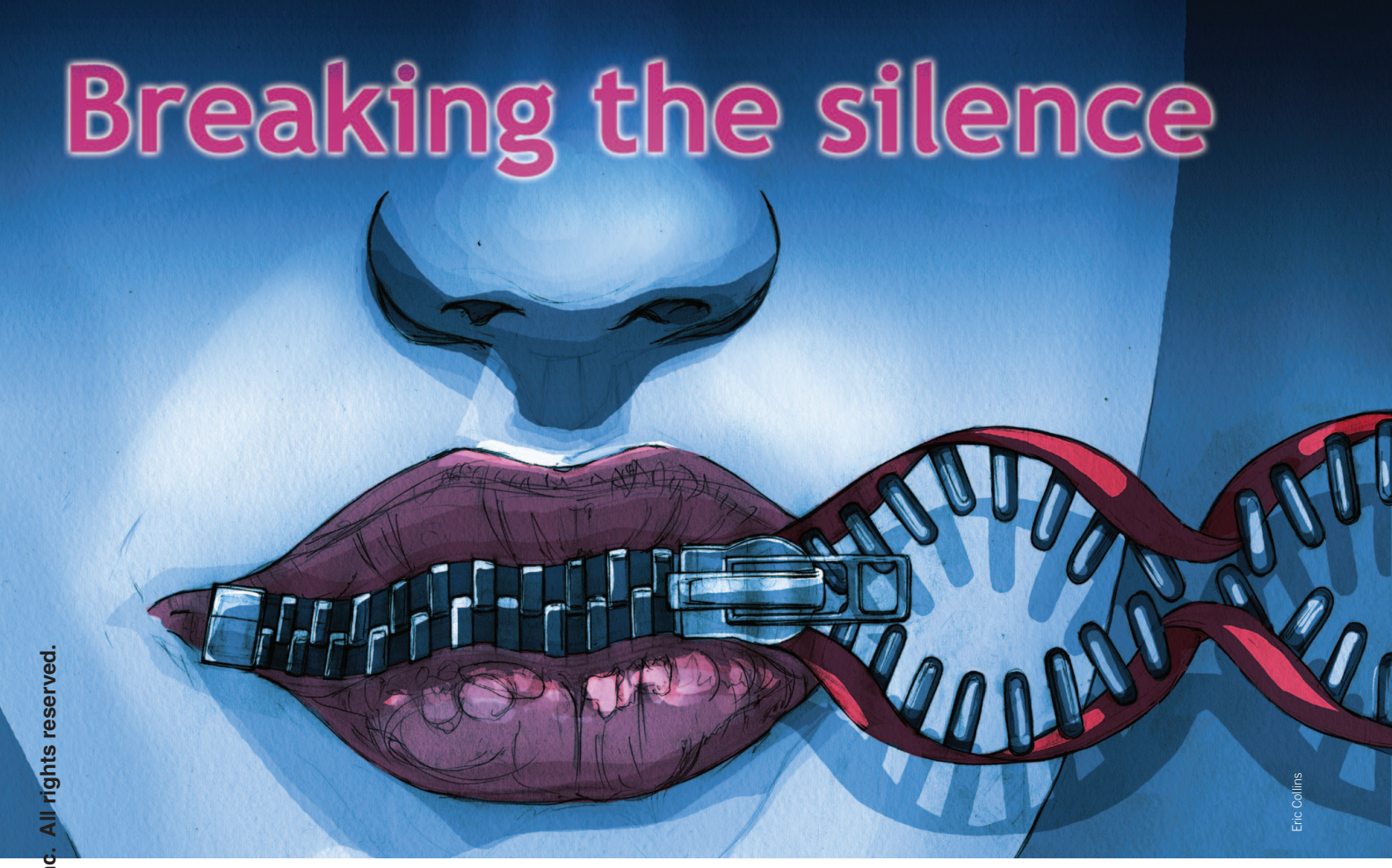
# Breaking the silence



Eric Collins



Scientists had long assumed that any genetic mutation that does not alter a protein sequence should have no impact on human health. But recent research has shown that such synonymous DNA changes can trigger disease in a number of ways. **Alla Katsnelson** talks to scientists and biotech companies who are speaking up about 'silent' mutations.

It all started with an expression problem. Michael Gottesman and his lab members at the US National Cancer Institute in Bethesda, Maryland were studying a membrane protein involved in drug metabolism called P-glycoprotein to understand why some people develop resistance to chemotherapy during cancer treatment. But when the scientists tried to express large quantities of the protein in bacterial cells, they hit a wall.

"It was a real mess," Gottesman recalls. "We couldn't do it."

