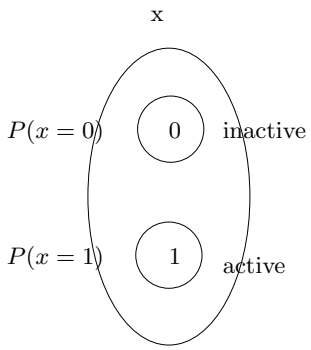
Models of transcriptional regulation

We have already discussed four simple mechanisms of transcriptional regulation,

- nuclear exclusion
- nuclear concentration
- modification of bound activator
- redirection of binding sites due co-regulator

Much of transcriptional regulation depends on protein-protein interactions that are modulated by protein modifications such as phosphorylation (we will only consider phosphorylation). A single protein may have multiple phosphorylation sites and its activity may depend combinatorially on the state of the individual sites. To represent a protein state we introduce variables

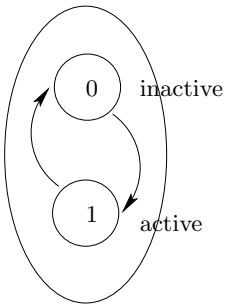


where the two states (0/1) are tied to protein activity (whether it is phosphorylated).

One way to build mechanisms is to use these variables as nodes in a Bayesian network. In many cases, however, Bayesian networks would not able to represent the mechanisms *explicitly* but would require appropriate setting of the parameters to characterize the behavior of the interacting variables. So while Bayesian networks could effectively capture the mechanism, they would not yield an appropriate visual representation of the mechanism (the parameters are not visible in the graph). Moreover, several mechanisms may be consistent with a single graph structure while differing substantially in terms of the choice of the parameter values. We focus here instead on a complementary way of capturing protein states and their dynamics using state transition diagrams.

Transition diagrams

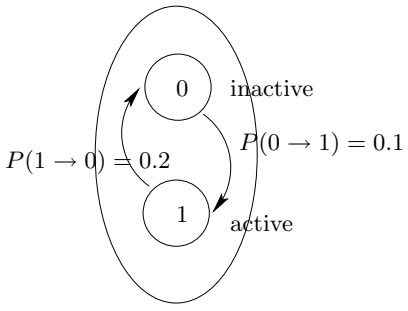
Protein states are often transient so our representation of state must involve time in some manner. We can, for example, explicate transitions from one state to another within some reference time interval:



where, in terms of phosphorylation, the transition from inactive to active state is due to kinase(s) while the reverse transition is facilitated by phosphatases. It is no longer sufficient to characterize the model in terms of static probabilities $P(x = 0)$ and $P(x = 1)$. Instead, we must consider the state $x(t)$ as a function of time and evaluate $P(x(t) = 0)$ and $P(x(t) = 1)$, i.e., the probabilities concerning the state at time t.

To fully specify the model we need to quantify when to expect the transitions. For example, we can specify the probability that the protein makes a transition to another state (or that it remains in the same state) within the reference time interval. The probabilities associated with making a transition and remaining in a state are obviously linked.

Here's an example of an annotated diagram:



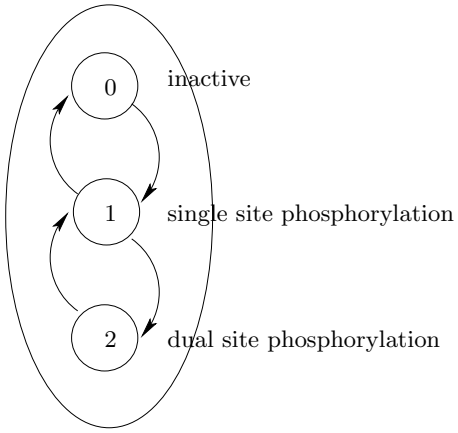
The remaining probabilities can be inferred: for example, the probability of remaining inactive within the reference period is $1 - 0.1 = 0.9$.

