF)	yes
ZSTK474	+	Cav1-prot-exp	no
ZSTK474	+	Chek2-phos(T68)	no
ZSTK474	-	Cox2-prot-exp	no
ZSTK474	-	CyclinB1-prot-exp	no
ZSTK474	+	CyclinD1-prot-exp	no
ZSTK474	-	Eif4ebp1-phos(S65)	yes
ZSTK474	-	Eif4ebp1-phos(T37/T46)	yes
ZSTK474	+	Fn1-prot-exp	no
ZSTK474	-	Gsk3b-phos(S9)	yes
ZSTK474	-	Gsk3s-phos(SFAE)	yes
ZSTK474	+	Irs1-prot-exp	yes
ZSTK474	-	Mek12s-phos(SMANS)	no
ZSTK474	+	Pai1-prot-exp	no
ZSTK474	+	Plk1-prot-exp	no
ZSTK474	-	Rb1-phos(S608/S807/S811)	no
ZSTK474	-	Rps6-prot-exp	no
ZSTK474	-	Rps6-phos(S235)	yes
ZSTK474	-	Rps6-phos(S240)	yes
ZSTK474	-	S6k1-phos(T412)	yes
ZSTK474	+	Stat5s-phos(DGYV)	no
ZSTK474	+	Tp53-prot-exp	no

Fig. 6. Summary of changes in response to the 5 known drugs.

The occurrences in the SKMEL133 dish.

`Akts@CLc, Axin1@CLc, Bim@CLc, BrafV600E@CLc, Btrc@CLc, Csnk1a1-act@CLc, Ctnnb1@CLc, Cul7@CLc, Eif4ebp1@CLc, Erk5@CLc, Erks@CLc, Fbxw8@CLc, Gsk3s-act@CLc, Ilk-act@CLc, Irs1@CLc, Maz@NUc, Mdm2@CLc, Mek1@CLc, Mlst8@CLc, Mtor@CLc, Pdpk1@CLc, Pi3k@CLi, PIP2@CLm, Pld1@CLi, Proteasome@CLc, Raptor@CLc, Rbx1@CLc, Rheb-GTP@CVc, Rictor@CLc, Rps6@CLc, Rsk1@CLc, S6k1@CLc, Sin1@CLc, Skp1@CLc, Tp53-gene-on@NUc, Tsc1:Tsc2@CVc, Ybx1@CLc, Ywhas@CLc`

References

1. B. A. Aksoy. BioPax prior information provider (PERA), 2017. Accessed: 2017-04-26.
2. M. D. A. C. Center. Antibody information and protocols, 2017. Accessed: 2017-04-25.
3. M. D. A. C. Center. Rppa core facility, 2017. Accessed: 2017-04-25.
4. M. Clavel, F. Durán, S. Eker, P. Lincoln, N. Martí-Oliet, J. Meseguer, and C. Talcott. *All About Maude: A High-Performance Logical Framework*. Springer, 2007.
5. R. Dienstmann, J. Rodon, V. Serra, and J. Tabernero. Picking the point of inhibition: a comparative review of PI3K/AKT/mTOR pathway inhibitors. *Mol Cancer Ther*, 13(5):1021–1031, 2014.
6. R. Donaldson, C. Talcott, M. Knapp, and M. Calder. Understanding signalling networks as collections of signal transduction pathways. In *Computational Methods in Systems Biology*, 2010.
7. J. Fisher and T. A. Henzinger. Executable cell biology. *Nature Biotechnology*, 25(11), 2007.
8. T. S. Gujral, R. L. Karp, A. Finski, M. Chan, P. E. Schwartz, G. MacBeath, and P. Sorger. Profiling phospho-signaling networks in breast cancer using reverse-phase protein arrays. *Oncogene*, 32(29):3470–3476, 2013.
9. S. M. Hill et al. Inferring causal molecular networks: Empirical assessment through a community-based effort. *Nature Methods*, 13(4), 2016.
10. A. Korkut, W. Wang, E. Demir, B. A. Aksoy, X. Jing, E. J. Molinelli, O. Babur, D. L. Bemis, S. S. Onur, D. B. Solit, C. A. Pratilas, and C. Sander. Perturbation biology nominates upstream-downstream drug combinations in raf inhibitor resistant melanoma cells. *Elife*, 18(4), 2015.
11. L. Logie, A. J. Ruiz-Alcaraz, M. Keane, Y. L. Woods, J. Bain, R. Marquez, D. R. Alessi, and C. Sutherland. Characterization of a protein kinase b inhibitor in vitro and in insulin-treated liver cells. *Diabetes*, 56(9):2218–2227, 2007.
12. V. Nigam, R. Donaldson, M. Knapp, T. McCarthy, and C. L. Talcott. Inferring executable models from formalized experimental evidence. In O. F. Roux and J. Bourdon, editors, *13th Computational Methods in Systems Biology*, volume 9308 of *Lecture Notes in Computer Science*, pages 90–103. Springer, 2015.
13. Pathway Logic, 2016. Accessed: 2016-11-10.
14. A. A. Samatar and P. I. Poulikakos. Targeting RAS-ERK signalling in cancer: promises and challenges. *Nat Rev Drug Discov*, 13(12):928–942, 2014.
15. K. Schmidt. LoLA: A Low Level Analyser. In M. Nielsen and D. Simpson, editors, *Application and Theory of Petri Nets, 21st International Conference (ICATPN 2000)*, volume 1825 of *LNCS*, pages 465–474. Springer-Verlag, 2000.
16. C. Talcott. The pathway logic formal modeling system: Diverse views of a formal representation of signal transduction. In Q. Wang, editor, *Workshop on Formal Methods in Bioinformatics and Biomedicine*, 2016.
17. J. Tsai and etal. Discovery of a selective inhibitor of oncogenic b-raf kinase with potent antimelanoma activity. *Proc Natl Acad Sci U S A*, 105(8):3041–3046, 2008.