PAPER • OPEN ACCESS

DARPA's Big Mechanism program

To cite this article: Paul R Cohen 2015 *Phys. Biol.* **12** 045008

View the article online for updates and enhancements.

Related content

- Modeling for (physical) biologists: an introduction to the rule-based approach
  Lily A Chylek, Leonard A Harris, James R Faeder et al.

- From physics to pharmacology?
  Richard J Allen and Timothy C Elston

- State--time spectrum of signal transduction logic models
  Aidan MacNamara, Camille Terfve, David Henriques et al.

Recent citations

- Emerging whole-cell modeling principles and methods
  Arthur P Goldberg *et al*

- Dynamic Influence Networks for Rule-Based Models
  Angus G. Forbes *et al*

- LitPathExplorer: a confidence-based visual text analytics tool for exploring literature-enriched pathway models
  Axel J Soto *et al*

Physical Biology

CrossMark

PAPER

DARPA's Big Mechanism program

Paul R Cohen

Defense Advanced Research Projects Agency (DARPA)

E-mail: paulrcohen@gmail.com

Abstract

Reductionist science produces causal models of small fragments of complicated systems. Causal models of entire systems can be hard to construct because what is known of them is distributed across a vast amount of literature. The Big Mechanism program aims to have machines read the literature and assemble the causal fragments found in individual papers into huge causal models, automatically. The current domain of the program is cell signalling associated with Ras-driven cancers.

When it comes to complicated systems such as the climate, and economic, ecological and biological systems, our influence runs ahead of our understanding. This can be perilous. The Defense Advanced Research Projects Agency (DARPA) created the Big Mechanism program (BMP) to develop technology for constructing, understanding and reasoning about big, complicated systems. The focus of the program is cancer signaling pathways, but the technology is intended to apply more generally. This paper is a progress report on the first year of research in the BMP.

Imagine a future in which computers read the literature of cancer biology, extract fragments of causal mechanisms from publications, assemble these mechanisms into executable models of unprecedented scale and fidelity, use these models to explain and predict aspects of cancer biology, and even test these predictions *in vitro*. The future is not yet here (the BMP started in July, 2014) but we can see its outlines and our first steps are clear. Roughly half of the program's effort goes into machine reading, the other half into assembling large, causal models and reasoning with them[1]. The BMP makes good use of existing online sources of biological knowledge (e.g., [4, 13, 14, 19, 53, 65]) as well as existing machine reading and information extraction methods (e.g.,

[2, 8, 41, 52, 54, 55, 63]); while it develops representations and inference methods for mechanistic biological models, methods for machines to understand the claims and evidence in papers, algorithms for inferring causal relationships given data, and robotic techniques for testing computer-generated hypotheses. The BMP focuses on Ras-driven cancers [31, 35, 67], as these are among the least well-understood and most lethal cancers.

1. What is a big mechanism?

Big mechanisms are mechanistic models of complicated systems. Here, 'complicated' systems are those that have too many elements and relationships, or too many possible current or future states, to be easily comprehended by people. Here, 'mechanistic' is a *pragmatic* attribute of models, having to do with how a model is used, rather than a *syntactic* or *semantic* attribute, having to do with the form or content of a model. Consider a model $M: f(x)(ax + \epsilon) \Rightarrow \hat{y}$. This says, a function f takes input x, multiplies it by a, adds ϵ, and returns the result as y. The 'hat' over y says the purpose of the model is to make predictions about y. The *accuracy* of M, as measured by $(\hat{y} - y)$, is a semantic attribute. The linearity of M is a syntactic attribute. A mechanistic model is one that makes a claim about mechanisms, so it might be written $M: \hat{f}(x)(ax + \epsilon) \Rightarrow y$, meaning that M claims that some mechanism f in the world is approximated by a mechanism $\hat{f}$. (Specifically, M is read as the claim that f in the world multiplies x by a and adds ϵ.)

---

[1] The participants are: University of Arizona (Surdeanu, PI); Carnegie Mellon University (Hovy, PI and Miskov-Zivanov, PI); University of Chicago (Rzhetsky, PI); Cycorp, (Lenat and Witbrock, PIs); Harvard University (Fontana, PI and Sorger, PI); IHMC (Allen, PI); USC/ISI (Marcu, PI); Leidos (Morgenstern and Demir, PIs); SIFT (Burstein, PI); SRI International (Freitag and Rohwer, PIs). MITRE (Hirschman, PI) conducts evaluations. The Army Research Lab (Purushotham, PI) provides contracting support.
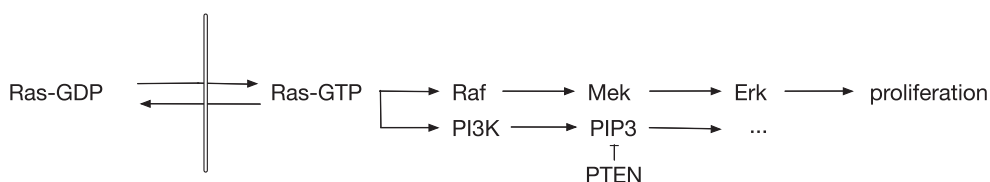
Figure 1. An incomplete Boolean network model of Ras signaling (excerpted from [67]).

Mechanistic models make claims about mechanisms; they might secondarily make claims about measurements or observations. Nonmechanistic models make no claims about mechanisms, only about measurements or observations. Many–perhaps most–applications of 'big data' depend on non-mechanistic models, whereas the BMP is explicitly about mechanistic models. The primary question in the BMP is not whether one can *predict* the outputs given the inputs (i.e., $(\hat{y} - y)$ is small) but whether one can *explain how* the inputs give rise to outputs (i.e., $(\hat{f} - f)$ is small).

