Ockham's broom

Opinion

Molecular machines or pleiomorphic ensembles: signaling complexes revisited

Bruce J Mayer*†‡, Michael L Blinov*‡ and Leslie M Loew*§

Addresses: *Richard D Berlin Center for Cell Analysis and Modeling, †Raymond and Beverly Sackler Laboratory of Genetics and Molecular Medicine, ‡Department of Genetics and Developmental Biology, §Department of Cell Biology, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, CT 06030-3301, USA.

Correspondence: Bruce J Mayer. Email: bmayer@neuron.uchc.edu

Abstract

Signaling complexes typically consist of highly dynamic molecular ensembles that are challenging to study and to describe accurately. Conventional mechanical descriptions misrepresent this reality and can be actively counterproductive by misdirecting us away from investigating critical issues.

A cell must constantly monitor cues from its environment and adjust its activities accordingly. Faithful and reliable signal transduction is not only essential for normal life, but its malfunctioning underlies many human health problems. Enormous strides have been made in the past several decades toward understanding how this process works at the molecular level. It is notable that when describing the fruits of that work, those of us who work on cell signaling would be hard-pressed to avoid terms such as 'machinery' and 'mechanism'. The analogy between cell signaling and man-made machines is all-pervasive, frequently adopting the imagery of elaborate clockwork mechanisms or electronic circuit boards. This perception is undoubtedly shaped by what we know: the machines that we use in our everyday life and the ways that we describe such machines in diagrams or in words. But is this really an accurate, or useful, description of the actual processes used by cells? We will argue that signaling complexes typically consist of pleiomorphic and highly dynamic molecular ensembles that are challenging to study and to describe accurately. Conventional mechanical descriptions not only misrepresent this reality, they can be actively counterproductive by misdirecting us from investigating critical issues.

First, let us define what we mean by a *bona fide* manmade machine. A key property of such a structure is that it can be described in terms of a parts list and a diagram or blueprint for how those parts fit together. Any machine, from a can-opener to a computer chip to an Airbus, can be rendered in a diagram with sufficient detail that someone who has never seen one could make it from the component parts.

Using the diagram, one could assemble any number of individual machines, each of which would be virtually identical in appearance and performance.

Cells contain a number of structures that conform quite well to this idea of a machine (see Box 1). Ribosomes, for example, or proteasomes, or nuclear pores, all have a clearly defined structure. Indeed, the ribosome has been subjected to X-ray crystallography, and the complex interlocking relationship of its many component proteins and structural RNAs has been revealed in molecular detail. The same list of components, in the same stoichiometry and physical relationship, is found in every ribosome in the cell (of course posttranslational modifications and accessory factors provide some variation, but the basic plan is the same). Because the parts interlock in a unique configuration, with multiple interactions between multiple components, the assembly of such structures is highly cooperative. This means that partly assembled structures are unstable and transient, whereas the fully assembled structure is very stable and unlikely to fall apart.

Now let us compare these machine-like structures with the complexes that mediate signal transduction in the cell. As an example, consider a transmembrane receptor for a mitogen such as platelet-derived growth factor (PDGF). How this receptor transduces signals has been worked out in great detail [1], and will briefly be summarized here (Figure 1). The receptor has intrinsic tyrosine kinase activity (that is, it can catalyze the transfer of phosphate from ATP to tyrosine groups on substrate proteins), but this activity is quiescent in the unstimulated receptor. Once the receptor binds its ligand, however, receptor dimerization or oligomerization increases the likelihood of transphosphorylation of the receptor by its new-found neighbors. Phosphorylation at a critical site in the catalytic domain induces conformational changes that lock the domain into an active conformation that can go on to phosphorylate other receptors, as well as other substrate proteins in the vicinity.

