Incorporating Biological Knowledge into Evaluation of Causal Regulatory Hypotheses

L. Chrisman, P. Langley, S. Bay

Pacific Symposium on Biocomputing 8:128-139(2003)

INCORPORATING BIOLOGICAL KNOWLEDGE INTO EVALUATION OF CAUSAL REGULATORY HYPOTHESES

LONNIE CHRISMAN,[a†] PAT LANGLEY,[†] STEPHEN BAY,[†]
and ANDREW POHORILLE[‡]
[†] *Institute for the Study of Learning and Expertise*
2164 Staunton Court, Palo Alto, CA 94306
[‡] *Center for Computational Astrobiology and Fundamental Biology*
NASA Ames Research Center, M/S 239-4, Moffett Field, CA 94035

Biological data can be scarce and costly to obtain. The small number of samples available typically limits statistical power and makes reliable inference of causal relations extremely difficult. However, we argue that statistical power can be increased substantially by incorporating prior knowledge and data from diverse sources. We present a Bayesian framework that combines information from different sources and we show empirically that this lets one make correct causal inferences with small sample sizes that otherwise would be impossible.

1 Introduction and Motivation

There is a growing interest in the development and application of new computational methodologies for analyzing genomic and proteomic data, ranging from clustering techniques[1] to algorithms for inferring regulatory networks.[2,3,4,5,6] However, most such methods concentrate on discovering regularities in individual data sets and operate in a knowledge-lean manner. This contrasts sharply with the strategies of most biologists, who focus on testing specific hypotheses formulated in the context of biological knowledge and previous studies.

This observation suggests that biologists would benefit from better computational aids for hypothesis evaluation. Many such tools already exist, but their statistical power remains generally weak because, like most computational discovery techniques, they focus on data collected from a single study and typically ignore available knowledge. In this paper, we demonstrate how one can utilize prior biological knowledge to substantially increase the statistical power of causal hypothesis evaluation. Along the way, we address a number of challenges that this idea raises, including the facts that knowledge may come from different sources under different experimental conditions, have varying levels of uncertainty, and involve quantities that are not measured directly.

In the section that follows, we provide a motivating example that describes a biological hypothesis and relevant background knowledge. We then present a computational framework and associated algorithm that lets us calculate the evidence in favor of such a hypothesis given both prior knowledge and data.

[a] Corresponding author: lonnie@apres.stanford.edu.

We take a Bayesian approach to hypothesis evaluation, since this paradigm provides ready mechanisms for combining data and knowledge from multiple sources. After this, we report experimental studies with the algorithm on synthetic data, to determine its robustness, and a specific biological hypothesis, to ensure its relevance. In closing, we review related work on causal models in biology and suggest some directions for future research in this area.

2 A Motivating Example

Mitogen-activated protein kinase signal transduction pathways process a wide range of extracellular stimuli to determine a cell's transcriptional response to environmental changes or inter-cellular messages. One example, the c-Jun NH_2-terminal kinase[7] (JNK/SAPK) pathway, responds to growth factors (e.g., TGF-β and EGF), cytokines (e.g., TNF and IL-1), and forms of environmental stress (e.g., osmotic and radiation). It terminates in the phosphorylation and activation of JUN-family transcription factors, which dimerize with FOS, ATF, or other JUN factors to form AP-1 leucine-zipper transcription factor complexes,[8] which in turn enhance or repress transcription of many immediate-early genes. The JNK pathway has been implicated in many cellular processes and pathologies, including embryonic morphogenesis, cancer, immune system response, apoptotic signaling, cardiac hypertrophic response, neurodegenerative disease, and diabetes complications.[9] The JNK pathway is also considered a promising intervention point for many pathological conditions.