Mechanistic models are *compositional* in two ways: the outputs of one component can be inputs of another, and components can be built from other components. Not all the components of big mechanisms will be mechanistic. Some will be non-mechanistic because their mechanisms are unknown; others will elide mechanistic details for the sake of brevity or relevance or computational tractability. Walter Fontana, one of the participants in the BMP, distinguishes between 'boxed' and 'unboxed' components. A boxed component specifies its inputs and outputs, whereas an unboxed component additionally specifies how the inputs are transformed into outputs. In the cell signaling domain, inputs and outputs are usually abundances of genes, proteins, complexes and so on; whereas mechanisms ground out in molecular interactions. Thus, a boxed model can say that increasing A causes an increase in complex C, whereas an unboxed version of the same model would have to specify how the complex forms. Unboxing a component entails adding mechanism, but unboxed models are not necessarily mechanistic in the sense defined above: entailing a claim that the mechanism is *correct*.

Ideally, the BMP will produce mechanistic models that make strong predictions about the behavior of cell signaling networks through a combination of logical or probabilistic reasoning and powerful simulations. Boolean networks (also called interaction maps [62])—like the one shown in figure 1, only much bigger—fall somewhat short of this ideal: they do not capture the kinetics of signaling, but can still be useful because they summarize 'who talks with whom'; they are the starting point for more expressive graphical representations and interactomes and scaffolds [19, 34, 39, 57, 64, 65]; and they are an appropriate target representation for some machine reading. Moreover, some network structures produce interesting behaviors given only qualitative dynamics; for example, feed forward networks can produce signal propagation that depends on relative rather than absolute levels of signal, consistent with Weber's law [25].

However, there are more expressive frameworks than Boolean networks, such as ordinary differential equation (ODE) models, curation languages such as BEL [7] and BioPAX [10, 19], graphical languages such as SBGN [44, 64], and rule-based or agent-based frameworks such as Kappa [16–18, 23], PySB [3, 46], BioNetGen [22] and Pathway Logic [69, 70]. Frameworks are commonly organized along a quantitative-to-qualitative dimension, with ODEs at the quantitative end and Boolean networks at the most qualitative end (e.g., [62]). Within the BMP, there is a sense that ODE models would be nice to have, but they are too unwieldy and hard to maintain for large systems and demand information about kinetics that we often lack; while Boolean networks say nothing about kinetics or complex control of processes; so, ideally, models should occupy a middle region of the quantitative-to-qualitative dimension. In this region, a host of other issues influence the choice of a modeling framework. Systems biologists make tradeoffs between the completeness of models, the uncertainty associated with the parameters of these models, the amount of data available for estimating these parameters; the kinds of inference supported by the models, and the tractability of computation with the models; and the 'readability' of a model by other biologists, and the extensibility, maintainability and reusability of models. Not surprisingly, the answer to the question 'which is the best modeling framework?' is 'it depends on one's purpose'. (See [1, 3, 29, 30, 62, 66] for good introductions to these issues.)

The BMP intentionally does not commit to a single standard for formal biological models. At present, BMP researchers are working with frameworks designed for curation (BEL, BioPAX, and Pathway Logic datums) and frameworks designed for modeling and simulation (Kappa, Pathway Logic rules, and PySB). The former are representations of what people write about biology, whereas the latter are representations of biology, itself. For example, in BEL one asserts A `directlyDecreases` B, which 'indicate[s] that increases in A have been observed to cause decreases in B' [7]. This is not a mechanistic model, it is a formal description of an experimental observation. BEL neither requires its users to specify mechanisms nor

Figure 2. A rough architecture for Big Mechanism systems.
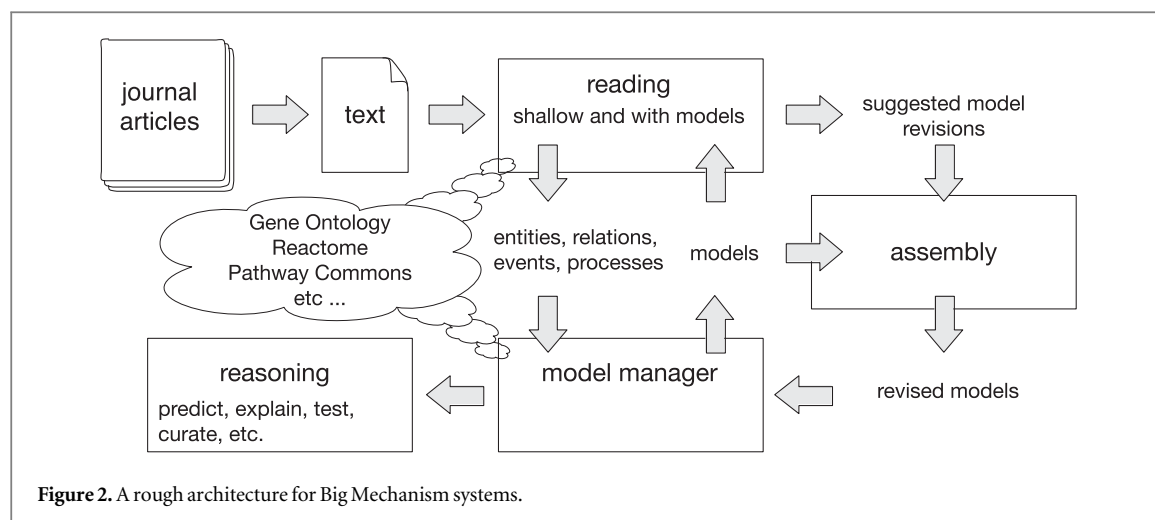
makes it easy to do so. In BioPAX, one can assert attributes of mechanisms; for example, one can say A inhibits B competitively, meaning that A binds to the enzyme that catalyzes B or binds to B and prevents catalysis. But, again, these are assertions *about* mechanisms, not mechanisms proper. In contrast, Kappa is not a language for curating experimental results or attributes of mechanisms; it is a language for representing mechanisms. In Kappa, A either inhibits B or it does not, depending on whether the rules in a Kappa model implicitly represent a chain of events in which more of A results in less of B. The inhibition might be competitive, if the Kappa model includes rules that allow A to bind to the substrate of B and block out the enzyme that would otherwise bind to B. In Kappa, one describes how the biology works, and then runs the model to discover what the biology does. The same can be said for rules in Pathway Logic and programs in PySB: they represent mechanisms, proper, not assertions about mechanisms.