<div style="border:1px solid; padding:10px;">

Box 1

Different classes of molecular assemblies

Molecular machines
- Stoichiometric
- Specific interactions
- Fixed size
- discrete molecular states
- Functional assembly requires complete set of subunits (assembly highly cooperative)
- Amenable to structural biology tools
- Examples: ribosomes, molecular motors, nuclear pore complex, flagella, proteasomes…

Pleiomorphic ensembles
- Non-stoichiometric
- Combinatorial interactions
- Heterogeneous size
- Spectrum of molecular states
- Lifetime of assembly greater than subunit residency (assembly may or may not be cooperative)
- New experimental and mathematical tools needed
- Examples: receptor complexes, adhesion complexes, mRNA splicing complexes, trafficking intermediates…

</div>

Heterogeneity due to phosphorylation status

So far so good - the receptor itself seems to be acting as a molecular machine, and indeed receptor catalytic domains have been crystallized, revealing in exquisite detail the conformational changes involved in activation. But here is where it gets tricky. The typical receptor has many different potential autophosphorylation sites (in the case of the PDGF receptor at least ten), and it is highly unlikely that all sites can be phosphorylated at the same time. Furthermore, abundant intracellular phosphatases are constantly working to remove phosphates as soon as they are added, so at any time a particular activated receptor molecule is likely to be phosphorylated only on a subset of the ten possible sites. If each of the 10 sites can be phosphorylated or dephosphorylated independently of the others, the total number of potential phosphorylation states per receptor will be 2^{10} (1,024). But because receptors must dimerize in order to activate, each activated receptor dimer has a much larger number of potential states - in this case, more than 500,000 different unique combinations of phosphorylation states (which is given by the expression $Y[Y + 1]/2$, where $Y = 2^{10}$).

The state of phosphorylation is critically important because it is these very phosphorylation sites that serve to transmit downstream signals from the activated receptor. They do so by binding to cytosolic effector proteins with phospho-tyrosine-binding motifs, most commonly Src homology 2 (SH2) domains [2]. By binding to the receptor, these signaling proteins are brought into close proximity to their substrates (which in many cases reside exclusively on the membrane), and they may also be phosphorylated by the receptor, which can modulate their activity. There are more than 100 of these cytosolic effector proteins that can bind to the receptor, but each of them binds to only a subset of the sites on the receptor with reasonably high affinity [3,4]. Thus, which effectors ultimately bind to the receptor will depend on the local concentration of each of the effectors and on which sites on the receptor are phosphorylated. Steric clashes and cooperativity among different binding partners may also affect which effectors are bound.

Effector binding leads to a tremendous increase in the number of potential states for the receptor. Even if we oversimplify and assume that each phosphorylated site can bind to only one effector (so the possible states for each site are now three: unphosphorylated; phosphorylated but unbound to effector; and phosphorylated and bound to effector), the total potential number of states for each receptor monomer increases to 3^{10} (around 60,000) and for the receptor dimer to almost 2 billion! This does not even take into consideration the possibilities that any bound effector may or may not be phosphorylated by the receptor, or be simultaneously bound to yet another effector. Clearly, the theoretical number of possible states is virtually infinite, certainly far more than the actual number of receptors in the cell (which is generally on the order of tens of thousands of receptor molecules). Of course, the actual number of possible states might be smaller because of steric clashes and other mechanical and physical constraints, but in most cases the experimental data necessary to eliminate improbable states are lacking.

