(R)evolution of Complex Regulatory Systems

Rune Linding (22 June 2010)

Science Signaling **3** (127), eg4. [DOI: 10.1126/scisignal.3127eg4]

Rune Linding is the team leader of the Cellular and Molecular Logic Team, Section of Cell and Molecular Biology, at The Institute of Cancer Research, London, UK.
E-mail: linding@icr.ac.uk

Cite as:
R. Linding, (R)evolution of complex regulatory systems. *Sci. Signal.* **3**, eg4 (2010).

10.1126/scisignal.3127eg4

(R)evolution of Complex Regulatory Systems

WHEN I THINK OF EVOLUTION OF BIOLOGICAL SYSTEMS, I IMAGINE HOW MILLIONS OF YEARS OF exposure to changing conditions have forced cellular molecular machines into robust and highly conserved complexes. In contrast, I bet signaling systems are much more dynamic and capable of revolutionizing cell behavior over a few months or days—for example, in response to the largely altered genomic environment of a cancer cell. Signaling systems are exciting to study precisely because they are some of the most complex and dynamical systems that we know. The study of evolution in signaling systems is thus challenging because of the range of time scales over which these systems both operate and evolve. The ability to compare, describe, and model the evolution or state changes in these systems will have profound impact not just on biology but also on drug development, engineering, and complex systems theory.

Eukaryotic signaling proteins consist of two different types of functional modules: protein domains and linear motifs. Whereas domains can be defined as large (>30 residues) globular structures with defined regulatory, binding, or catalytic activities, linear motifs are short (<10 residues) colinear sequences that often reside in disordered regions. Posttranslational modifications of linear motifs are keys to cellular information processing through directional and dynamic protein interaction networks. Protein phosphorylation can, for example, modulate the binding of protein domains to a tyrosine- or serine-containing motif and thereby control the dynamics, timing, and strength of a physical interaction.

Advances primarily in mass spectrometry have led to the identification of thousands of cellular posttranslationally regulated linear motifs. However, most of these are uncharacterized with respect to their roles in signaling, because the enzymes responsible for their modification and the domains that recognize them are unknown. Although computational tools (such as ELM, ScanSite, NetworKIN, and NetPhorest) have been developed to analyze, predict, model, and classify linear motifs and their binding partners, these tools would benefit from incorporation of evolutionary information. But first we must understand the full dynamic range of module and signaling system evolution.

Evolution of signaling proteins and systems can be studied at three levels. First, we can evaluate evolution at the level of sequence conservation of the modules. Second, we can analyze evolution at the level of modular organization or protein architecture (that is, the combination and order of the domains and motifs). For example, we can explore how SH2 domains coevolve with kinase domains to "build" the overall specificity of a tyrosine kinase. Third, we can study evolution at the level of signaling networks, where different biological implementations may yield similar outputs. At the sequence level, domains are frequently conserved over long evolutionary distances and change through divergent evolution. Much less is known about the evolution of linear motifs, although they likely evolve more rapidly and through convergent evolution due to their short length.

Sequence conservation analysis of proteome-wide phosphorylation data revealed that many previously unidentified phosphorylation sites are not well conserved, leading to the proposal that many (up to 50%) are nonfunctional. Furthermore, phosphorylation sites with unknown function seem typically less conserved at the sequence level than functionally characterized sites. I see at least two problems with the conclusion that lack of conservation equates with lack of function. First, it is assumed that the modulation of protein function by phosphorylation is position-dependent; however, both experimental and computational studies have demonstrated that linear motifs can "jump" within the sequence throughout the evolution of a protein. Second, the observation that functionally categorized sites are more conserved than "novel" sites is most likely a trivial result of study biases, because biologists conventionally rely on sequence conservation to pick sites before embarking on a functional characterization study. However, some sites may

"… to fully appreciate the impact of modular reorganization, linear-motif jumping, and noise or operational freedom in cell signaling, we have to monitor and compare the signaling networks over both cellular and evolutionary time."

well be functionally silent or biological stochastic noise (in contrast to false-positives from mass spectrometry, which remains a challenge!). Finally, I would argue that cellular systems need operational freedom and, thus, many motif-domain interactions might not be "hard-wired" through evolution.

At the protein architectural level, domain recombination may facilitate a high degree of biological innovation. For example, experimentally engineered shuffling of protein domains resulted in a larger dynamic signaling range and greater phenotypic diversity than was achieved by engineered gene or domain duplication events. Thus, changing the order and combinations of modules in a protein is a powerful evolutionary tool kit. At the network level, comparative proteomics is a powerful approach, which has provided insight into disease-associated signaling networks through the identification of phosphorylation events that, while not necessarily positionally conserved in orthologous proteins, are maintained by evolutionarily conserved kinase-substrate interactions. This type of analysis may reveal how the signaling networks in diseased cells "evolve" and adapt to changing genomic or environmental conditions—for example, within a tumor or in response to a drug.

This brings us to realize that in order to fully appreciate the impact of modular reorganization, linear-motif jumping, and noise or operational freedom in cell signaling, we have to monitor and compare the signaling networks over both cellular and evolutionary time.

Biological entities are often conserved across millennia, and this property serves as the foundation for selecting such entities for functional studies. However, the cell does not live in evolutionary time; it lives here and now. Thus, operational freedom is needed in biological systems to provide responsive and emergent properties that enable cells to respond to changes in the environment or genomic lesions. This is why we need to study signaling (r)evolution!

–Rune Linding