Each of these modeling frameworks is designed to capture some but not all of the details of the underlying biochemistry. These approximations are necessary because of the combinatorial complexity of molecular interactions and the lack of empirical data to calibrate model parameters for these interactions. Also, there is some concern that the behavior of a model does not constrain its parameters because, empirically, many sets of model parameters can produce similar model behaviors [26], although causes of this effect are disputed [72]. So for the time being, BMP models of signaling will be 'coarse' in various ways. The challenge will be to manage multiple drafts of models in a way that makes them amenable to revision by both humans and machines [19, 43, 46, 47].

In addition to models of biological processes, the BMP is developing novel representations for the contents of tables, for the structure of experimental protocols [58], for typical statistical analyses, for rhetorical devices and epistemic or qualified assertions [20, 37, 63], and for other kinds of contents of research

papers. These kinds of content are often 'meta' to the biological models, but they are important elements of publications because they establish the reasons to believe or doubt models. If a paper says, 'mutant and wildtype responded differently to *x*, suggesting a therapeutic effect of *x* on *y*,' then a Big Mechanism system should extract both a model fragment that relates *x* and *y* and also the uncertainty conveyed by the 'suggesting' construction.

2. Big Mechanism systems

Several consortia of academic and industry researchers are building Big Mechanism systems. The technologies and use cases for these systems vary, but all of the systems are being developed along the lines shown in figure 2. Source texts will be read by machines, and the content of these texts will suggest revisions to prior models. Machine reasoning about whether and how to modify models is called assembly. Some modifications of models might create distinct new models (e.g., variants of a model might be created for different species of a protein) so the BMP will have to develop model management techniques. Finally, BMP systems will reason with models; for example, finding similar models, or finding generalities across models, or identifying potentially druggable proteins or pathways. To the extent possible, all the objects in BMP models will be normalized to appropriate ontologies and other resources.

At this early point in the program (which started in July, 2014) there has not been much progress on model management and reasoning (figure 2), but work on reading and assembly is underway.

Big Mechanism systems are expected to read as researchers do: with models in mind. The model a system is reading about is called its *prior* model. We distinguish *shallow* reading—which discovers the entities, relations, events and processes in text—from *deep* reading (also called *reading with models*) which discovers what the text says about prior models.

Ideally, this distinction should not exist, but it is a curious shortcoming that most algorithms for text mining and information extraction do not use prior models. Indeed, one goal of the BMP is to make shallow reading algorithms 'deeper' by using prior models to help identify entities and relations in text. For example, if text contains an ambiguous term, but one sense of the term denotes a protein in a prior model, then it's a good guess that this is the correct sense in the context of the prior model.

One might think that machines would have little difficulty reading papers in cancer biology because the language they contain seems very 'close' to biological phenomena. To a human, a sentence like, 'Ras acts as a molecular switch that is activated upon GTP loading and deactivated upon hydrolysis of GTP to GDP' seems clear enough, but consider what a machine must do to understand it (see also [32]):

Named entity recognition: figure out the entities—Ras, GTP and GDP—and associate them with standard ontology entries when possible, noting that some terms denote families.

Coreference resolution and anaphora: find multiple references to the same entity (e.g., the second mention of GTP refers to the same entity as the first one) and anaphoric references (e.g., 'that' refers to 'molecular switch,' which itself refers to Ras);

Event resolution and event coreference resolution: recognize that words such as activation, loading, and hydrolysis denote events (even though these are nouns!) and that two different terms (e.g., GTP loading and hydrolysis) can refer to different parts of one bigger event.

Parsing and/or semantic role labeling: correctly place entities in their appropriate roles in these events; for example, the subject of activation is the molecular switch, but the subject of GTP loading is not GTP but the switch.

Discourse analysis: weave these events into a causal story that unfolds over time. In this case, time is ordinal and events are treated as instantaneous (indicated by 'upon'), but in general, sentences may contain kinematic information and finer temporal distinctions.

The Big Mechanism consortia are adapting prior information extraction and machine reading methods to deal with cancer biology texts. In most cases, the approach is to create an intermediate formal representation (e.g., [2, 5, 15, 33, 45, 54, 60]). Much thought has been devoted to these Janus-like representations, which face both natural language text and elements of biological models in formal frameworks such as BEL, BioPAX, Kappa, PySB and Pathway Logic. One requirement of these representations is to store information that requires further processing to incorporate into models; particularly, imprecise and incomplete information about biological entities and processes at different scales. For example, the sentence 'c-Raf expression is required for the onset of non-small-cell lung carcinoma (NSCLC)' does not disclose the mechanism of c-Raf expression, nor the role of c-Raf in NSCLC, nor precisely what's meant by onset; and c-Raf expression obviously works at a different temporal and spatial scale than NSCLC. It would be challenging for trained humans to incorporate the meaning of this sentence into a formal biological model, and even more challenging for machines to do it, automatically. And yet, the sentence makes an important causal assertion that should not be discarded. Intermediate representations serve as a sort of 'purgatory' where the meanings of sentences await introduction into formal models.