This combinatorial explosion of possible states makes it very difficult to pin down exactly what we mean by 'activated PDGF receptor': each receptor dimer or cluster of activated receptors is likely to be different from other activated receptors in terms of exactly which sites are phosphorylated, and which effectors are bound to those sites. In reality, the activated receptor looks less like a machine and more like a pleiomorphic ensemble or probability cloud of an almost infinite number of possible states, each of which may differ in its biological activity. In this sense, the activated receptor is rather like the genomes of RNA viruses, which because of the inherent inaccuracy of their replication can only be described in terms of 'average' sequence, from which each individual genome will deviate to some extent [5]. Although not explicitly discussed here, the same arguments could be applied to other complex but heterogeneous assemblies that regulate such diverse cellular processes as adhesion to the extracellular matrix and other cells, mRNA splicing and transport, localized actin remodeling, and many others (see Box 1).
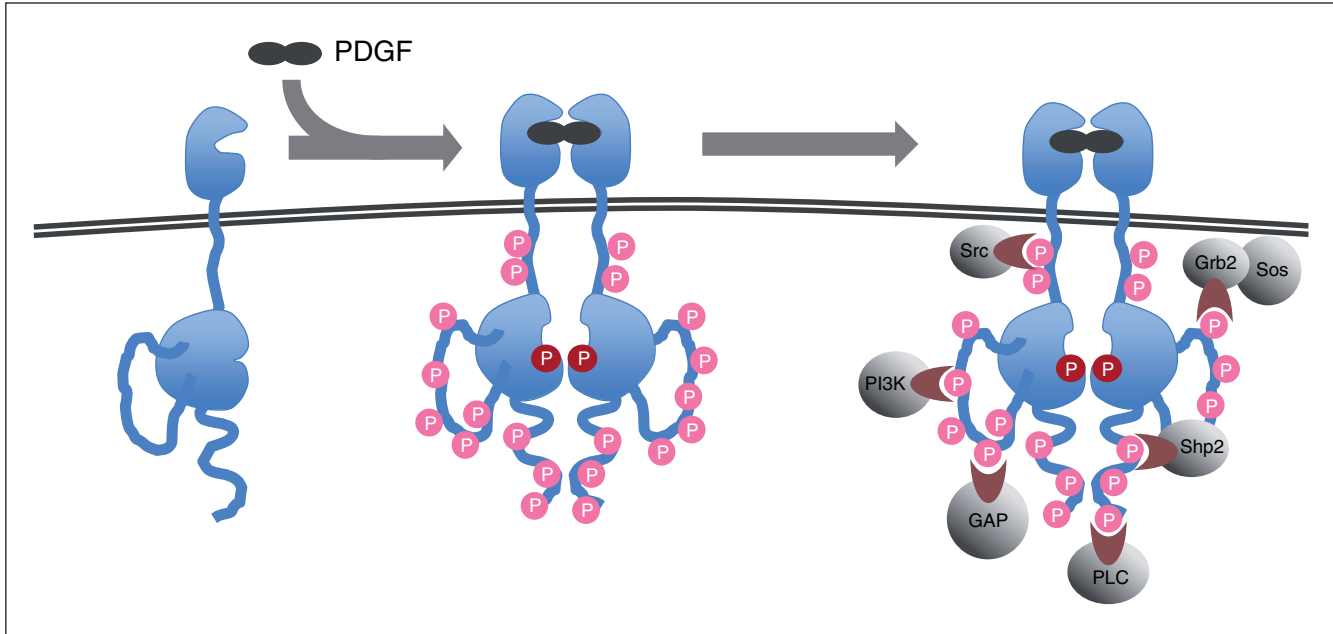
Figure 1

Signaling by the platelet-derived growth factor (PDGF) receptor. The unliganded receptor is monomeric and its tyrosine kinase catalytic activity is low (left). On binding to dimeric PDGF, the receptor dimerizes, its catalytic activity increases, and receptors transphosphorylate each other on a number of different sites, represented by pink circles (center). These phosphorylated sites (with one exception) serve to recruit cytosolic effector proteins (gray) that contain phosphotyrosine-specific modular binding domains (right). The exception is the activating phosphorylation, located on the catalytic domain of the receptor adjacent to the active site (red circle). Representative effectors depicted are: Src, Src-family non-receptor tyrosine kinases; PI3K, regulatory subunit of phosphatidylinositol 3-kinase; GAP, RasGAP, a GTPase-activating factor for Ras; PLC, phosphatidylinositol-specific phospholipase C-γ; Shp2, SH2-containing tyrosine phosphatase; Grb2, adaptor protein that recruits the Ras guanine-nucleotide exchange factor Sos.