In humans and mice, the JNK family of kinases is derived from three genes, each of which elicits distinct responses under distinct conditions.[10,11] Jnk3 is found almost exclusively in brain, heart, and testes, whereas Jnk1 and Jnk2 are present in all tissues.[10] JNKs are known to phosphorylate several components of AP-1 complexes, including c-Jun, JunD, and Atf2,[9] although the different JNKs differ in their ability to phosphorylate each target.[12,13,10] The JUN family of transcription factors consists of c-Jun, JunB, JunD, and the viral oncogene v-Jun. They differ in their activation conditions, the AP-1 complexes in which they participate, and their transcriptional targets. To date, most laboratory studies involving JNK pathways have studied the involvement of JNK or JUN as a group, rather than looking at specific JNK or JUN variants. However, understanding the interactions between specific variants is essential to untangling functional roles of these pathways. We use as our motivating example the hypothesis that c-Jun is uniquely activated by Jnk2, and therefore is not activated by Jnk1.[b]

[b] For example, Kallunki et al.[13,11] found that Jnk2 binds to c-Jun 25 times more efficiently than Jnk1, and Gupta[10] found that Jnk2 isoforms tend to have higher affinities for c-Jun and Atf2 than do Jnk1 isoforms. Until recently, no kinase other than JNKs had been found

3 Representing Background Knowledge, Hypotheses, and Data

A computational system that evaluates causal biological hypotheses in the context of background knowledge and data must first represent such knowledge, data, and hypotheses. We encode these in terms of relations among discrete variables that can take on the values + (up-regulated), − (down-regulated), or 0 (unchanged) relative to a control condition. Facts and experimental data are divided into *scenarios*, with $X_{i,j}$ denoting the value of variable $\mathbf{x}_i$ in scenario j. When referring to a specific variable by name, we need only the scenario subscript, e.g., c-Jun$_j$ for the level of phosphorylated c-Jun in scenario j.

Consider the background knowledge that increasing TPA causes IL-11 expression to increase[16] [with other factors held constant]. To encode this, we assign this to a scenario, say $j = 1$, then define *causal conditions*, $C_1 = \{\text{TPA}_1 = +, \text{TNF-}\alpha_1 = 0\}$ and *known effects*, $E_1 = \{\text{IL-11}_1 = +\}$. Encoding of experimental data is done in a similar manner. Consider an experiment in which cells are exposed to an increased amount of TNF-α with TPA exposure unchanged, and in which expressions of nur77, FL1, and IL-11 increase, whereas expressions of p19 and p53 decrease. Assigning this experiment to scenario $j = 2$, we encode the *experimental conditions* or interventions as $I_2 = \{\text{TNF-}\alpha_2 = +, \text{TPA}_2 = 0\}$ and the *observed effects* as $D_2 = \{\text{EL\_nur77}_2 = +, \text{EL\_p19}_2 = -, \text{EL\_FL1}_2 = +, \text{EL\_IL-11}_2 = +, \text{EL\_p53}_2 = -\}$. If Jnk1 were also knocked out, the experimental conditions would be expressed as $I_3 = \{\text{TNF-}\alpha_3 = +, \text{TPA}_3 = 0, \text{Jnk1}_3 = 0\}$, where $\text{Jnk1}_3 = 0$ states that Jnk1 is controlled to be unchanged.

Background knowledge also indicates which direct causal links are plausible. We say that variable $\mathbf{x}_1$ is a causal *parent* of $\mathbf{x}_2$ when externally changing $\mathbf{x}_1$ can effect a change in $\mathbf{x}_2$ when all other variables are held unchanged. We specify *plausible parent relationships*, since there may be uncertainty over any specific causal link. In our example, we specify that each stimulus variable (TNF-α or TPA) is a plausible parent of every kinase (Jnk1 or Jnk2), that each kinase is a plausible parent of every transcription factor (JunD or c-Jun), and that each transcription factor is a plausible parent of each expression variable. The hypothesis that Jnk2 uniquely activates c-Jun is represented by the assertion that there is no link from Jnk1 to c-Jun, which may be true or false in any particular hypothetical causal model. Additional background knowledge takes the form of numeric assessments, α, that encode subjective beliefs about model parameters, including link likelihoods, reliability of causal statements, expected noise levels, and beliefs that particular links should be positive or negative influences.