Some examples of intermediate representations can be seen in table 1 and figure 3. They are syntactically diverse, but all are hierarchical and composed of predicate-argument assertions, where the predicate is an event such as binding or a modifier such as a concentration and the arguments are generally biological entities.
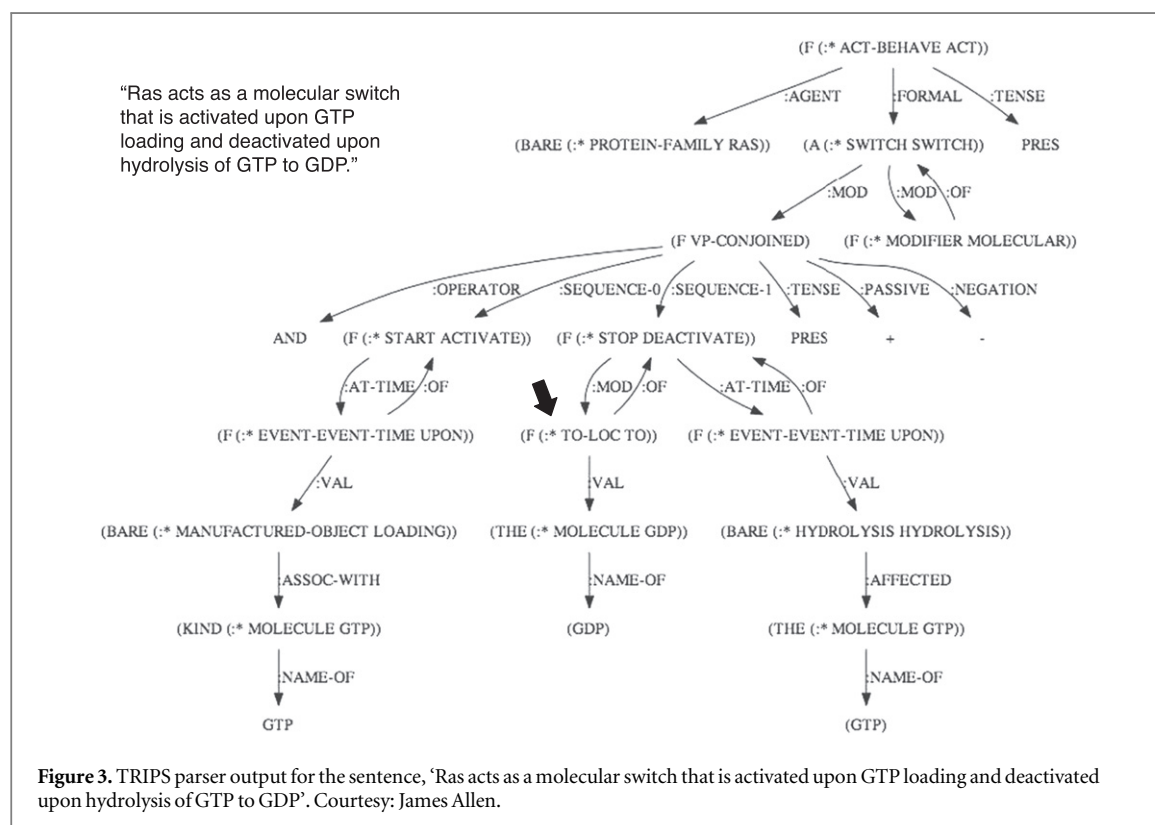
A semantically deep intermediate representation called logical form is produced by James Allen's TRIPS parser [2]. Figure 3 shows the logical form representation produced when TRIPS parses the sentence, 'Ras acts as a molecular switch that is activated upon GTP loading and deactivated upon hydrolysis of GTP to GDP'. TRIPS gets much of the semantics of the sentence right, and its subtle errors illustrate why semantic analysis is difficult. It correctly understands the sentence to be expressing something about how Ras acts or behaves. The node at the top of the tree in figure 3 tells us that the sentence is about the activity or behavior of an agent, Ras, and that the behavior is of a formal type called switch. The rest of the tree in figure 3 expands on, or modifies, this idea of the switch, by noting that it is a molecular switch, and by representing a sequence of events that the switch switches. One of these events is a start event, an activation; the other is a stop event, a deactivation. Events have associated times, and TRIPS associates the first instance of the word 'upon' with the time at which the activation event begins. 'Upon' what? Upon the manufacture of something—in this case, GTP—by loading. Similarly, the deactivation event happens upon hydrolysis of GTP.

However, TRIPS interprets the prepositional phrase in 'to GDP' incorrectly as identifying a location (highlighted by the heavy arrow in figure 3), as if the phrase were, 'hydrolysis of GTP goes to a place called GDP'. This illustrates the profound ambiguity of prepositions. TRIPS knows the correct sense of 'to GTP,' but doesn't choose it, here. Also, although TRIPS knows that the GTP in 'GTP loading' is the same as the one in 'hydrolysis of GDP to GTP,' it doesn't know

Table 1. Some intermediate representations of the meanings of sentences. Courtesy: Andrey Rzhetsky, Mihai Surdeanu, Daniel Marcu.

'Nicotine induces the binding of Raf-1 to Rb in HAEC and A549 cells as seen by IP/Western blotting'.	Statement1 = bind(entity1=Raf1, entity2=Rb) Statement2 = induce(entity1=Nicotine, process1 = statement1)
'Rb binds to E2F' 'Transactivation capacity of E2F' 'Rb ... inhibits the transactivation capacity of E2F'.	ComplexAssembly(In(Rb, E2F), Out(Complex(Rb, E2F))) TemplateRegulation(In(E2F)) DownRegulation(Controller(Rb), Controlled(TemplateRegulation(In(E2F))))
'MEK2/ERK2 catalyzes its own phosphorylation'.	(c / catalyze-01 :ARG0 (m / macro-molecular-complex :part (e / enzyme :name (n / name :op1 'MEK2')) :part (e2 / enzyme :name (n2 / name :op1 'ERK2'))) :ARG1 (p / phosphorylate-01 :ARG1 m))

Figure 3. TRIPS parser output for the sentence, 'Ras acts as a molecular switch that is activated upon GTP loading and deactivated upon hydrolysis of GTP to GDP'. Courtesy: James Allen.

that GTP loading and hydrolysis are essentially inverse processes, and that the switch controls not merely two events but alternating, inverse processes. This illustrates that automated, deep semantic analysis is limited by deep domain knowledge: an ordinary human who lacks this knowledge is unlikely to do any better than TRIPS at constructing a logical form representation of the sentence.