Despite the many potential states of the receptor, we might safely ignore this complexity if it had no real impact on signaling. This might be the case if only a few of the many possible states were actually populated (that is, present in significant amounts in the cell). Alternatively, we would not need to account for the precise state of each of the individual receptors if the effective output from the many individual receptors in the cell is averaged over the whole population. So it is worth looking at what is known about these two possibilities. Unfortunately, the short answer is very little: virtually all the analytical methods now used to study signaling proteins can only tell us about the average state of the population, not the state of individual molecules. Such methods necessarily fail to capture information on the distribution of different states (Figure 2). The technique of top-down mass spectrometry is just beginning to be used to quantify different posttranslationally modified isoforms of histones [6,7], but this approach has yet to be applied to signaling molecules such as activated receptors. So for the moment, we really do not have the kind of experimental data we need to estimate the seriousness of the problem.

We do know enough, however, to suggest that we ignore this issue at our peril. Let us consider a few specific cases. Things would not be so bad if the receptor, for example, actually existed in only two predominant states: inactive, in which no sites are phosphorylated; and active, in which all possible sites are phosphorylated. This is not an unreasonable idea, and in fact many quantitative models of receptor tyrosine kinase (RTK) signaling make just this assumption [8]. But there really is no solid experimental evidence to support this model, and even if it were true, at the next level of signaling (the binding of SH2-containing effectors), it is almost certain that the relatively low affinity of such interactions, and the likely steric clashes with multiple proteins trying to bind to a number of closely spaced sites, would make it unlikely that all sites would ever be fully occupied by a complete set of effectors. Thus, it is hard to escape the conclusion that activated receptors are, by necessity, heterogeneous, non-stoichiometric ensembles.

We still might be able to ignore this heterogeneity if signal output depended only on the aggregate or average state, summed over all of the activated receptors in the cell. In

Figure 2

Averaging leads to loss of information. In the panel on the right, each pixel is the average of the properties of all the individual pixels in the panel on the left. By averaging, all information on the range of properties of individual pixels, and their spatial distribution, is lost. Most biochemical methods used to probe signaling complexes, such as immunoprecipitation followed by immunoblotting or mass spectrometry, average the properties of complexes over the entire population.

other words, if half the receptors bound effector 1 and half bound effector 2, signal output would be equivalent no matter how those effectors were distributed among the individual receptors - for example, half of the receptors bound to both 1 and 2 and the other half bound none, versus half bound to 1 and the other half bound to 2 (Figure 3). While this may be true in some situations, in others it clearly is not. For example, different effectors often interact positively or negatively, reinforcing or canceling out each other's activity. Take the case of Grb2 (an adaptor that recruits Sos, which in turn activates a key downstream effector, Ras), and RasGAP, which inactivates Ras (Figure 3a). Clearly, the extent and spatial distribution of Ras activity would be quite different if both Grb2 and RasGAP were recruited to the same receptor, compared with the case when the two are recruited to different spatially separated receptors (Figure 3c). Another example illustrates the importance of the temporal order of assembly of complexes. The effector phospholipase C-γ (PLC-γ) cleaves the phospholipid phosphatidylinositol 4,5-diphosphate ($PI(4,5)P_2$) into two second messengers (diacylglycerol and inositol trisphosphate (IP_3)), whereas a second effector, phosphatidylinositol 3-OH-kinase (PI 3-kinase), uses the same substrate but phosphorylates it, generating yet another second messenger, $PI(3,4,5)P_3$. It is known that the products of each of these effectors cannot be used as substrates by the other. This implies that whichever effector is recruited first will rapidly deplete the substrate in the vicinity of the receptor before the second one is recruited.

Heterogeneity due to protein-protein interactions

In the example of RTK signaling we have emphasized the complexity and heterogeneity induced by differential phosphorylation. A second major source of heterogeneity in signaling complexes is protein-protein interactions. Often these two are inextricably linked, as one of the major roles of posttranslational modifications such as phosphorylation is to regulate protein-protein interactions [9]. But more generally, we know that signal processing almost always involves the regulated assembly of multi-protein complexes, often mediated by modular protein binding domains [10]. Such interactions can be highly specific, but in many cases a particular site may bind to several (or many) different proteins with similar affinity - for example, the binding of tyrosine-phosphorylated peptides to the SH2 domains of multiple proteins [3]. It is self-evident that if more than one of these potential partners is present in the local environment, the actual complexes formed will be a mixture of different species.