---

capable of phosphorylating c-Jun;[12] but some evidence has emerged that an ERK kinase may phosphorylate c-Jun in specific cell types and developmental conditions.[14,15]

4 From Background Knowledge to Prior Probabilities

Background knowledge consists of plausible causal links between variables, causal conditions C, known effects E, and numeric assessments, α. Because we are working in a Bayesian framework, we must transform all this background knowledge into prior probabilities. Also, we must be able to express causal relations like "TNF-α (directly or indirectly) up-regulates JunD" and the results of experimental manipulations like applying TPA and knocking out Jnk1. Traditional notations for conditional probability do not capture the distinction between an estimate when a variable is observed and an estimate when a variable is manipulated. Thus, we introduce the notation $P(X|Y:Z)$ for the probability of X when Y is believed or observed to be true and Z is exogenously forced to be true. With our background knowledge consisting of α, E, and C, we define the *prior causal probability* as $P(\cdot|E, \alpha : C)$.

A causal model structure, M, is an acyclic subset of directed links between variables. Each variable $\mathbf{x}_i$ has an associated conditional probability, $P(\mathbf{x}_i|par_M(\mathbf{x}), M, \theta_M)$, which specifies its probability distribution conditioned on possible values of its parents, provided $\mathbf{x}_i$ is not exogenously controlled. The conditional probability is parameterized by a set of parameters, θ_M, and does not depend on the scenario, so that together an instance of (M, θ_M) defines a Bayesian network. For example, when M contains a link from Jnk1 to JunB, θ_M includes the probabilities that JunB is up-regulated, stays the same, or is down-regulated given that Jnk1 is up-regulated. When JunB has multiple parents according to M, our parameterization combines the contribution from each incoming link as a weighted mixture. We provide a detailed elucidation of our specific parameterization in a supplement.[17] We handle causal intervention by setting the local model to be $P(X_{i,j} = z|par_M(\mathbf{x}), M, \theta : X_{i,j} = z, \ldots) = 1$ whenever a variable is exogenously controlled, effectively severing the influence from the variable's parents.[18,3] Following the Bayesian network product rule and combining scenarios, the joint distribution over all variables under the causal interventions in C is the product of the conditional probabilities:

$$P(X|M, \theta_M : C) = \prod_i \prod_j P(X_{i,j}|X_{par_M(\mathbf{x}_i),j}, M, \theta_M : C_{i,j}) \qquad (1)$$

Knowing M and θ_M is sufficient for estimating probabilities in situations that involve hypothetical causal interventions, such as $P(\text{c-Jun} = +|\text{EL\_IL-11} = 0, M, \theta_M : \text{TNF-}\alpha = +)$. In terms of M and θ, we rewrite the prior causal probability as

$$P(X, M, \theta_M|E, \alpha : C) = P(M)P(\theta_M|\alpha)P(X|E, M, \theta_M : C)\ , \qquad (2)$$

where $P(M)$ expands to $k \cdot exp(-|M|)$ to provide a simplicity bias and $P(\theta_M|\alpha)$ consists of products of Dirichlet priors. The final term is encoded by the Bayesian network (M, θ_M) from Equation (1).

5 Evaluating Causal Biological Hypotheses

As we have noted, biologists are often concerned with evaluating whether experimental data, encoded by experimental conditions I and observed effects D, support some particular causal hypothesis, H. It is important to distinguish between the degree to which (a) the data alone support the hypothesis, (b) the data and prior knowledge together support the hypothesis, or (c) the data support the hypothesis in the context of the prior knowledge. Classical statistical hypothesis testing generally focuses on (a) and Bayesian analysis usually focuses on (b), whereas we focus on (c). To this end, we first combine the prior probability of a causal model M with experimental data to determine the posterior probability that H is true, as in (b), then we utilize this to define a p value that codifies (c). For example, the scientist may which to publish results demonstrating how significantly his new data supports (or refutes) a hypothesis in the context of what was previously known.