The logical form in figure 3 is an intermediate representation in the sense that it is neither text nor mechanistic model, but is a formal representation of the meaning of text from which elements of mechanistic models can be extracted.

Six months after it began, the BMP conducted an evaluation of shallow and deep reading[2]. Shallow

reading refers to named entity recognition and event recognition. The evaluation focused on recognizing genes and proteins, protein families, complexes, drugs, drug classes, cell lines, pathways and subcellular components; as well as protein post-translational modifications, binding and dissociation, direct and indirect regulation and modulation, translocation, and changes in quantities. Previous evaluations had shown that machines can recognize these entities and events with middling accuracy (e.g., [8, 9, 36, 41, 42, 56]). The best BMP systems performed at par, which is not surprising because in most cases they simply adapted technologies from these earlier evaluations. The consortia will try to improve on these technologies by exploiting the context provided by prior models. For example, if one has a prior model of molecular switches that includes 'slots' for activating and de-activating events, and the model specifies

that the same entity can participate in both activating and de-activating the switch, then a likely interpretation of the two mentions of GTP is that they are co-referring: GTP is one entity in both the activating and de-activating events.

While the evaluation of shallow reading involved standard test items and metrics, the evaluation of deep reading required new methods, befitting the BMP emphasis on deep language semantics and mechanistic models. The deep reading question was, 'given a prior model and some text, what does the text tell you about the model?' Four kinds of relationships between texts and prior models were probed: the text might corroborate what's already in the model; it might contradict something in the model; it might introduce a new mechanism (e.g., x phosphorylates y at location z); or it might introduce a new relationship between entities in the model (e.g., x activates y, but no mechanism is specified). One prior model was constructed and distributed to the consortia in BEL, Bio-Pax, Kappa and Pathway Logic formats. Six passages were also distributed, and the automated reading systems were asked to identify the assertions in these passages and say for each assertion which of the four kinds of relationship held between the assertion and the prior model, or whether none held. Before the evaluation began, biology experts on the evaluation team prepared a gold standard: a list of assertions, each of which was given one of five labels (one for each of the four relationships plus 'none holds'). After the reading systems submitted their results, the evaluation team calculated recall and precision scores. Recall is the fraction of relationships that should have been found that were actually found. For example, if eight 'new mechanism' relationships were found but sixteen were mentioned in the texts, then recall would be 0.5. Precision is the fraction of the relationships found by a reading system that were in the gold standard; for example, if the system found six 'new mechanism' relationships but only four of these were in the gold standard, then precision would be 0.67.

Two expert human biologists took this test. Remarkably, their recall scores were less than 0.5, which means they failed to notice roughly half of the relationships between the texts and the prior models. However, their precision was very high: 86–100%. Moreover, they noticed *different* relationships, so they disagreed with each other. They also noticed some relationships that the evaluation team had not. Some of these were added to the gold standard before the reading systems were evaluated. The best recall score for a reading system was 0.4 with an associated precision score of 0.67; that is, this system found 40% of the relationships in the gold standard, and 67% of the relationships it found were in the gold standard. The least effective system achieved 3% recall at 33% precision.

Although this evaluation was small and the grading was lenient, the fact that a machine was able to identify almost as many relationships between texts

and a prior model as the two expert biologists—albeit with lower precision—is promising, though perhaps unsurprising: human expertise probably includes an ability to *not* notice assertions that are 'obvious' or 'unimportant'. Still, if the result holds up in larger, more stringent evaluations planned for the summer of 2015, it will strengthen the case that machines are perhaps the only intelligences that hold out the promise of reading huge literatures broadly, quickly and thoroughly.

Reading to identify relationships between texts and prior models is the first step in the process of revising models, also called assembly (figure 2). At present, no Big Mechanism system does assembly on a large scale, although most participants in the BMP have developed technology to extract model fragments from intermediate representations; and organize large numbers of logical or probabilistic assertions into theories, or consistent sets, or most-probable groupings.

The greatest challenge posed by assembly will be managing the proliferation of models and meta-data that stand in various relations to each other (the model manager box in figure 2). It is unrealistic to think that there is 'one true model' of anything and that the job of Big Mechanism systems is to discover it. In fact, one advantage of machines being able to build models is to have them compare and contrast many models. There might be one consensus model at some level of abstraction (e.g., the interaction map of Ras signaling in [67]) which most researchers agree upon, while acknowledging a penumbra of variant and special-case models; but even a consensus model would not spring into being fully formed like Athena from the brow of Zeus, but would be developed in stages, by consulting multiple publications, and would be revised in light of new evidence. How are multiple drafts of models maintained? How shall we organize the evidence for these models? How shall we recognize the overlapping or contradicting parts of models [21, 30, 40, 59]? How shall we organize models of 'one thing'—say, Ras signaling—that are actually families of models specific to organisms, tissue types, and mutations? The complexity of model management is compounded by the need for qualitatively different kinds of models for different kinds of reasoning. As noted, interaction maps might serve to summarize molecular interactions, while physicochemical models—more or less coarse-grained—are required to simulate the kinetics of these interactions.

Abstraction, modularity and re-use of models will probably facilitate model management [46], although to the best of our knowledge, model management by machines is a new area of science.

Any solution to the model-management problem should depend on the anticipated uses of models (see the box labeled 'reasoning' in figure 2). The big mechanism consortia are focusing on several use cases, of which the most promising are curation and predicting and testing the effects of interventions.

