

An occasional column, in which Caveman and other troglodytes involved in cell science emerge to share their views on various aspects of life-science research. Messages for Mole can be left at mole@biologists.com. Any correspondence may be published in forthcoming issues.



More than enough cake

I've been thinking about infinity. Well, not infinity, but really, really big numbers. Infinity per se is just too big for my poor insectivore brain to comprehend. (What? You didn't know I was insectivorous? What did you think I was? *Rodent*?) Infinity isn't just too big, it's too weird. In an infinite universe of infinite desserts, you can eat an infinite amount of cake *and* have it too. I shouldn't have to do this, but the argument goes like this: number all the cakes from 1 to forever. Now only eat the infinite odd-numbered cakes; you still have an infinite quantity of even-numbered cakes to lick the frosting off. And don't even *start* about cutting one into infinite pieces.

No, like I say, I've been thinking about big numbers. And it started with a large but smaller-than-expected number: 40,000. That's the estimated upper limit on the number of genes we have. Okay, there are some problems with that, since we know that we can make thousands of antibody genes by insidious B-cell developmental trickery, and none of

these gets counted in the estimate, but still it seems that 40,000 is about right. And if it is, then it also seems that we can do an awful lot with just that number of genes, because that's the number (or less) we need to make people, or moles – or naked mole rats, lesser kudus, comb jellies, tree kangaroos, monarch butterflies, sea squirts, rhinoceros hornbills, golden tamarins, gorgons and timice. And if we buy into the anecdotal evidence, small variations in these genes can produce a taste for chocolate or grubs, a fear of spiders, or a penchant for wearing red socks.

So here's what I'm thinking. Why restrict ourselves to the genes that we've got in the gene world or the genes that we fiddle with, one base pair at a time? Why not just make *all* of them? Every gene that can be made. We set an upper limit on size, vary each base pair, and just produce not only every single gene that exists, but every single gene that *can* exist. It's a big number, really big, but it's not even close to infinity. We could put these in a really big store – a super store, a MOLE-cular super store. And here's the cool thing. Why this is a fundamentally fantastic

store: since we produce every gene that can exist, and then produce from them every protein that can exist, *some* of the proteins will be regulatory proteins (just by chance) that control the expression and functions of some of the other proteins, and among this huge (but not infinite) set, there will be one, or more than one, protein that will spontaneously select all of the ones we need to coordinate, organize, and catalyze the self-assembly of absolutely *any* sort of life form you might want to buy.

See what I'm getting so excited about? A self-organizing, very large (but not infinite) pipeline for the production of life-to-order. The MOLE-cular superstore has it all, by *definition* (catchy, huh?). Every DNA sequence, and as a consequence, every protein and, as a further consequence, every enzyme pathway that ever was, will be, or *can* be, with a self-generating machinery to make it all work, just because it's really, really big. Hurry, operators are standing by, because the offer to invest in this *now* ends midnight tomorrow.

Okay, I know, this is really stupid.

We're all pretty giddy these days about some recent innovations in biology and

chemistry – sequenced genomes, DNA arrays, proteomic arrays, protein interactisomes, combinatorial drug libraries, programs that predict structure and function – real power. And somewhere along the line, some of us have gotten the idea that this makes science easier. We don't have to do the hard stuff, we just shop in the MOLE-cular superstore and, because it's so big, the information will self-assemble and just leap out at us – probably in ready-to-publish/patent/utilize/cure form.

It is true that there is an explosion of information, but some of us are already finding out what we *should* have known all along: this makes our work *harder*! Identifying every gene that is expressed in every cell in the body at every moment of development of any organism is suddenly do-able. But, like building my superstore, we can't expect the information to self-select, self-assemble, and present itself to us as understanding. We're falling into the trap that obtaining information is the same as understanding it. But the truth is we've got to figure out what to *do* with it. Four-dimensional pharmacogenomic protein interaction arrays pulsate with information, but it is only one part information and a very large (but not

infinite) amount of junk. Many of us (not you or me, of course, but many) seem to be waiting for it just to sort itself out.

One answer seems to be to build more machines to analyze the information for us and tell us what's important, but that's another trap (if the information isn't enough, let's just turn it into *more* information). But we know where the real answer lies: an old fashioned answer. We have to frame hypotheses, explore them, refine them, and confirm them, do experiments (yes, with controls), and others have to examine those results and ask if they're correct. Let's not forget, in our enthusiasm, that this is the way we've learnt an awful lot in the last two hundred years. Much of it is indeed easier (even doing the experiments, sometimes), but asking the good questions has gotten hugely more difficult. Too bad – nobody ever said that science was easy – well, certainly not *me*.

So, we *can't* have our cake and eat it too. But who would want too? There's more than enough to go around.

Mole

Journal of Cell Science 117, 975-976
Published by The Company of Biologists 2004
doi:10.1242/jcs.01012

Letters

JCS welcomes correspondence provoked by articles in all sections of the journal. Responses to articles in the Sticky Wicket section should be sent directly to Caveman (email: caveman@biologists.com). Correspondence relating to Research Articles, Commentaries and Cell Science at a Glance should be addressed to the Executive Editor and sent to

***Journal of Cell Science,*
140 Cowley Rd,
Cambridge,
CB4 0DL,
UK.**