In our framework, a causal hypothesis is an expression that is either true or false as a function of the model M, its parameters θ_M, and the values of latent variables X.[c] The hypothesis used in our example, $(Jnk1, \text{c-Jun}) \notin M$, is a function only of M. If we let

$$\begin{aligned} prior(H) &= P(H|E, \alpha : C) \\ posterior(H|D) &= P(H|D, E, \alpha : I, C) \end{aligned}$$

the latter expression, the *posterior causal probability*, can be rewritten as

$$P(X, M, \theta_M|D, E, \alpha : I, C) = \xi P(M) P(\theta_M|\alpha) P(X, D, E|M, \theta_M : I, C) \quad (3)$$

$$\text{where} \qquad \xi = 1/\sum_{M \in \mathcal{M}} \int P(M) P(\theta_M|\alpha) P(D, E|M, \theta_M : I, C) d\theta_M$$

and where $\mathcal{M}$ is the set of possible causal models.

A posterior probability may be an optimal metric to employ in decision making contexts, but it often does not measure the real concerns in scientific data analysis. The posterior can be hard to interpret due to subjective assessments in prior knowledge and does not specifically reveal the data's support for the conclusions. A typical data analysis question is whether the experimental

[c] We also let H depend on a hypothetical control condition, which lets us evaluate the outcomes of hypothetical causal interventions, but we have omitted this case here for the sake of simplicity.

data support the hypothesis in the context of prior knowledge, which, as we have explained, differs from whether the data and prior knowledge together support the hypothesis. Many scientists are more familiar with the frequentist notion of a p value, which is conventionally not applicable to such knowledge-rich contexts. However, by combining the p value with the posterior, we obtain

$$p(H) = Pr\,[\, posterior(H|D') > posterior(H|D) \mid D' \sim prior(\cdot|\neg H)\,]$$

where $D' \sim prior(\cdot|\neg H)$ denotes that D' are hypothetical observations drawn at random from $prior(\cdot|\neg H)$. As in classical statistics, the p value gives the probability that random observations would appear to support H as much as the actual data even when H does not hold, i.e., the probability that one can be misled by the data into thinking H is true when it is actually false. A p value near zero indicates that the data provide strong support for H. If $p(\neg H) \approx 0$, then they provide strong evidence against the hypothesis.

We utilize a Metropolis-Hastings sampler[19] to compute $posterior(H|D)$, based directly on Equation 3. The basic strategy, which is an instance of a Markov chain Monte Carlo algorithm,[20] rapidly samples as many plausible models as it can in a short time. The method samples each model with a probability that asymptotically approaches its posterior probability, and uses the resulting counts to estimate how often the hypothesis H is true. We sample M, θ, and X simultaneously, as opposed to most other approaches to Bayesian network induction,[2,3] which usually use one method to search over possible model structures, M, a separate nested method (often based on EM) to fit parameters, θ, and yet different algorithms for inference over X.

Our algorithm begins with a starting model structure M, parameters θ, and value assignments to all latent variables, X. One point in the sample space, (M, θ, X), can be conceptualized as one possible model of the underlying biological system and how its latent variables respond in each situation. The method then proceeds to sample new values for $z_t = (M_t, \theta_t, X_t)$, based on z_{t-1}, to converge on the posterior distribution given by Equation 3.

At each step a new point, $z' = (X', M', \theta')$, is randomly proposed according to a proposal probability $g(\cdot|z_{t-1})$, and the Hastings acceptance probability

$$a(z_{t-1}, z') = \min\left\{1, \frac{\pi(z')}{\pi(z_{t-1})}\frac{g(z_{t-1}|z')}{g(z'|z_{t-1})}\right\}$$

is computed, where $\pi(z)$ is the right hand side of Eq. 3, and where z_t is set to z' with probability $a(z_{t-1}, z')$ and otherwise to z_{t-1}. The Hastings update rule guarantees *detailed balance*,[20] and thus that $P(z_t) \to \pi(z)$ as $t \to \infty$, provided that g can reach all points of the sample space.