The genetic code is read in triplets called codons, 64 of them representing just 20 amino acids. That means there is more than one codon for each amino acid, and different organisms preferentially use certain codons to make translation faster. One standard trick for boosting the expression of human genes in other organisms is to swap around nucleotides to get the DNA triplets most often used by the host's cellular machinery. But a colleague of Gottesman's suggested a different tack: as proteins elongate, the translation process needs to slow down and speed up to achieve proper folding, and perhaps the distribution of frequent and rare codons might control that rhythm.

The idea got Gottesman thinking about a niggling problem. The gene that encodes P-glycoprotein, called multidrug resistance 1 (*MDR1*), has about 50 single nucleotide polymorphisms, a handful of which are located in the coding region but at a position where they don't affect the protein's amino acid sequence. One, for example, in exon 26 of this 209-kilobase-long gene switches an ATC codon to ATT, both of which encode the amino acid isoleucine. Scientists routinely assume that such 'silent' or 'synonymous' mutations don't affect the protein's function, but clinical data clearly showed that people carrying these mutations metabolize drugs differently. "We were trying to think of how it could be that these synonymous mutations caused these changes," Gottesman says. Maybe, he thought, they were meddling with the rhythm, thereby changing the protein produced.

The researchers then expressed the ATT codon along with two other naturally occurring polymorphisms and saw that the expression levels of messenger RNA (mRNA) and protein remained the same, but the protein's activity was altered. Just as Gottesman had hypothesized, the evidence pointed to a shape shift in the resultant P-glycoprotein, caused by altered timing of translation.

The findings, published online in *Science* in late 2006 (ref. 1), weren't the only report of this type of mutation at work. A paper published at the same time in the same journal reported that synonymous polymorphisms in a gene encoding a protein called catechol-O-methyltransferase, which modulates responsiveness to pain, affected the loops and turns that make up the structure of the gene's corresponding mRNA, and, with it, the level of protein expression[2]. The two studies were the first published examples of human genes in which naturally occurring mutations produced proteins with an unchanged amino acid sequence but clearly different functional effects on disease. And, in the five years since, the idea that such mutations can have dramatic and far-reaching effects is beginning to take off.

Upwards of 50 disorders—including depression, schizophrenia, multiple cancers, cystic fibrosis and Crohn's disease—have now been linked to synonymous mutations. And although genome-wide association studies, the workhorse of medical genetics, have routinely excluded synonymous polymorphisms,

researchers are conceding the need to take a closer look. In one recent inspection of more than 2,000 human genome studies, for example, a team from Stanford University School of Medicine in California found that synonymous mutations were just as likely as nonsynonymous ones to play a part in disease mechanisms[3].

"Five years ago, you just wrote off silent mutations," says Zuben Sauna, an author on the *MRD1* paper now at the Laboratory of Hemostasis of the US Food and Drug Administration (FDA) in Bethesda. "But a critical mass of literature has now emerged that we should start looking at it more seriously—both from the disease point of view and in generating biologics."

It's true that manipulating synonymous codons to produce medicines such as therapeutic antibodies, vaccines and gene therapies has become routine in biotechnology. By some estimates, in fact, therapeutic proteins are sometimes altered by as much as 80% from their native form. If synonymous mutations can and do affect protein conformation and function, then companies better take a closer look to make sure they know what they're doing, Sauna says.

## Synonymous with disease

Hints that not all codons are created equal began to emerge roughly 40 years ago, not too long after researchers developed the first techniques for sequencing DNA. Evolutionary geneticists noted that organisms from viruses and fruit flies to rabbits and humans favored certain codons over others for specifying particular amino acids, especially in highly expressed genes. And these sites accumulated fewer mutations compared to other areas of the genome. Somehow, the frequent codons seemed to be important for highly expressed proteins—perhaps by making the process faster or more accurate. "It appears that evolution is saying, 'Please do not monkey with these sites; they are optimized for some reason,'" says Allan Drummond, an evolutionary cell biologist at the University of Chicago.

Doubts lingered, though, about whether this so-called 'codon bias' really mattered in mammals. Many argued that the evolutionary pressure to maintain the bias would be much weaker, owing to the smaller population sizes typical of mammals compared to the simpler model organisms studied in the laboratory. But in 2005, Laurence Hurst and his colleagues at the University of Bath in the UK found signatures of selection at synonymous sites throughout the genomes of mice, chimpanzees and humans, particularly at points marking where exons get spliced together[4]. In all, Hurst estimates that around 10% of synonymous mutations have a big enough benefit in mammals that they are under strong selective pressure. "In retrospect," says Hurst, "it's actually rather obvious—not just that these things happen, but that they happen rather regularly."

