**Practical aspects of building models from modules 2-7-23**

The overall idea is to build models of each module and then link them together. Because each module is insulated, they should act independently. However, you still have to include affinity and capacity so that modules can respond differently when the strength of upstream signals (modules) change.

If you look at my original diagram of modules, I outline the modulation of input-output relationships. However, there is nothing on sensitivity or specificity of the module.

Thus, the fundamental parameters of a module description should include:

1. Sensitivity: the amount of input that produces half-maximal output (Kd)
2. Specificity: the list of inputs that can activate each module
   1. Sensitivity and specificity are coupled. Each different input (such as a ligand) will have its own sensitivity-specificity pairing (different Kd for each input).
   2. I am not sure if we should include a “dwell time” – how long the stimulus persists. Based on enzyme kinetics, this would be the “off rate”, which is pseudo first order. Rebinding of dissociated ligands reduces apparent off rates, so it might be more useful to represent Kd in terms of its component kf and kr because that would allow us to include the impact of membrane density or clustering.
3. Activation time: the amount of time that is required to activate the module. This might simply reflect the probably of assembling/dissembling all of the module components relative to the mean lifetime of the limiting step.
4. Offset: the amount of input that is required to induce a measurable change in the output
5. Gain (for proportional controllers): relationship between output and input minus the offset
6. Threshold (for switches): the amount of input needed to switch states of the module.
7. Backswitch (for switches): amount of input that will switch module off from on.
8. Controls: List of molecular processes that control input-output relationships.
   1. Change in sensitivity-specificity & relationship.
   2. Change in offset – specificity & relationship.
   3. Gain/threshold – specificity & relationship.
9. Limit: the maximum output of a module
10. Output: the specific information that is generated by a module

As conceived, the input-output behavior of any specific module is a function of the available input (species, concentration) and the “state” of its controlling species (e.g., HPER protein). If a HPER protein has 3 different phosphorylation sites/states, then it will be necessary to map the relationship between each state and module properties.

One of the key aspects of this description is time: how long does it take to generate an input from an output. This should be in terms of half-maximum activation time from a step-function input. The fact that you can have different inputs is key. Each input has its own sensitivity. Does it also have its own gain? Limit should be the same regardless of input sensitivity. Think about receptor internalization – ligand internalization did not seem to depend on ligand type. It just affected the binding time and thus concentration of the occupied surface receptors.

Let’s expand on the idea of internalization: If an activated EGFR can only generate a signal at the cell surface, then the mean time that a ligand-receptor complex can generate a signal is dependent on both dissociation and internalization (proportional to 1/(kd + ke)). If kd<<ke, then this approaches 1/ke. Conversely, if kd>>ke, then it will approximate 1/kd. We used to assume that all other reactions in the EGFR module are fast relative to either ke or kd, but if this was not the case, then large values of ke or kd might negatively impact downstream signaling.

The above suggests that we still need the functional equivalent of kf and kr for the module interface, or the process that couples the concentration of input to module activation. We should ignore non-limiting parameters such as MEK concentrations, considering them irrelevant as long as they are within normal parameter ranges. For example, the Raf module would start with the interaction between Rafs and Ras species in terms of protein concentrations and affinities. The module output is pERK levels, parameterized by measurements, but as long as MEK levels are not limiting, module properties should not be sensitive to them. It must contain the 4 min activation time. We do not intrinsically require a mechanistic explanation for the activation time, but it might be the case that if we understood the time-dependencies of the different module reactions, the 4-minute delay would naturally arise from them.

**General concepts**

A module has a set input-output relationship that is modulated by alterations in its core (HPER) protein. There are other components of the module that are not limiting but are necessary for their function. For example, Grb2 must bind to SOS1 and SOS2 for those proteins to be active, but it is in excess and rarely phosphorylated, so it is not limiting. The same is the case with 14-3-3 proteins and scaffolding proteins such as SHOC2 and deactivating proteins such as phosphatases or RASA1. These proteins can become important when mutated, so should probably be included in a mechanistic description of the module function. We should probably have a discussion on which proteins we should include in each module.

Because of the complexity of signaling pathways, it is probably best to build and parameterize each module independently. The fundamental parameters described above should be used as a guide.

**Nomenclature**

To keep everything straight, I propose that we keep a list of genes and shorthand for each (e.g., EGFR = R). A protein with an activation phosphorylation is preceded by a “p” (e.g., pR), a double-phosphorylated one by a “pp” and protein-protein interactions are represented by an underscore (e.g., a complex of the EGFR and Shc1 would be represented as pR\_L\_Shc because Shc only binds to a phosphorylated EGFR. When the bound Shc is phosphorylated by the EGFR, the complex would become pR\_L\_pShc). I am not sure how to represent a protein with a modulated phosphorylation (e.g., T693 in the EGFR, because they would represent a new protein species. Perhaps “R-T693” for the EGFR and “Shc-S139” for Shc? This would allow us to continue to use the “p” prefix and “\_” for complex formation.

**Rules:**

Systems tend to consolidate regulation to as few nodes as possible because of the cost of connections (leading to modularity and hierarchy) and the need for evolvability.

Regulation of a node can only be effected in a limited number of ways:

* Change in abundance – activity should scale with abundance.
* Change in localization – being in the right place at the right time.
* Reversible modification – changes the structure of a node.
* Conversion – irreversible change in a critical component.

If a system evolves by consolidating regulation to a few nodes, then it holds that it must somehow prevent changes in other important, non-regulated nodes from altering the system. This is especially true for the core components. How is this done? Obviously, preventing a change in abundance, localization, modification and conversion is essential, but mechanistically, how is this accomplished? Here is where I think that feedback plays a crucial role in generating compensation for the stochastic fluctuations in these properties.

In the case of protein abundance, a change in abundance is only regulatory if the abundance is limiting. Having a non-limiting (i.e., high) abundance is an obvious way to ensure a non-regulatory role of a protein, even if the protein is essential.

So, what we are saying is that the pattern of protein (or nucleic acid) expression, localization, modification and conversion across time and cell type contains information on the role that the molecule plays in the cell and the mechanisms the cell is using to regulate its function.