We will not specify in detail our choice for g, but note that it involves a mixture of strategies, one for altering M by adding or deleting links, one for updating θ, one for changing X, and a few more specialized methods. When a new sample changes only a few links, parameters, or values, $\pi(z')$ can be updated incrementally, so that computational complexity depends on the number of changes rather than on the size of the model. Moreover, the normalization factors for π and g are not relevant and thus are not computed. Finally, we follow the standard practice for Markov chain Monte Carlo algorithms and begin with a 'burn-in' period before tallying statistics (usually 1/10th the total number of iterations), and we usually tally only every 100th sample in a 100,000 sample run. Our Java-based implementation typically explores about 2,000 proposals per second on a 1.5 GHz Pentium IV, with a typical proposal acceptance rate around 35 percent.

To compute the p value for a hypothesis H, we first utilize the above algorithm to draw a sample (M, θ_M, X), enforcing $\neg H$ during the process and treating all variables in D as latent (i.e., ignoring their original observed values). We then extract D' from the sampled X, use D' in place of D, and run the above procedure to compute $posterior(H|D')$. We repeat this N times, tallying the mean and variance of $posterior(H|D')$ across the different choices of D'. With small N, the number of times that $posterior(H|D') > posterior(H|D)$ is a poor estimate of the p value, so we instead assume $posterior(H|D')$ is normally distributed and use the area under the normal curve with the tallied mean and variance to estimate the probability that $posterior(H|D') > posterior(H|D)$.

6 Experimental Evaluation

Our basic claim is that incorporating prior knowledge into causal biological hypothesis evaluation enhances statistical power, therefore making it possible to infer causal effects that would otherwise be undetectable in small sets of experimental data. To support this claim, we utilized the hypothetical, but biologically plausible, model in Figure 1 to generate synthetic data under six different experimental conditions. These experiments involved one wild type organism and two knockout conditions, each under stimulation by TPA or TNF-α. The 'true' model takes the same form as those described earlier, that is, a causal Bayesian network in which each random variable has the domain $\{+, 0, -\}$ and each causal influence is stochastic (with 10% of the observed values being altered by noise). We generated two sets of data for our evaluation: the first assumed a model in which Jnk1 does not influence c-Jun, whereas the second assumed this influence does occur.

Our first study aimed to demonstrate that prior background knowledge lets one obtain statistical support for a causal hypothesis even when some of

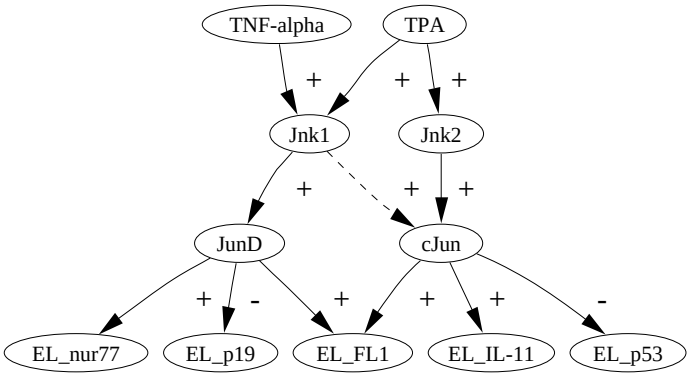

Figure 1: A hypothetical regulatory system used to generate synthetic data. The variables TNF-α and TPA are extracellular stimulus levels controlled by experimental conditions. Jnk1 and Jnk2 denote activation levels for the respective kinases, each of which are knocked out in two of six experiments. JunD and c-Jun, which denote activation levels of the respective transcription factors, are unobserved in the experiments. Variables prefaced with EL denote the observed expression levels of specific genes. Enhancement and suppression influences are indicated with pluses and minuses, respectively.

