

Comment

What's in a name?

Gregory A Petsko

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

Published: 26 March 2002

Genome Biology 2002, **3**(4):comment1005.1–1005.2

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2002/3/4/comment/1005>

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

Perform a literature search for articles concerning ARF, a small GTP-binding protein that is involved in vesicular transport, and you will find to your surprise that it is also a tumor suppressor gene product that binds to p53-DNA complexes. Except that it isn't. There are two completely different proteins with the same name. One ARF is ADP-ribosylation factor, the small GTPase. The other is derived from the name of a genetic locus, *INK4a-ARF*. I would like to believe that the cancer researchers who named their protein ARF in the mid-1990s were not aware that the name had already been used for over a decade for another important eukaryotic gene product; that they didn't deliberately try to appropriate the name for their own discovery and thus muddy the waters of protein annotation. Nothing, however, seems to engender more passion and provoke more quarrels than the matter of assigning names to things. Scientists defend their choices with the tenacity of a mother tiger protecting her cubs, with the result that the scientific literature is awash with names that range from the cute to the stupid. Duplications abound, and the information content of most gene names is nil. This was tolerable when the number of genes and proteins one had to worry about was manageably small, but in the era of genomics something has to be done about it. And whatever is done, it is clear to me that we cannot cede responsibility to the cell biologists.

When it comes to naming things, cell biologists seem to have about as much imagination as the American actor/screenplay writer/director Sylvester Stallone, who came up with *Rocky*; *Rocky II*; *Rocky III*; *Rocky IV*; ... and then *Rambo*; *Rambo II*; *Rambo III*. (By the way, have you ever noticed that, with a few exceptions, the quality of a movie is usually inversely proportional to the number of jobs the star has in addition to acting?) Examples of the cell biologists' jaunty wit and evocative command of language include CD1, CD2, CD3, CD4 and so on - names that, at a glance, tell one everything about what the proteins do, don't they? - and my personal favorite, p53. I mean, what can you learn about function from a name like p53, except that the person who

thought it up obviously didn't have a clue what the function was at the time? It isn't even a good uninformative name; didn't it occur to this person that there might just be a few other proteins around with a molecular weight of about 53,000? What are the cell biologists going to do when they encounter those - call them p53 II, p53 III and so on?

My personal preference - indicative, no doubt, of middle-aged nostalgia - is for the old style of naming things, where the name actually told you something about the function of the protein. Haemoglobin has a haem in it. The HIV protease is a protease from the human immunodeficiency virus. Triosephosphate isomerase isomerizes triosephosphates. These names were assigned by biochemists and enzymologists, who didn't feel they had the right to name something until they had some idea what it did. Another, perhaps more whimsical, alternative is to assign names arbitrarily from human names, just as is done for hurricanes and typhoons. In such a scheme, we would just name the *INK4a-ARF* gene product "Fred". p53 could be "Mary" or perhaps "Fatima" (we don't want to restrict ourselves to Anglo-Saxon names). There's precedence for this approach, actually. Jack Peisach, a biophysicist at Albert Einstein College of Medicine in New York, named the copper-containing electron-transfer protein he discovered in 1967 stellacyanin after his wife, thereby causing generations of biochemists to be grateful that he hadn't married someone named Gertrude. But the whole business is too important to be left to my preferences, sensible and imaginative though they may be.

What's clearly needed is a new, international commission on gene-product naming. This commission must not, under any circumstance, be connected with any International Union of Pure and Applied Anything, for those are the cretins who force-fed us the SI units. If you are my age - which, from the way my students treat me, means that you have personal recollection of the Middle Ages and used to go on double dates with Charlemagne - you probably spend much of your time in fond yearning for the Ångstrom, the atmosphere, the

kilocalorie, and other units that actually suggest something. The Ångstrom is a wonderful unit for subcellular distances because the length of a carbon-hydrogen bond is almost exactly 1 Ångstrom, so when someone says that the thickness of the hydrophobic portion of a lipid bilayer is about 30 Ångstroms it immediately refers that thickness to something of the same size range that one can visualize. The atmosphere is a great unit for pressure, because if somebody says a pressure of 1000 atmospheres it immediately refers that pressure to something with which we are all familiar. The kilocalorie is a great unit for energy, because the amount of kinetic energy available at ordinary temperatures is about 1 kilocalorie, and the energy of most weakly-polar, noncovalent interactions in biology is also about 1 kilocalorie. But thanks to the SI gang, a group of mostly European nobodies who sit around all day with nothing better to do than to think of ways to make science even more jargon-laden and obscure than it already is, the Ångstrom has been replaced by 0.1 nanometers, the atmosphere by the easy-to-remember 101,325 Pascals, and the kilocalorie by 4.184 kilojoules, all of which are units that have no simple frame of reference whatsoever. (By the way, all of them were originally invented in France and two of them were named after Frenchmen; come to think of it, SI stands for Systeme International, which is French, so is this in fact some kind of linguistic revenge for English having given the French terms like *le traffic jam* and *le weekend*?) Not because of xenophobia, but simply as a way of resisting any further nomenclature hegemony, the SI crowd, who enforce the use of their system with great fervor, must not be allowed to get their hooks into the gene-naming game. A new and independent commission is essential.

This commission must have absolute authority to revoke stupid names - p53 ought to be the first to go, with H-Ras not far behind - and assign clever new ones that suggest what the protein actually does while being, if possible, entertaining as well. Thus p53 could be renamed, for example, Guardian, as it has been characterized as the "guardian of the genome". Some of these names might get quite long - I imagine H-Ras could end up as something like Ubiquitous Eukaryotic Protein that Binds and Hydrolyzes Guanine Nucleotides and Signals to Everything - but is that so bad? Many Spanish men have five or six names, and that doesn't seem to get in their way. To function as intended, the commission must be constituted very carefully. It can, and should, have advisory bodies of scientists, but no scientist should be allowed to sit on the commission except for whoever thought up the protein class 'chaperone' and a *Drosophila* geneticist (any community that can come up with gene names like *Son of Sevenless* and *Sonic Hedgehog* has shown it can be trusted). The remainder of the commission should be constituted as follows. One stand-up comedian. An advertising copy writer; it's true that we all hate them, but they do this for a living and they're good at it. Film writer and director George Lucas, because there has never been a better name for a villain than Darth Vader - clearly

this man has the right stuff for the job. And finally my mother. I know this last suggestion smacks of rank nepotism, but I like the name Gregory, she chose it, and when you write your own column you can suggest your mother. I believe these people would take the job suitably seriously and would provide us with an improved working vocabulary for post-genomic biology.