Again, the tools at our disposal to study protein interactions make it difficult to ascertain how big a problem this might be. But it is important to keep in mind that any binding interaction is dependent on the concentration of the partners, and the affinity (dissociation constant, K_D) of each interaction. Strong interactions can be insignificant if the concentration of the partners is very low, or if many competing binders are present; conversely, relatively weak interactions can be critically important for biological processes when the local concentration of the partners is sufficiently high (this is often seen, for example, when relatively weak intramolecular interactions hold a protein in one conformation until they are disrupted by competition with another binding partner in *trans*). Furthermore, cooperative interactions among multiple binding partners can also strongly affect the complexes that form preferentially [11,12].

For these reasons, comprehensive lists of protein-protein interactions (or more grandiosely, the so-called 'interactome') should be viewed with some skepticism. Such data are almost always based on some simple assay (such as yeast two-hybrid, or pull-down of one component followed by mass spectrometry), and anything rising above the detection limit for that particular assay is scored as positive. Although thinking of binding in binary terms (binds/does not bind) makes sense in a mechanical world (a part either fits or it does not), it really does not make sense in a world where the amount of a specific complex can only be predicted if we know the local concentration and affinity of all possible interaction partners. More important, it is rare that such interaction data can be validated for functional relevance. In the absence of independent evidence that the proposed interaction has real biological consequences, such as a known genetic interaction that is consistent with the observed biochemical interaction, global interaction maps provide only a crude guide to what is possible.

Once again we should ask whether this is really a serious practical concern, or whether it can safely be swept under the rug. This issue has been addressed more or less directly in the case of SH3 domains, another modular protein-binding domain of which there are more than 300 examples in the human proteome [13]. Because most SH3 domains bind to a common peptide consensus of PxxP (P is proline, x is any amino acid), usually flanked by a basic residue, and early studies with purified domains and peptide ligands showed clearly overlapping specificities, it was long suspected that these domains may be rather promiscuous in their binding *in vivo* [14]. Lim and colleagues looked at specificity of SH3 domains in the yeast *Saccharomyces cerevisiae* (which has fewer than 30 SH3 domains in total), and their results suggested that, for the most part, each SH3 domain binds non-overlapping targets

Figure 3

Individual receptor states can influence signal output. **(a)** Grb2 and RasGAP (GAP) bind to distinct sites on the PDGF receptor (blue line). For clarity, only one receptor molecule is shown; the actual activated form of the receptor is a dimer. The consequence of Grb2 binding is Ras activation, through the Ras guanine-nucleotide exchange factor Sos. The consequence of GAP binding is Ras inactivation (by stimulating its intrinsic GTPase activity). Thus, these two effectors have opposing effects on Ras activity. **(b)** Three different possible distributions of GAP and Grb2 on receptors are depicted; in all cases, an average of 0.5 molecule of Grb2 and 0.5 molecule of GAP are bound per receptor. Left, binding of Grb2 and GAP are positively correlated; middle, binding of Grb2 and GAP are independent; right, binding of Grb2 and GAP are negatively correlated. **(c)** Effects of different distribution of effectors on Ras activity are depicted. Left, where binding of GAP and Grb2 are positively correlated, Ras activity will be relatively low and uniform. Right, where binding of Grb2 and GAP are negatively correlated, areas of high Ras activity and low Ras activity will be interspersed.

in vivo. They suggested that this specificity arose not only by positive selection for useful interactions, but also through negative selection against nonproductive or counterproductive competing interactions [15]. A more recent comprehensive study of the yeast SH3 binding repertoire partially supports this conclusion, showing that while the majority of putative SH3 binding partners are

likely to interact with high affinity with only a single SH3 domain, a significant fraction have multiple possible partners [16]. One can, however, imagine that in human cells, endowed with ten times the number of SH3 domains (and a proportional increase in potential binding partners), the likelihood of multiple competing partners is considerably higher. Furthermore, as mentioned above, most interaction screens cannot detect relatively low-affinity interactions that may nonetheless be biologically important. Thus, the experimental data now available are equivocal, and certainly are consistent with competition among binding partners during the assembly of signaling complexes.