the relevant variables are latent. We used background knowledge about the JNK/JUN signaling pathways collated from published literature, subdivided into 12 facts: c-Jun is a transcription factor (i.e., parent) of IL-11[16] and p53;[21] JunD is a transcription factor of nur77;[22] TPA up-regulates IL-11,[16] JunD;[23] c-Jun;[24] and nur77;[22] JunD up-regulates nur77;[25] c-Jun up-regulates IL-11[16] and down-regulates p53;[21] and Jnk1 and Jnk2 are kinases, so they positively influence their targets. We used this knowledge in evaluating two hypotheses

H_1 : There is no direct causal link from Jnk1 to c-Jun.

H_2 : There is a direct causal link from Jnk1 to c-Jun.

against the two data sets. We performed each evaluation twice, once using the entire corpus of background knowledge, and once using no prior background knowledge (except for the set of variables and plausible links). Table 1 shows the posterior probabilities and p values that resulted from these runs.

Statistically significant support for the correct hypothesis occurs at the 0.05 level only for the two cases that utilize background knowledge. From the data alone, neither hypothesis is supported at a significant level. Since c-Jun and JunD are both latent, the knowledge is necessary to relate them to the observed data; otherwise, the system finds no support for the hypotheses that involve those variables.

Table 1: Statistical significance levels for causal hypotheses with and without the incorporation of background knowledge. The first number denotes the posterior probability of each hypothesis, whereas the second denotes the p value.

True Model	All Prior Knowledge		No Prior Knowledge	
	H_1 : no link	H_2 : link	H_1 : no link	H_2 : link
No Influence	0.99 / 0.003	0.01 / 0.77	0.42 / 0.46	0.58 / 0.51
Influence	0.03 / 0.500	0.97 / $< 10^{-7}$	0.44 / 0.55	0.56 / 0.44

Figure 2 shows the results of a more extensive study, using only data generated by the model in which Jnk1 does not influence c-Jun, that relates statistical power to the amount of background knowledge available. In these runs, we ordered the 12 facts randomly and evaluated the hypothesis H_1 (that the link is not present) while varying the number of facts given to the system from 0 to 12. The graphs reveal that a certain amount of background knowledge, about half the corpus, must be available before strong support for the hypothesis becomes evident. The sixth fact, which happened to be that c-Jun influences EL_IL-11, was a major clue the system needed to disambiguate between two main competing systems-level scenarios that appeared possible prior to this fact. With nine or more facts, the system detects that the data sets indeed support the hypothesis statistically at the 0.05 level.

Without this ample biological knowledge, it would have been impossible to validate or refute the hypotheses from experimental data alone. However, when combined with such prior information, the data reveal clear support for the hypothesis at statistically significant levels. These results are consistent with our central thesis — that utilizing preexisting background knowledge can be crucial for effective hypothesis testing when only small samples are available or the variables are only partially observable.

7 Discussion

Our approach to interpreting biological data contrasts sharply with most computational work in this area. The most common methodology passes available data to an induction algorithm, which extracts regularities that, hopefully, are biologically meaningful. However, work in this paradigm typically assumes data are plentiful and makes little use of background knowledge about biology. Our framework instead takes advantage of prior knowledge and previous experimental results in analyses of new observations, increasing their statistical power even when few samples are available.