So far we have viewed this as a model for a single protein (interpretation 1). Alternatively, we could view this as a model for a set of proteins and interpret the current probability distribution (belief) over the states as a frequency in a population of proteins (interpretation 2). In other words, $P(x(t) = 0)$ represents the fraction of proteins of this type that are inactive at time t. The state transitions consequently represent how the sates of the population of protein evolve within the reference time period. For example:

$$P(x(t+1) = 0) = P(x(t) = 0)P(0 \to 0) + P(x(t) = 1)P(1 \to 0)$$

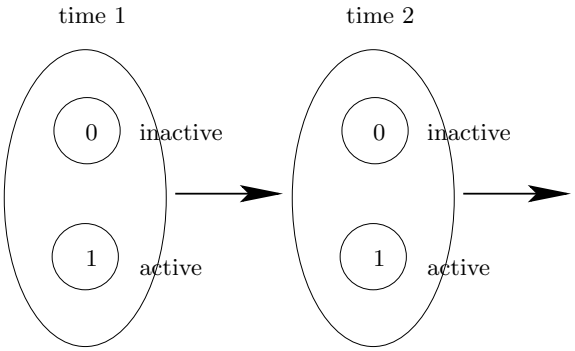
where t indexes the time interval and x denotes the state. While it may seem inconsequential at this stage to either interpret the distribution 1) as a belief over the state of a single protein or 2) as a fraction in a population, the choice of the interpretation does make a difference later on when we consider the coordinate influence of multiple proteins.

We can naturally extend this representation to multiple phosphorylation sites:



where the missing transition from inactive to dual phosphorylation indicates that the transfer of phosphate groups cannot occur in tandem but has to be done sequentially, perhaps with a different kinase.

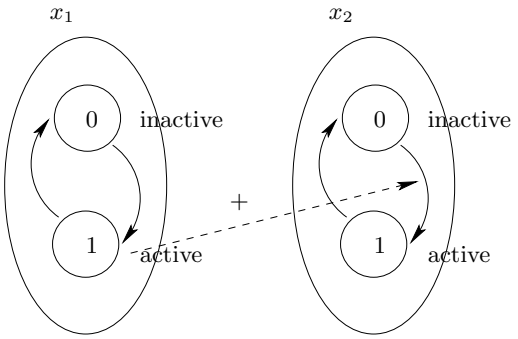
Under the single protein interpretation the large ovals enclose variables whose values vary in a time dependent manner. The structure within the ovals explicate how the values change. In order to capture the "dependences" in the model, we can cast the model as a (dynamic) Bayesian network:



where the ovals refer to protein states at successive time intervals and the thick arrow represents the fact that the state at the next time interval depends on the current state. The additional structure highlighted in the transition diagram is often helpful in our context (more explicit).

Interactions

To capture an interaction between two protein states, e.g., resulting from a phosphorylation reaction, we could write



where only an active kinase (left) can phosphorylate the substrate (right). In other words, the probability associated with the inactive-active transition of the substrate changes in the presence of an active kinase. We can also annotate the influence in terms of its effect on the transition probability (positive or negative) as in the figure.

More formally, the effect of the kinase on the transition probability depends on the interpretation:

(1) Under the single protein interpretation the transition probability from inactive to

active state depends on the state of the kinase or $x_1(t)$. For example,

$$P_2(0 \to 1|x_1(t)) = \delta_{x_1(t),1} P_2(0 \to 1)$$

so that the transition is possible only in the presence of an active kinase.

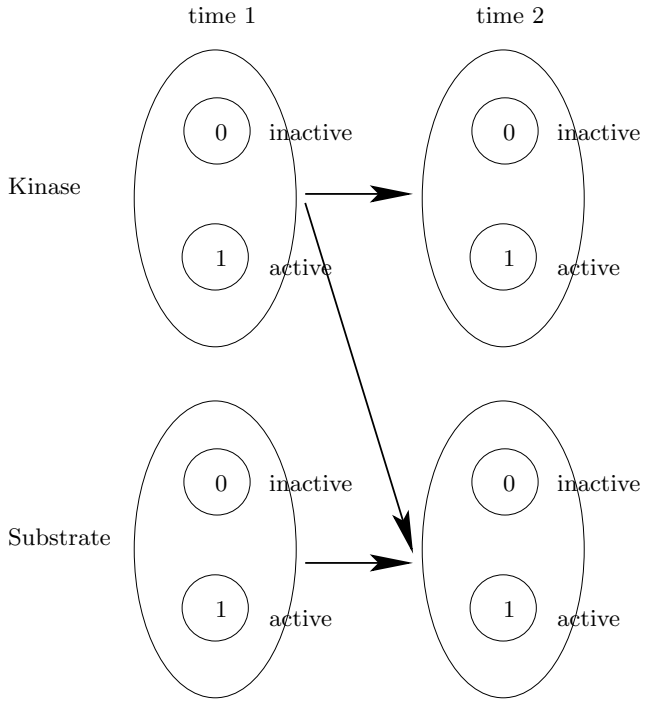
(2) When we refer to populations of proteins within a single cell or across different cells, we no longer have access to $x_1(t)$ (this is the state of a single kinase) but only the fractions $P(x_1(t) = 0)$ and $P(x_1(t) = 1)$. The influence on the transition probability can be defined in proportion to the concentration of active kinase as in