The ephemeral nature of signaling complexes

Another important and underappreciated attribute of signaling complexes is their ephemeral nature. Many of the protein-protein interactions that drive signaling are of modest affinity (typically high nanomolar to low micromolar K_D values), and this necessarily implies that such complexes are highly dynamic, with half-lives on the order of seconds or less. Posttranslational modifications such as phosphorylation are likely to be similarly transient, as kinases and phosphatases continually battle it out in the cytosol. In the case of tyrosine phosphorylation, this dynamic nature is illustrated by what happens when the phosphatase inhibitor vanadate is added to cells: there is an enormous and quite rapid increase in levels of protein tyrosine phosphorylation, implying a very rapid cycle of phosphorylation and dephosphorylation under normal conditions. Thus, signaling complexes, formed by posttranslational modifications and protein interactions, are unlikely to be stable in any traditional sense of the word, but will rather flicker rapidly between many different states.

Perhaps the most significant barrier to appreciating the dynamic, heterogeneous aspect of signaling complexes is the lack of a good analogy from our daily experience. This contributes to a second related problem, our inability to depict such interactions diagrammatically. Indeed, the typical 'cartoons' of signaling pathways, with their reassuring arrows and limited number of states (as seen here in Figure 1), could be the real villain of the piece. Instead of simplifying an inherently complex system so that the key points can be grasped, we would argue that such diagrams actively mislead, implying a specificity and homogeneity that does not at all reflect the messy reality of actual signaling complexes. To some extent this can be blamed on historical precedents (those yellowed diagrams of metabolic pathways hanging on the wall), and on the prosaic demands of publishing our results. It is much easier to write and publish a paper suggesting Protein X is necessary for transmitting a signal from A to B, than one showing that Protein X is one of many potential components of a heterogeneous ensemble of signaling complexes that together couple A to B. Two currently popular representations, protein-interaction networks or reaction network diagrams, are little better. Protein-interaction networks capture the heterogeneity of possible interactions, but in most cases the connections (edges) between proteins (nodes) provide no information on the likelihood of interaction between proteins, or how those interactions may depend on others, or any temporal aspect of interactions. Reaction network diagrams are clear and unambiguous, but fundamentally are similar to cartoons such as Figure 1. Details pertaining to the heterogeneity of complexes are lacking, and adding more details only adds to the confusion by making the diagram unreadable.

Are there any answers?

Is there a way around this conceptual hurdle? One approach is to use a unified, consistent graphical notation standard - Systems Biology Graphical Notation (SBGN) - to depict functional relationships among components in signaling pathways and networks [17]. This is a promising development, but the complexity of this task has already led to several distinct formats of SBGN - 'Process Diagrams', 'Entity Relationship Diagrams' and 'Activity Flow Diagrams', each of which captures only some aspects of complexity. Furthermore, quantitative aspects of interactions such as affinities cannot be captured and depicted in these formats, as SBGN aims merely at capturing qualitative, or functional, relationships among entities.

Computational models may provide another approach to capturing the dynamic, heterogeneous aspect of signaling complexes. For such models to provide an accurate and comprehensive representation of the system and its interconnections, each biological component (protein, RNA, and so on) would have attributes specifying its physical and chemical activities and interactions with all other components (such as on-rates and off-rates of binding interactions, K_m of enzymatic reactions, cooperative relationships). Development of community standards for data exchange among databases can greatly facilitate the construction of models. These could include standards (such as BioPAX) to access qualitative data within multiple pathway databases, as well as standards for exchange of quantitative data (such as models encoded in the SBML or CellML formats) among multiple model databases (for example, the Virtual Cell Database and BioModels.net) [18-22].