Curation refers to a growing movement to make biological knowledge machine-readable [11, 12, 27]. Even if Big Mechanism technologies do not automate curation, they could reduce the human effort expended on mundane aspects of the job [28, 54, 61]. And humans could help solve problems that the machines may not be able to solve alone, such as figuring out which assertions in papers are relevant to a user's curation task [56], or detecting subtle elisions or obfuscations that are intended to make results seem more general or convincing. Machines might have trouble understanding things that are not said or are said artfully, and might do well to rely on humans to uncover rhetorical subtleties. One particularly attractive aspect of human-computer curation is that each chunk of work provides training data for machine learning, hastening the day when machines can curate all by themselves.

The second major use case for the BMP is testing interventions. Several BMP teams are focusing on predicting effects of interventions and testing these predictions in simulations that run on conventional computers as well as on field programmable gate array platforms [50]. The BMP will also be able to test predictions *in vitro*, using a fully-automated robotic system that can set up experiments to test hypotheses generated by Big Mechanism systems [38, 49].

3. Big Mechanism and the practice of science

The methods of experimental science give us knowledge of two kinds: they establish causality among tiny numbers of factors, and they quantify the influence of other factors as variance. But of course, this is an idealization. *p* values generally are pointless, few people report effect sizes, and experiments are done in dramatically different contexts. All of which means that the literature comprises thousands of *narrow* causal assertions of varying quality. We try to figure out complicated systems assertion by assertion. No wonder replicability can be a problem.

However, the purpose of the BMP is not to overturn the experimental methods that have given us great scientific insights. In fact, the BMP assumes that, for now, humans are the best practitioners of experimental methods, and takes researchers at their word until there is a reason not to. This is one reason why the BMP begins with research papers and tries to assemble models that integrate many causal assertions, *collectively*. This approach addresses implicitly the vulnerability of taking authors at their word. Often, the assertions of authors matter less than how these assertions stand in relation to others. Sometimes, these assertions are inconsistent or mutually contradictory [21, 40]. The better we can assemble causal assertions from many sources into large models, the better we will be able to evaluate new assertions. Of course, 'we' includes computers, as only computers can evaluate

how a causal assertion stands in relation to thousands of others derived from an enormous literature.

With the exception of some work on algorithms for inferring causal structure from data, and the possibility that data from BMP-proposed experiments will feed back into modeling, the BMP does nothing with primary data. Its 'data' are the assertions made by authors. There is plenty of work on inducing biological models directly from data (e.g., [6, 24, 51, 57, 68, 73]), and clearly the bewildering diversity of oncogenic mutations recommends big data approaches to therapy [71]. But the BMP is not a big data program, a cancer therapy program, or even a cancer biology program. It is a program designed to develop technologies to help scientists understand very complicated systems, generally.

The biggest contribution of Big Mechanism might be to change how knowledge is organized and communicated. We still work in a medieval tradition in which each of us searches the literature for very large numbers of narrow results and pulls them into our heads for synthesis into big mechanisms or causal understandings. This 'pull scholarship' is enormously wasteful. It works less and less well because there are simply too many results to synthesize in our heads. We can't read any faster than we already do, but the literature grows at an increasing rate, so we read more narrowly. Just when we need to understand highly connected systems *as systems* our research methods force us to focus on little parts. If the BMP works, it will be the first demonstration of a new kind of 'push scholarship' where, instead of pulling results into our heads, we push results into machine-maintained big mechanisms, where they can be examined by anyone. This could change science profoundly.

References

[1] Aldridge B B, Burke J M, Lauffenburger D A and Sorger P K 2006 Physicochemical modelling of cell signaling pathways *Nat. Cell Biol.* **8** 1195–203

[2] Allen J *et al* 2008 Deep semantic analysis of text *Symp. on Semantics in Systems for Text Processing (STEP) (Venice, Italy)* (http://www.aclweb.org/anthology/W08-2227.pdf)

[3] Bachman J and Sorger P 2013 New approaches to modeling complex biochemistry *Mol. Syst. Biol.* **9** 646 (http://www.nature.com/nmeth/journal/v8/n2/pdf/nmeth0211-130.pdf)

[4] Bada M *et al* 2012 Concept annotation in the CRAFT corpus *BMC Bioinformatics* **13** 161

[5] Banarescu L, Bonial C, Cai S, Georgescu M, Griffitt K, Hermjakob U, Knight K, Koehn P, Palmer M and Schneider N 2013 Abstract meaning representation for sembanking *Proc. ACL Linguistic Annotation Workshop* (http://amr.isi.edu/a.pdf)

[6] Bebek G and Yang J 2007 PathFinder: mining signal transduction pathway segments from protein–protein interaction networks *BMC Bioinformatics* **8** 335

[7] Biological Expression Language V1.0 (http://wiki.openbel.org/download/attachments/819563/BEL)

[8] Arighi C N, Lu Z, Krallinger M, Cohen K B, Wilbur W J, Valencia A, Hirschman L and Wu C H 2011 Overview of the biocreative III workshop *BMC Bioinformatics* **12** (Suppl 8)

[9] http://2013.bionlp-st.org/tasks/pathway-curation

[10] Biological Pathways Exchange Language (http://biopax.org/release/biopax-level3-documentation.pdf)

[11] Bourne P E and McEntyre J 2006 Biocurators: contributors to the world of science *PLoS Comput. Biol.* **2** e142

[12] Burge S, Attwood T K, Bateman A, Berardini T Z, Cherry M, O'Donovan C, Xenarios l and Gaudet P 2012 Biocurators and Biocuration: surveying the 21st century challenges *Database* **2012** bar059

[13] Cerami E G, Gross B E, Demir E, Rodchenkov I, Babur O and Anwar N 2011 Pathway Commons, a web resource for biological pathway data *Nucleic Acids Res.* **39** Database issue D685–90