$$P_2(0 \to 1|x_1(t)) = \alpha\, P(x_1(t) = 1) P_2(0 \to 1)$$

where $\alpha \in (0, 1]$ is a proportionality constant.

Note that the state distributions $P(x_i(t) = j)$, $j = 0, 1$, $i = 1, 2$ will evolve differently depending on the interpretation.

We can try to capture the same setting with a Bayesian network by introducing a new state variable for each time point:

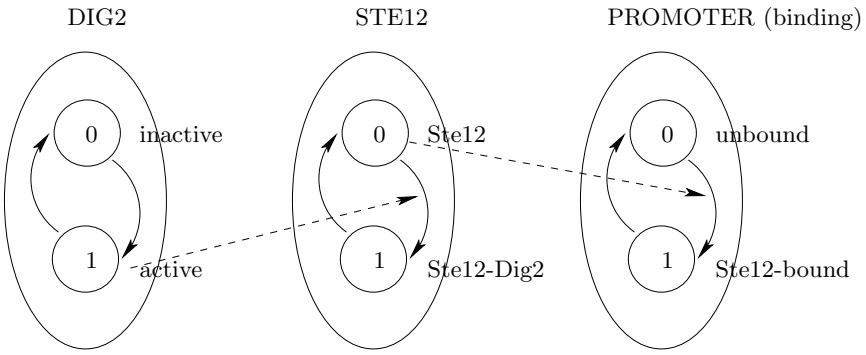


but this description hides the details about how the interaction occurs (only with active kinase and the substrate that is not already phosphorylated).

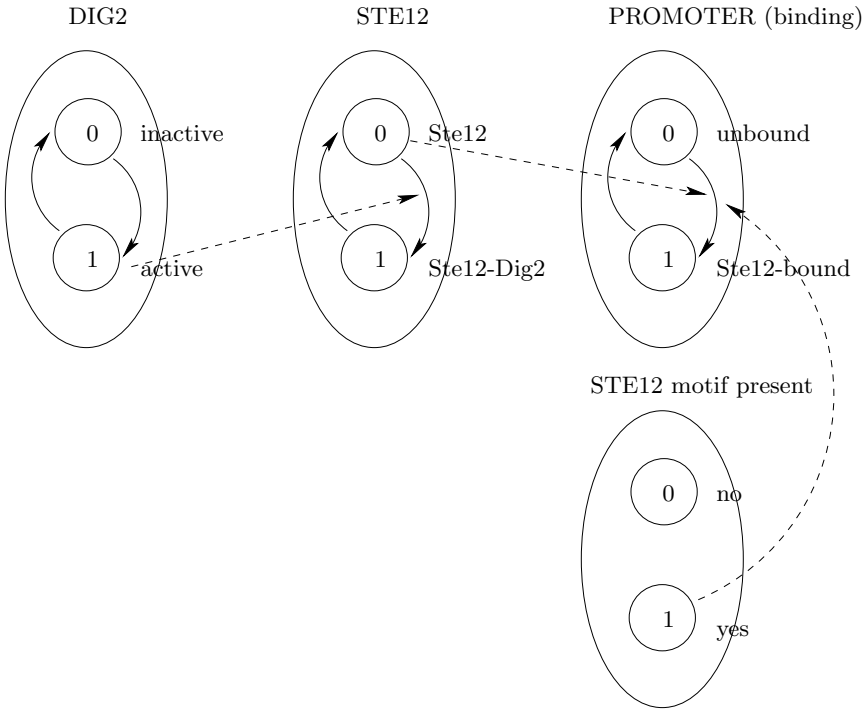
Example: Ste12 and Dig1/Dig2

Ste12 is a activator whose activity is modulated by two inhibitors, Dig1 and Dig2. Dig1 and Dig2 bind to different parts of Ste12 and have a different mechanism of inhibition. Dig1 presumably acts similarly to how Gal80 inhibits Gal4 activity. In other words, Dig1 does not preclude Ste12 from binding to promoters but rather blocks its activation domain. Dig2, in contrast, blocks the DNA binding domain of Ste12 thus preventing it from binding to the relevant sequence elements. The activity of Dig1 and Dig2 may be further modulated by phosphorylation. Dig1 is expressed constitutively while Dig2 may be up/down regulated substantially depending on the context (e.g., a 2-fold increase in pheromone response).