Thus, computational models can serve not only as tools for quantitative predictions of experimental outcomes, but also as repositories of precisely the kind of detailed information that is lacking in a typical cartoon diagram of a signaling mechanism. One can envisage logging in to a public model where clicking on a component of interest brings up a battery of potential modifications, interactions and activities, and the likelihoods and potential consequences of each under a variety of 'typical' sets of conditions, or specific conditions set by the user. Although designing user interfaces that would be helpful and

intuitive for experimental biologists may be a challenge, surely this goal is achievable in the relatively near future.

Using quantitative models that fully account for the heterogeneity of signaling complexes to actually predict signaling outputs is still rather challenging, however, in part because the proliferation of possible states for the system makes calculating the concentrations of each of these states extremely computationally intensive. Tricks now being developed to get around the specific enumeration of each state, such as rule-based modeling, are likely to help in this regard [18,23]. Stochastic and on-the fly simulations that can include all populated states is a particularly promising approach that can accommodate the concept of pleiomorphic ensembles instead of signaling machines. Given the ubiquity of cooperative interactions among proteins in signaling, we are also likely to need new mathematical tools to predict and quantitatively estimate the effects of cooperativity on the composition and activity of signaling complexes.

In addition to the development of quantitative models that can more accurately predict what can happen, new analytic methods are also urgently needed to expand our ability to monitor what actually does happen, at the single-molecule level, in the cell. Mass spectrometry and other approaches have begun to be able to quantify the number of molecules with specific combinations of posttranslational modifications, or specific binding partners, under different conditions. Imaging methods and biosensors with single-molecule resolution will begin to provide similar information within the spatial and temporal context of the living cell [24].

The pleiomorphic, heterogeneous, non-stoichiometric nature of signaling complexes provides a serious conceptual challenge for biologists, who are naturally more comfortable thinking of mechanical devices with states that are clearly defined and limited in number. But the current practice of avoiding these properties because they are difficult to study and to describe is likely to be a mistake. Only by confronting this issue head-on will be able to assess, once and for all, its real impact on signal transduction.

Acknowledgements

References

1. Heldin CH, Westermark B: **Mechanism of action and *in vivo* role of platelet-derived growth factor.** *Physiol Rev* 1999, **79:**1283-1316.

2. Pawson T: **Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems.** *Cell* 2004, **116:**191-203.

3. Jones RB, Gordus A, Krall JA, MacBeath G: **A quantitative protein interaction network for the ErbB receptors using protein microarrays.** *Nature* 2006, **439:**168-174.

4. Liu BA, Jablonowski K, Raina M, Arce M, Pawson T, Nash P: **The human and mouse complement of SH2 domain proteins - establishing the boundaries of phosphotyrosine signaling.** *Mol Cell* 2006, **22:**851-868.

5. Domingo E, Holland JJ: **RNA virus mutations and fitness for survival.** *Annu Rev Microbiol* 1997, **51:**151-178.

6. Garcia BA, Pesavento JJ, Mizzen CA, Kelleher NL: **Pervasive combinatorial modification of histone H3 in human cells.** *Nat Methods* 2007, **4:**487-489.

7. Phanstiel D, Brumbaugh J, Berggren WT, Conard K, Feng X, Levenstein ME, McAlister GC, Thomson JA, Coon JJ: **Mass spectrometry identifies and quantifies 74 unique histone H4 isoforms in differentiating human embryonic stem cells.** *Proc Natl Acad Sci USA* 2008, **105:**4093-4098.

8. Birtwistle MR, Hatakeyama M, Yumoto N, Ogunnaike BA, Hoek JB, Kholodenko BN: **Ligand-dependent responses of the ErbB signaling network: experimental and modeling analyses.** *Mol Syst Biol* 2007, **3:**144.