The fact that selection picks on these regions suggests that splicing changes could be the main mechanism by which synonymous mutations cause disease. Indeed, in the majority of disease associations where a mechanism has been proposed, interference with a splice site is the leading culprit[5]. Hurst concedes, though, that it could simply be the easiest mechanism to identify.


"This codon-optimized gene was the cornerstone for the entire biotechnology revolution."
– Claes Gustafsson

Additional mechanisms abound, such as one identified earlier this year by researchers in France. A team led by Patrick Brest at the University of Nice showed that a synonymous mutation in the coding region of a gene called 'immunity-related GTPase family, M', which codes for a protein involved in removing intracellular bacteria, is actually a microRNA binding site. The mutation, in which a CTG codon is changed to a TTG (both code for leucine), is one of three polymorphisms in the gene that have been associated with Crohn's disease. In inflamed epithelial cells of the gut, the TTG form blocks the microRNA binding site; this, in turn, keeps gene expression 'on', which inhibits the cell's antibacterial activity and exacerbates the illness[6]. "This kind of phenomenon can occur in any disease," says Brest.

In another example, a group from the University of Alabama at Birmingham showed last year that a synonymous codon change found in the most common form of cystic fibrosis results in mRNA misfolding[7]. And then there's the scenario suggested by *MDR1*, in which a codon change may interfere with the pauses that characterize RNA passing through the ribosome, thereby changing how the growing amino acid chain folds. "At the moment I think we're in a phase in which we are discovering the major mechanisms by which synonymous mutations can be associated with disease," says Hurst. "And they are vastly more diverse than people thought."

Until recently, most studies have been largely observational, but new high-throughput technologies can help scientists assemble a complete picture. "For the first time, we can do very large scale systematic experiments that allow you to measure the effects of different synonymous choices either on the actual details of translation or on the fitness of the cell," says Joshua Plotkin, a mathematical biologist at the University of Pennsylvania in Philadelphia. Plotkin's team recently created 154 different variants of the gene coding for

**Actions speak louder:** Predicted mechanisms by which silent mutations cause disease.

| Class of human disease | Splicing | RNA structure | Rate of translation | Unknown |
|---|---|---|---|---|
| Vision | • | • | | • |
| Muscle | • | | | • |
| Kidney | • | • | | |
| Skin | | • | | |
| Lung | • | • | • | |
| Heart | | • | | |
| Blood | • | • | • | • |
| Bone | • | | | |
| Cancer | • | • | • | • |
| Immune | • | • | | |
| Diabetes | | • | | • |
| Neurological | • | • | | • |
| Longevity | • | • | | • |
| Liver | • | • | | • |

Source: *Nat. Rev. Genet.* **12**, 683–691 (2011).

the green fluorescent protein, each varying at one synonymous site. When the researchers expressed the genes in *Escherichia coli*, they saw a 250-fold range in protein expression levels that seemed to be determined not by codon bias but by whether the mutations caused changes in the secondary structure of mRNA[8]. For now, however, the only conclusion Plotkin is willing to draw is that the surface has only been scratched.

### The silent majority

Long before researchers knew about the wily ways of silent mutations, switching up synonymous codons became standard practice in industry. In 1977, biochemist Herbert Boyer, who had co-founded Genentech, the world's first biotech company, the previous year, used a bacteriophage to synthesize the human gene encoding the hormone somatostatin. Boyer knew the protein's amino acid sequence, but not the gene's nucleotide sequence, so he and his team strung together codons favored by *E. coli*, the organism in which they planned to express somatostatin[9]. Although the South San Francisco biotech never developed the protein into a commercial product, "this codon-optimized somatostatin gene was the cornerstone for the entire biotechnology revolution," says Claes Gustafsson, chief operating officer of DNA2.0, a gene synthesis company in Menlo Park, California.

As researchers got better at expressing genes in various cellular hosts and the cost of gene synthesis dropped, optimizing codons to improve expression became the norm. Yet the basic technique for optimizing proteins remains largely unchanged from Genentech's early days: recoding DNA so that the translation process uses the codons favored by highly expressed genes in the host organism.

Dorene Nielsen, director of business development for Blue Heron Biotechnology, the gene synthesis arm of the Rockville, Maryland–based gene product company OriGene Technologies, estimates that around 70% of her company's clients who need to express their proteins—mostly pharmaceutical or biotech firms studying the properties of protein drugs or diagnostics in their pipeline—optimize their genes in this way. The results, however, are notoriously hit or miss. Nielsen, Gustafsson and others agree on the problem: lots of people play around with silent positions in the genome, but nobody has derived a clear-cut system for predicting expression.