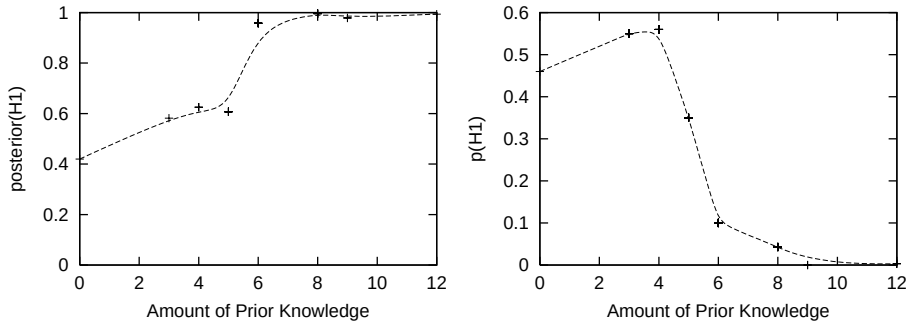

Figure 2: Evidence that statistical power increases with available background knowledge, as reflected by the number of biological facts provided to the system. The data set supports the (correct) hypothesis at a significant level ($p(H_1) < 0.05$) only when at least nine of the 12 domain facts are utilized during evaluation.

Despite this crucial difference, our approach has clear links to earlier research in computational biology. The strongest connection is to methods for learning Bayesian networks,[2,26,27,28] which also search a space for causal models that match the experimental data. For the special case in which each possible link is a hypothesis, we can view these systems as evaluating causal hypotheses with respect to their support by the data. Previous efforts within this framework have also dealt with the technical complications that arise with latent variables[29,27] and causal interventions;[3,18] both addressed in our own work.

Nevertheless, most research on inducing Bayesian networks of biological systems has taken a knowledge-lean, data-mining approach. Despite a few exceptions that incorporate knowledge about promoter sequences,[4,30,31] typical work in this paradigm attempts to construct a causal model from scratch, rather than evaluate particular causal relations in the context of background knowledge. Our own previous research on computational methods for revising causal biological models[5] comes much closer, but still emphasizes model discovery rather than hypothesis evaluation. A few researchers[32,33] have focused on model evaluation as opposed to model discovery. In particular, the JustAid system[32] also supports the use of experimental data to evaluate qualitative causal hypotheses, although its underlying algorithms are quite different and it addresses neuroendocrinology rather than gene regulation.

At a computational level, our approach draws heavily on Markov chain Monte Carlo techniques for hypothesis testing from the statistical literature.[34] However, our use of p values in a Bayesian context appears novel, in that the standard approach to Bayesian hypothesis testing involves comparing the *Bayes Factors*[34] or *Bayesian Information Criteria*[35] for alternative models.

Both approaches are legitimate from a statistical perspective, but we have chosen to utilize p values because they are generally more familiar to biologists.

Despite our encouraging results, we must extend our computational framework along a number of dimensions before it can become a useful tool for biologists. For example, we should explore other representations that let us encode qualitative causal knowledge about biological systems, especially notations that make stronger contact with established biological concepts like phosphorylation and dimerization. Moreover, we should incorporate this extended formalism into a user interface that lets biologists visualize and manage their background knowledge, hypotheses, and experimental data.

Another limitation of our current implementation concerns efficiency, in that its sampling strategy does not scale well to very large corpora of background knowledge. Also, since our posterior distributions often exhibit isolated regions of high probability, achieving a workable mixing rate from the Hastings algorithm is challenging. In future work, we plan to make our inference methods more efficient by incorporating additional ideas from the literature on Markov chain Monte Carlo and Metropolis-Hastings algorithms.[36,20] Our approach would also benefit from computational methods that generate plausible hypotheses automatically by reasoning over biological knowledge.[37] Finally, we must demonstrate the utility of our framework on data from actual biological experiments and on hypotheses they were designed to test.

In summary, we have presented a computational approach to evaluating causal hypotheses that takes advantage of background knowledge and previous experimental results to increase statistical power. Our framework encodes biological knowledge, hypotheses, and data in terms of qualitative relations between variables, and it utilizes Bayesian inference to calculate the evidence for and against each candidate hypothesis. We illustrated this approach to hypothesis evaluation in the context of knowledge and data about the JNK/JUN signaling pathways, and we demonstrated the increase in statistical power that background knowledge provides in this setting. We believe that the techniques we have described will make future tools for computational biology more robust and let them use available data more effectively.