We can now use the representation discussed above to model how Dig2, if activated, blocks Ste12 from binding to a specific promoter element.



Moreover, we can include the presence of a motif to integrate binding assays and sequence analysis.

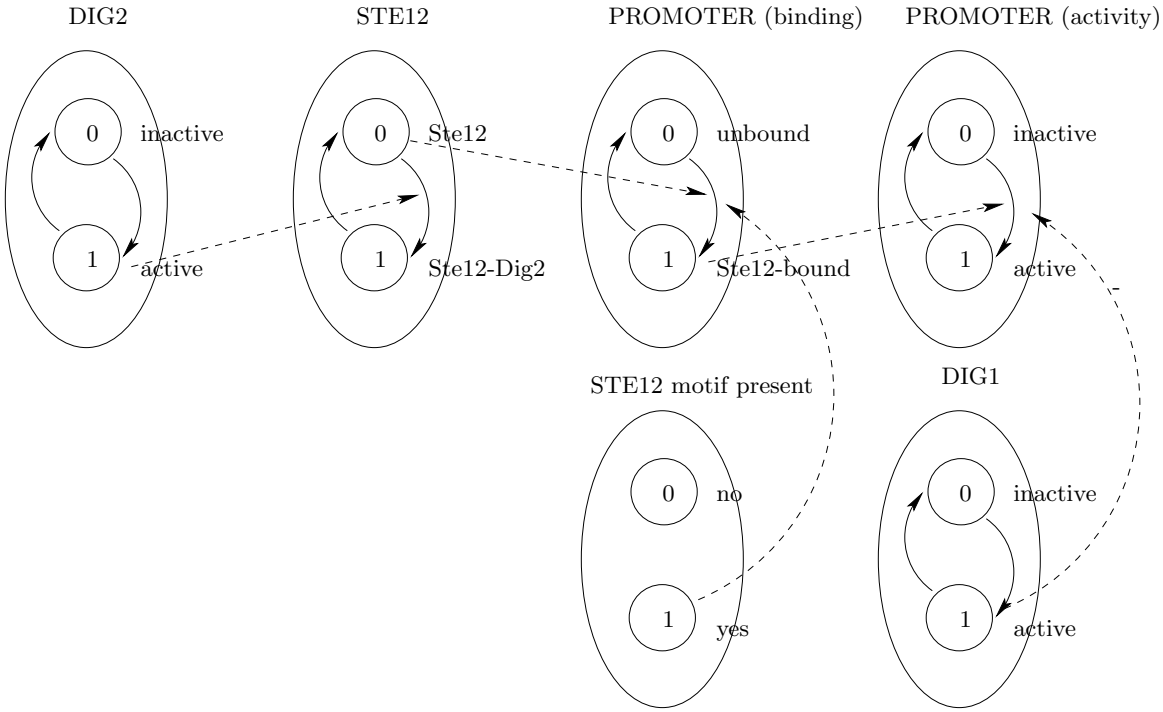


How do we determine the joint influence of the motif and the available Ste12 on whether Ste12 binds to the corresponding promoter? Under the second interpretation (population), we could define

$$P_{\text{promoter}}(0 \to 1) = \alpha\, P(\text{Ste12-Dig2}(t) = 1)\, P(\text{Ste12-motif} = 1)$$

In other words, the combined influence is proportional to the probability that the individual conditions hold.

We can finally add the effect of Dig1, for example, as follows:



where

$$P_{\text{prom-activity}}(0 \to 1) = \alpha' \, P(\text{Ste12-bound}(t) = 1) \, (1 - P(\text{Dig1} = 1))$$

where the transition probability is propotional to the fraction of inactive Dig1 since Dig1 is a repressor.