

Comment

## Location, location, location

Gregory A Petsko

Address: Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02454-9110, USA.  
E-mail: [petsko@brandeis.edu](mailto:petsko@brandeis.edu)

Published: 6 June 2001

*Genome Biology* 2001, **2(6)**:comment1008.1–1008.2

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2001/2/6/comment/1008>

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

My father always said one should buy the cheapest house in the best neighborhood one could afford. He reasoned that the quality of the neighborhood would ensure that the value of the property would increase with time, and the relatively low price of the house would mean that it would not be overpriced and so would have a high percentage return. When I once questioned this advice on the grounds that the cheapest house might be cheap because it was poorly constructed or in need of repairs, he responded that one could always sell a relatively low-priced house to someone else if the neighborhood was good. “The only thing that really matters,” he said, “is location.” Years later I had a friend who bought and sold commercial property. He seemed to do well at it, and I asked him what his secret was. He replied with a maxim that he claimed was followed by everyone who was successful in the real-estate business. “The three most important words in real estate,” he said, “are location, location, location.”

Soon, if you ask cell biologists what they would most like to know about the products of the genes they are studying, you may get the same response. If you asked them now, most would probably say “The structures” or “The function(s),” but for proteins for which structural information is already available either “Location” or “The proteins it interacts with” are, I think, also very good choices. And I wager that before long these two alternatives will be almost synonymous, not only with each other but also with beginning to know the function.

I believe we can now make the categorical statement that there is no such thing as a freely floating protein in a eukaryotic cell. Everything is tied up: in complexes with other macromolecules, in cargo vesicles, by attachment to membranes, or as passengers on actin railroads in the cytoskeleton. Perhaps prokaryotes can be accurately described as bags of enzymes (though I wouldn't bet on that) - but eukaryotic cells are organized. In fact, it is increased organization, not increased gene number, that is the real hallmark of the complexity of eukaryotic cells. (The fission yeast *Schizosaccharomyces pombe*, whose genome has just been sequenced, has fewer

genes than the bacterium *Pseudomonas aeruginosa*.) And the key to this organization is location. In eukaryotic cells, proteins are targeted to the sites where they are needed in a dynamic fashion. Changes in targeting are used to alter protein function at the cellular level, even when the biochemical function of the gene product does not change.

Nowhere is this fact more evident than in signal transduction pathways. The number of protein kinases in, say, the human genome is large, but it is nowhere near the number of protein kinase substrates. Since we do not appear to have one kinase for each substrate, kinases must have less than absolute specificity. But in that case, how are they prevented from phosphorylating the ‘wrong’ protein at an inappropriate time? Location is one answer. If the kinase is targeted to the same location as its ‘correct’ substrate, a location different from that for any other potential substrate, then the action of that kinase can be made specific in a dynamic fashion, changing as needed by simply relocalizing kinase and/or substrate.

Or consider the small monomeric GTPase Tem1 from the budding yeast *Saccharomyces cerevisiae*. A member of the Ras superfamily, Tem1 is an essential gene product in yeast. It is involved, *inter alia*, in termination of M phase of the cell cycle. Many yeast proteins have been subjected to a systematic investigation of their interactions with other proteins by genome-wide two-hybrid analysis. Tem1 is one of these, and it has been found to interact physically with 24 different yeast gene products. Now the average protein-protein interface has been shown, by crystal structure determination of many complexes, to be at least 400 square Angstroms in contact area. If we assume that Tem1 can be approximated by a sphere of 25 Angstroms radius, then the protein has about 2,000 square Angstroms of surface area available for interaction at any one time. (Tem1 is not really spherical and its surface is far from smooth, but for our purposes these oversimplifications don't matter.) One concludes from this simple consideration that no more than about four proteins

can possibly bind to Tem1 at the same time, so how do we account for the fact that 24 proteins are able to do so? Differences in the timing of gene expression can account for some of the control of specificity, but most of that control has to come from targeting of Tem1, and its partners, to different locations in the cell at different times.

Although the most frequently employed targeting mechanism seems to be phosphorylation or the binding to a phosphorylated site on another protein, two other common types of targeting are by binding to membranes and binding to scaffold proteins. It is often difficult to recognize either a scaffold protein or something that will bind to one from examination of the sequence or even the structure of a protein, although some scaffolds (say, a protein with seven SH3 domains) are obvious. More work on the computational identification of possible sites of protein-protein interaction is clearly needed. Membrane-binding modules, on the other hand, can often be detected by sequence-gazing (although new ones are turning up all the time, and some of them are also used to bind other proteins instead). Covalent attachment of a protein to a lipid molecule that in turn localizes the protein to the membrane, as in the case of Ras, which is farnesylated at its carboxyl terminus, is also common. I have always been uncomfortable with the idea that these lipid anchors just insert into membranes willy-nilly by virtue of their hydrophobicity. I doubt that membranes in eukaryotic cells are really just random soups of lipids; that likelihood seems as remote to me as the possibility that eukaryotic cells are random soups of proteins. I think that membranes will be found to have many patches where specific lipids congregate, forming islands that target the lipid anchors, and lipid binding domains, of proteins not just to the membrane but to very specific places on the membrane. Control of the location and size of these patches by enzymatic modification and hydrolysis of phospholipids is likely to be a major area of research in the genomic era. So is the question (which in my view has received too little attention) of how proteins come off the membrane when they are to be targeted to a new location. Much more work is needed on all this.

Which brings me to my final thought: that understanding location and how it is used to control the action of gene products means a lot more than just doing the yeast two-hybrid screen on all the proteins in a genome. It is the dynamics of localization that matter: not just where something is, but when it is there. It does you very little good to buy the cheapest house in the best neighborhood if the following year your neighbors' houses disappear and are replaced by a shopping mall, or a prison. Ask anyone who has ever bought property for investment purposes and they will tell you that it is folly to assume that neighborhoods will always stay the same. For too long we have made that same assumption, unconsciously, in our thinking about how proteins function in the cell. But in the age of genomics we will all have to consider location as something that is not only the key to much of biology, but something we, like real-estate agents, can never take for granted.