8 Acknowledgements

We thank Jessica Ross, Jeff Shrager, and Hilary Spencer for their input and contributions to this project. This research was supported in part by Grants NCC 2-1202, NCC 2-5471, and NCC 2-1335 from NASA Ames Research Center and in part by NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation.

9 References

1. M. B. Eisen *et al.* *Proc.Natl.Acad.Sci.*, 95(25):14863–8, 1998.
2. N. Friedman *et al.* *Comp.Biol.*, 7:601–20, 2002.
3. D. Pe'er *et al.* *Bioinformatics*, pages 1–9, 2001.
4. A. J. Hartemink *et al.* In *Pacific Symp.Biocomp.*, pages 437–449, 2002.
5. J. Shrager *et al.* In *Pacific Symp.Biocomp.*, pages 486–97, 2002.
6. T. Akutsu *et al.* *Bioinformatics*, 16:727–34, 2000.
7. C. R. Weston *et al.* *Curr.Op.Gen.Dev.*, 12(1):14–21, 2002.
8. M. Z. L. Karin *et al.* *Cur.Op.Cell Biol.*, 9(2):240–6, 1997.
9. R. J. Davis. *Cell*, 103(2):239–52, 2000.
10. S. Gupta *et al.* *J. EMBO*, 15(11):2760–70, 1996.
11. F. Bost et al. *Mol.Cell Biol.*, 19(3):1938–49, 1999.
12. A. Minden *et al.* *Biochemica et Biophysica Acta*, 1333:F85–F104, 1997.
13. T. Kallunki *et al.* *Genes Dev.*, 8(24):2996–3007, 1994.
14. S. Leppa *et al.* *J. EMBO*, 17:4404–13, 1998.
15. Y. Levkovitz *et al.* *J.Neurosci.*, 22(10):3845–54, 2002.
16. W. Tang *et al.* *J.Biol.Chem.*, 273(10):5506–13, 1998.
17. http://cll.stanford.edu/∼lonnie/papers/psb2003sup.ps.
18. C. Yoo *et al.* In *Pacific Symp.Biocomp.*, pages 498–509, 2002.
19. W. K. Hastings. *Biometrica*, 57:97–109, 1970.
20. R. M. Neal. Tech.Rep. CRG-TR-93-1, Dept. C.S., Univ. Toronto, 1993.
21. M. Schreiber *et al.* *Genes Dev.*, 13:607–19, 1999.
22. C.Y. Young *et al.* *Oncol.Res.*, 6(4–5):203–10, 1994.
23. T. Li *et al.* *J.Biochem.*, 277:32668–76, 2002.
24. Zhou H. *et al.* *J.Biochem.*, 275(30):22868–75, 2000.
25. C. O. Stocco *et al.* *J.Biochem.*, 277(5):3293–3302, 2002.
26. D. Heckerman. Tech.Rep. MSR-TR-95-06, Microsoft Res., 1996.
27. P. Spirtes *et al.* *Causation, Prediction and Search.* MIT Press, 2001.
28. G. Cooper *et al.* *Machine Learning*, 9:309–47, 1992.
29. N. Friedman. In *UAI*, pages 129–38, 1998.
30. I. M. Ong *et al.* *Bioinformatics*, 18:241–8, 2002.
31. E. Segal *et al.* In *RECOMB*, 2002.
32. A. Mahidadia *et al.* In *Proc. Discovery Science*, pages 214–27, 2001.
33. A. J. Hartemink *et al.* In *Pacific Symp.Biocomp.*, pages 422–33, 2001.
34. A. Raftery. In *[36]*, pages 163–188. 1996.
35. H. Akaike. *Trans.Auto.Control*, 19:716–27, 1974.
36. W. R. Gilks *et al.* *Markov Chain Monte Carlo in Practice.* Chapman & Hall, 1996.
37. P. Romero *et al.* *Pacific Symp.Biocomp.*, pages 471–82, 2001.