[14] Croft D *et al* 2011 Reactome: a database of reactions, pathways and biological processes *Nucleic Acids Res.* **39** D691–7

[15] http://cyc.com/cyc/cycl/syntax

[16] Danos V and Laneve C 2004 Formal molecular biology *Theor. Comput. Sci.* **325** 69–110

[17] Danos V, Feret J, Fontana W, Harmer R and Krivine J 2007 Rule-based modeling of cellular signaling concurrency theory *Lecture Notes Comput. Sci.* **4703** 17–41

[18] Danos V, Feret J, Fontana W, Harmer R and Krivine J 2009 Rule-based modelling and model perturbation *Lecture Notes Comput. Sci.* **5750** 116–37

[19] Demir E *et al* 2010 The BioPAX community standard for pathway data sharing *Nat. Biotechnol.* **28** 935–42

[20] de Waard A and Maat H P 2012 Epistemic modality and knowledge attribution in scientific discourse: a taxonomy of types and overview of features *Proc. Workshop on Detecting Structure in Scholarly Discourse* pp 47–55

[21] Divoli A, Mendonça E A, Evans J A and Rzhetsky A 2011 Conflicting biomedical assumptions for mathematical modeling: the case of cancer metastasis *PLoS Comput. Biol.* **7(10)** e1002132

[22] Faeder J R, Blinov M L and Hlavacek W S 2009 Rule-based modeling of biochemical systems with BioNetGen *Methods Mol. Biol.* **500** 113–67

[23] Feret J, Danos V, Krivine J, Harmer R and Fontana W 2009 Internal coarse-graining of molecular systems *Proc. Natl Acad. Sci. USA* **106** 6453–8

[24] Friedman N 2004 Inferring cellular networks using probabilistic graphical models *Science* **303** 799–805

[25] Goentoro L, Shoval O, Kirschner M W and Alon U 2009 The incoherent feedforward loop can provide fold-change detection in gene regulation *Mol. Cell* **36** 894–9

[26] Gutenkunst R N, Waterfall J J, Casey F P, Brown K S, Myers C R and Sethna J P 2007 Universally sloppy parameter sensitivities in systems biology models *PLoS Comput. Biol.* **3** 1871–8

[27] Hirschman J, Berardini T Z, Drabkin H J and Howe D 2010 A MOD(ern) perspective on literature curation *Mol. Genet. Genomics* **283** 415–25 published online 11 March 2010

[28] Hirschman L *et al* 2012 Text mining for the biocuration workflow *Database* **2012** bas020

[29] Hlavacek W S, Faeder J R, Blinov M L, Posner R G, Hucka M and Fontana W 2006 Rules for modeling signal-transduction systems *Sci. STKE* **2006** re6

[30] Hlavacek W S 2009 How to deal with large models *Mol. Syst. Biol.* **5** 240 (published online 20 January 2009)

[31] Holderfield M, Deuker M M, McCormick F and McMahon M 2014 Targeting RAF kinases for cancer therapy: BRAF-mutated melanoma and beyond *Nat. Rev. Cancer* **14** 455–67 (published online 24 June 2014)

[32] Hunter L and Cohen K B 2006 Biomedical language processing: perspective what's beyond pubmed? *Mol. Cell* **21** 589–94

[33] Hunter L, Lu Z, Firby J Jr, Johnson W B, Ogren H, P and Cohen K 2008 OpenDMAP: An open source, ontology-driven concept analysis engine with applications to capturing knowledge regarding protein transport, protein interactions and cell-type-specific gene expression *BMC Bioinformatics* **9** 78

[34] Ideker T and Lauffenburger D 2003 Building with a scaffold: emerging strategies for high- to low-level cellular modeling *Trends Biotechnol.* **21** 255–62

[35] Karnoub A E and Weinberg R A 2008 Ras oncogenes: split personalities *Nat. Rev. Mol. Cell Biol.* **9** 517–31

[36] Kim J-D, Pyysalo S, Ohta T, Bossy R, Nguyen N and Tsujii J 2011 Overview of BioNLP Shared Task 2011 *Proc. BioNLP Shared Task 2011 Workshop (Association for Computational Linguistics)* pp 1–6 (http://aclweb.org/anthology/W/W11/W11-18.pdf)

[37] Kim S M and Hovy E H 2006 Identifying and analyzing judgment opinions *Proc. Human Language Technology/North American Association of Computational Linguistics Conf. (HLT-NAACL 2006)* (http://www.isi.edu/natural-language/people/hovy/papers/06HLT-JudgementOpinion.pdf)

[38] King R D *et al* 2009 The automation of science *Science* **324** 85–9

[39] Kitano H, Funahashi A, Matsuoka Y and Oda K 2005 Using process diagrams for the graphical representation of biological networks *Nat. Biotechnol.* **23** 961–6

[40] Kirouac D C, Saez-Rodriguez J, Swantek J, Burke J, Lauffenburger D A and Sorger P K 2012 Creating and analyzing pathway and protein interaction compendia for modelling signal transduction networks *BMC Syst. Biol.* **6** 29

[41] Krallinger M, Valencia A and Hirschman L 2008 Linking genes to literature: text mining, information extraction, and retrieval applications for biology *Genome Biol.* **9** (Suppl 2)

[42] Krallinger M, Vazquez M, Leitner F, Salgado D, Chatraryamontri A, Winter A and Valencia A 2011 The protein–protein interaction tasks of biocreative: III. Classification/ranking of articles and linking bio-ontology concepts to full text *BMC Bioinformatics* **12** (Suppl 8)

[43] Novere L *et al* 2005 Minimum information requested in the annotation of biochemical models (MIRIAM) *Nat. Biotechnol.* **23** 1509–15

[44] Le Novere N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A and Kitano H 2009 The systems biology graphical notation *Nat. Biotechnol.* **27** 735–41