Slowly, a complex picture is beginning to form. DNA2.0 recently showed, by making many systematic mutations in a pair of genes, that protein expression in *E. coli* is predicted by certain biochemical properties of transfer RNA, which reads the mRNA message, not by the frequencies of overlapping codons[10]. Another study, by researchers at the German gene synthesis company GeneArt (now part of California's Life Technologies), characterized the effect of the company's complex optimization strategy on 50 different genes[11].

But, beyond getting optimization to work, some worry that fiddling with codons in therapeutic proteins could have unpredictable effects on people's health. "From our experiments now, we do not believe that you can do that to any protein and have the protein behave as it did in its native form," says Chava Kimchi-Sarfaty, who works with Sauna at the FDA. The changed form could cause immunogenicity, for example, which wouldn't be seen until late-stage clinical trials or even after approval. Already, she says, there are plenty of examples of small changes in manufacturing having dramatic effects on the quality and safety of biologic drugs, and there's no reason that protein alterations caused by synonymous codon changes wouldn't have similar effects. It's a theoretical concern, admits Jeremy Minshull, chief executive of DNA2.0. But he argues that there are so many variables being changed in therapeutic proteins that synonymous codons themselves are unlikely to pose special danger.


"We do not believe that you can optimize codons and have the protein behave as it did in its native form."
– Chava Kimchi-Sarfaty

The problem could be even greater for gene therapy, however, where optimizing expression has become routine in preclinical work and is beginning to trickle out into clinical trials. "We've reached a point where certain delivery systems have a reasonable margin of safety, so there's a new emphasis on optimizing the potency," notes Terence Flotte, dean of the University of Massachusetts Medical School in Worcester. Flotte's lab, for example, has a gene replacement therapy that uses an adeno-associated virus for delivery in phase 2 trials for a genetic form of emphysema. In its native state, the protein expresses at about 10% of its normal rate when administered to patients; Flotte is considering optimizing the gene sequence to try and bump it up to about 50%. The basic strategy is the same as for protein therapeutics—researchers change a handful of codons that are frequent in the human protein but rare in the viral vector. But because both the protein and the nucleic acid sequence are part of the therapy, he says, researchers will have to watch for unexpected effects of the codon changes on the protein that the vector is designed to pump out.

Meanwhile, Kimchi-Sarfaty, who led Gottesman's *MDR1* study, and Sauna are studying optimized versions of several genes that code for coagulation factors—some of them encoding biologic drugs already in clinical use—and characterizing how they differ from native proteins. But to really glean what's going on will require more data than the FDA alone can amass. Soon, they hope, companies developing biologics might submit their own observations about the effects of codon optimization to the FDA under the condition that the agency won't use the information against them when evaluating their drug applications. Such a voluntary submission process for exploratory data, originally designed for genomic and biomarker data, already exists, they note.

At the moment, companies developing recombinant therapies must verify that the DNA sequence designed by their scientists is the one that's producing their proteins, but they aren't required to note how different that is from the native genetic code. "We do not have any guidance with regard to the [DNA] sequence," Kimchi-Sarfaty notes. That's one piece of data that could be tracked by the system she is proposing. Such knowledge, in turn, could ultimately help define better strategies for optimization and possibly even make biologic drugs safer for people.

*Alla Katsnelson is a freelance science writer in New York.*


1. Kimchi-Sarfaty, C. *et al. Science* **315**, 525–528 (2007).
2. Nackley, A.G. *et al. Science* **314**, 1930–1933 (2006).
3. Chen, R., Davydov, E.V., Sirota, M. & Butte, A.J. *PLoS ONE* **5**, e13574 (2010).
4. Parmley, J.L., Chamary, J.V. & Hurst, L.D. *Mol. Biol. Evol.* **23**, 301–309 (2006).
5. Sauna, Z.E. & Kimchi-Sarfaty, C. *Nat. Rev. Genet.* **12**, 683–691 (2011).
6. Brest, P. *et al. Nat. Genet.* **43**, 242–245 (2011).
7. Bartoszewski, R.A. *et al. J Biol. Chem.* **285**, 28741–28748 (2010).
8. Kudla, G., Murray, A.W., Tollervey, D. & Plotkin, J.B. *Science*, **324**, 255–258 (2009).
9. Itakura, K. *et al. Science* **198**, 1056–1063 (1977).
10. Welch, M. *et al. PLoS ONE* **4**, e7002 (2009).
11. Fath, S. *et al. PLoS ONE* **6**, e17596 (2011).