9. Seet BT, Dikic I, Zhou MM, Pawson T: **Reading protein modifications with interaction domains.** *Nat Rev Mol Cell Biol* 2006, **7:**473-483.

10. Bhattacharyya RP, Reményi A, Yeh BJ, Lim WA: **Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits.** *Annu Rev Biochem* 2006, **75:**655-680.

11. Gibson TJ: **Cell regulation: determined to signal discrete cooperation.** *Trends Biochem Sci* 2009, **34:**471-482.

12. Whitty A: **Cooperativity and biological complexity.** *Nat Chem Biol* 2008, **4:**435-439.

13. Kärkkäinen S, Hiipakka M, Wang JH, Kleino I, Vähä-Jaakkola M, Renkema GH, Liss M, Wagner R, Saksela K: **Identification of preferred protein interactions by phage-display of the human Src homology-3 proteome.** *EMBO Rep* 2006, **7:**186-191.

14. Mayer BJ: **SH3 domains: complexity in moderation.** *J Cell Sci* 2001, **114:**1253-1263.

15. Zarrinpar A, Park SH, Lim WA: **Optimization of specificity in a cellular protein interaction network by negative selection.** *Nature* 2003, **426:**676-680.

16. Tonikian R, Xin X, Toret C, Gfeller D, Landgraf C, Panni S, Paoluzi S, Castagnoli L, Currell B, Seshagiri S, Yu H, Winsor B, Vidal M, Gerstein M, Bader G, Volkmer-Engert R, Cesareni G, Drubin D, Kim P, Sidhu S, Boone C: **Bayesian modeling of the yeast SH3 domain interactome predicts spatiotemporal dynamics of endocytosis proteins.** *PLoS Biol* 2009, in press.

17. Le Novère N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, Demir E, Wegner K, Aladjem MI, Wimalaratne SM, Bergman FT, Gauges R, Ghazal P, Kawaji H, Li L, Matsuoka Y, Villéger A, Boyd SE, Calzone L, Courtot M, Dogrusoz U, Freeman TC, Funahashi A, Ghosh S, Jouraku A, Kim S, Kolpakov F, Luna A, Sahle S, Schmidt E, *et al.*: **The Systems Biology Graphical Notation.** *Nat Biotechnol* 2009, **27:**735-741.

18. Blinov ML, Ruebenacker O, Moraru I: **Complexity and modularity of intracellular networks: a systematic approach for modelling and simulation.** *IET Syst Biol* 2008, **2:**363-368.

19. Garny A, Nickerson DP, Cooper J, Weber dos Santos R, Miller AK, McKeever S, Nielsen PM, Hunter PJ: **CellML and associated tools and techniques.** *Philos Transact A Math Phys Eng Sci* 2008, **366:**3017-3043.

20. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr JH, Hunter PJ, Juty NS, Kasberger JL, Kremling A, Kummer U, Le Novère N, Loew LM, Lucio D, Mendes P, Minch E, Mjolsness ED, *et al.*: **The systems**

biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 2003, **19:**524-531.

21. Le Novère N, Bornstein B, Broicher A, Courtot M, Donizelli M, Dharuri H, Li L, Sauro H, Schilstra M, Shapiro B, Snoep JL, Hucka M: **BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems.** *Nucleic Acids Res* 2006, **34(Database issue):**D689-D691.

22. Moraru II, Schaff JC, Slepchenko BM, Blinov ML, Morgan F, Lakshminarayana A, Gao F, Li Y, Loew LM: **Virtual Cell modelling and simulation software environment.** *IET Syst Biol* 2008, **2:**352-362.

23. Hlavacek WS, Faeder JR, Blinov ML, Posner RG, Hucka M, Fontana W: **Rules for modeling signal-transduction systems.** *Sci STKE* 2006, **2006:**re6.

24. Huang B, Bates M, Zhuang X: **Super-resolution fluorescence microscopy.** *Annu Rev Biochem* 2009, **78:**993-1016.