[45] Lincoln P D and Talcott C 2010 Symbolic systems biology and pathway logic *Symbolic Systems Biology* ed S Iyengar (Sudbury, MA: Jones and Bartlett)

[46] Lopez C F, Muhlich J L, Bachman J A and Sorger P K 2013 Programming biological models in Python using PySB *Mol. Syst. Biol.* **9** 646 published online 19 February 2013

[47] Mallavarapu A, Thomson M, Ullian B and Gunawardena J 2009 Programming with models: modularity and abstraction provide powerful capabilities for systems biology *J. R. Soc. Interface* **6** 257–70

[48] (http://obofoundry.org)

[49] Yihui Liu, Martin K, Sparkes A and King R D 2010 The analysis of yeast cell morphology using a robot scientist *Int. Conf. on Computational Intelligence and Security (CIS) (11–14 December 2010)* pp 10–14

[50] Miskov-Zivanov N, Bresticker A, Krishnaswamy D, Venkatakrishnan S, Marculescu D and Faeder J R 2011 Emulation of biological networks in reconfigurable hardware *Proc. 2nd ACM Conf. on Bioinformatics, Computational Biology and Biomedicine* pp 536–40

[51] Mitra K, Carvunis A-R, Ramesh S K and Ideker T 2013 Integrative approaches for finding modular structure in biological networks *Nat. Rev. Genetics* **14** 719–32

[52] Movshovitz-Attias D and Cohen W W 2012 Bootstrapping biomedical ontologies for scientific text using NELL *Proc. 2012 Workshop on Biomedical Natural Language Processing (Association for Computational Linguistics)* (http://www.aclweb.org/anthology/W12-2402)

[53] Smith B *et al* 2007 The OBO foundry: coordinated evolution of ontologies to support biomedical data integration *Nat. Biotechnol.* **25** 1251–5

[54] Rak R, Rowley A, Black W and Ananiadou S 2012 Argo: an integrative, interactive, text mining-based workbench supporting curation *Database* **2012** bas010

[55] Rak R, Batista-Navarro R T, Carter J, Rowley A and Ananiadou S 2014 Processing biological literature with customizable web services supporting interoperable formats *Database* **2014** bau064 published online 8 July 2014

[56] Ravikumar K E, Wagholikar K B and Liu H 2014 Challenges in adapting text mining for full text articles to assist pathway curation *Proc. 5th ACM Conf. on Bioinformatics, Computational Biology, and Health Informatics (BCB '14)* (New York: ACM) pp 551–8

[57] Rhodes D R, Tomlins S A, Varambally S, Mahavisno V, Barrette T, Kalyana-Sundaram S, Ghosh D, Pandey A and Chinnaiyan A M 2005 Probabilistic model of the human protein–protein interaction network *Nat. Biotechnol.* **23** 951–9

[58] Russ T A, Ramakrishnan C, Hovy E H, Bota M and Burns G A P C 2011 Knowledge engineering tools for reasoning with scientific observations and interpretations: a neural connectivity use case *BMC Bioinformatics* **12** 351

[59] Rzhetsky A, Zheng T and Weinreb C 2006 Self-correcting maps of molecular pathways *PLoS One* **1** e61

[60] Rzhetsky A, Koike T, Kalachikov S, Gomez S M, Krauthammer M, Kaplan S H, Kra P, Russo J J and Friedman C 2000 A knowledge model for analysis and simulation of regulatory networks *Bioinformatics* **16** 1120–8

[61] Rodriguez-Esteban R, Iossifov I and Rzhetsky A 2006 Imitating manual curation of text-mined facts in biomedicine *PLoS Comput. Biol.* **2** e118

[62] Samaga R and Klamt S 2013 Modeling approaches for qualitative and semi-quantitative analysis of cellular signaling networks *Cell Commun. Signaling* **11** 43

[63] Sandor A and de Waard A 2012 Identifying claimed knowledge updates in biomedical research articles *Proc. Workshop on Detecting Structure in Scholarly Discourse* pp 47–55 (http://dl.acm.org/citation.cfm?id=2391174)

[64] http://sbgn.org/Documents/Specifications

[65] Shannon P, Markiel A, Ozier O, Baliga N S, Wang J T, Ramage D, Amin N, Schwikowski B and Ideker T 2003 Cytoscape: a software environment for integrated models of biomolecular interaction networks *Genome Res.* **2003** 2498–504

[66] Sneddon M W, Faeder J R and Emonet T 2011 Efficient modeling, simulation and coarse-graining of biological complexity with NFsim *Nat. Methods* **8** 177–83

[67] Stephen A G, Esposito D, Bagni R K and McCormick F 2014 Dragging ras back in the ring *Cancer Cell* **25** 272–81

[68] Stuart J M, Segal E, Koller D and Kim S K 2003 A gene-coexpression network for global discovery of conserved genetic modules *Science* **302** 249–55

[69] Talcott C, Eker S, Knapp M, Lincoln P and Laderoute K 2004 Pathway logic modeling of protein functional domains in signal transduction *Proc. Pacific Symp. on Biocomputing* (http://pl.csl.sri.com/Papers/04psb-pl.pdf)

[70] Talcott C 2006 Formal executable models of cell signaling primitives *2nd Int. Symp. on Leveraging Applications of Formal Methods, Verification and Validation ISOLA06* (http://pl.csl.sri.com/Papers/06isola.pdf)

[71] Tenenbaum J M and Shrager J 2011 Cancer: a computational disease that AI can cure *AI Mag.* **32** 14–26 (http://www.sciencemag.org/content/302/5643/249.long)

[72] Tónsing C, Timmer J, Kreutz C and Christian T 2014 Cause and cure of sloppiness in ordinary differential equation models *Phys. Rev.* E **90** 023303

[73] Wu G, Feng X and Stein L 2010 A human functional protein interaction network and its application to cancer data analysis *Genome Biol.* **11** R53