responses. It is composed of a series of folded domains characteristic of those of the immunoglobulin (Ig) superfamily of proteins, as well as two unstructured regions (the PEVK and N2B regions). When a muscle extends, the induced stress first straightens the I-band from an initial coiled conformation, and then unravels the PEVK and N2B regions (Fig. 1). Once all the unstructured regions have extended, a few Ig domains might also unravel — an effect that dissipates energy and thus minimizes any damage to the protein caused by overstretching. When the muscle contracts, titin reverts to its tangled conformation.

Titin's capacity for reversible deformation is responsible for the passive elastic properties of the contractile system in muscle cells and for the reversible reshaping of these cells during contraction and extension. At low extensions, titin's high resilience reflects the mechanical responses of the PEVK and N2B regions, but at high extensions titin is tough and damps forces. The ability to make a swift transition between resilient and force-damping states is what makes titin such a crucial constituent of cardiac and skeletal muscle.

Lv *et al.*[1] used a creative approach to engineer artificial elastomeric proteins that mimic titin. Their proteins consisted of globular protein domains (GB1 proteins) that mimic titin's Ig regions, and repetitive amino-acid sequences derived from resilin that mimic the unstructured N2B regions. The small GB1 protein, derived from *Streptococcus* bacteria, is a highly effective molecular spring capable of reversible, rapid, high-fidelity refolding, with low mechanical fatigue over many stretching–relaxation cycles[8]. The authors constructed genetically modified *Escherichia coli* bacteria to express the GB1–resilin proteins, and then photochemically crosslinked the molecules to produce a gel-like material. The resulting GB1–resilin polyproteins are the first materials to incorporate folded, mechanically resistant globular domains as serial molecular springs.

The authors observed that, like titin, their polymers were highly resilient at low extensions because of the presence of resilin sequences, but that the resilience decreased with increasing extension to provide a force-damping response. Significantly, the materials rapidly recovered their resilience when the deforming forces were removed, reflecting the speed with which the non-covalent bonds responsible for the folding of GB1 domains re-form as the extended domains revert to their folded structure. Although efforts are under way to mimic the complex three-dimensional structure of proteins using non-biological polymers, synthetic elastomers, derived from the assembly of non-protein-based molecular springs, have yet to be produced with resilience that varies depending on how much the material is stretched.

Lv and colleagues' material[1] is certainly impressive, but is it a true muscle mimic? Muscles are complex molecular machines, in which several components are assembled into well-ordered structures capable of converting a stimulus into motion. Titin is a major constituent of muscle, but a titin mimic alone does not reproduce all the properties of muscle — such as its tensile strength, or force-generating and force-sensing abilities. In the absence of a self-repair mechanism, protein-based materials are also inherently susceptible to biological degradative processes after implantation, which could release 'foreign' protein fragments into the host. For biomedical applications, such materials therefore need to be carefully assessed to ensure that no fragments cause adverse immune reactions. Future work will undoubtedly address these issues, leading to creative designs and fabrication techniques for assembling artificial muscle elements that reproducibly and repeatedly respond on command, perform work, and function well after surgical implantation.

As first surmised by Staudinger, life may be a mortal coil comprised of a diverse set of biopolymers. By demonstrating new ways to wind protein threads into unique conformations, Lv *et al.* and other investigators are crossing scientific frontiers into an undiscovered country of molecular machines, nanometre-scale devices and molecularly engineered tissues that possess tailored macroscopic properties. ■

Elliot L. Chaikof is in the Department of Surgery, Emory University School of Medicine, Atlanta, Georgia 30322, USA, and the School of Chemical and Biomolecular Engineering, and the Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, USA.
e-mail: echaiko@emory.edu

1. Lv, S. *et al. Nature* **465,** 69–73 (2010).
2. Urry, D. W., Haynes, B., Zhang, H., Harris, R. D. & Prasad, K. U. *Proc. Natl Acad. Sci. USA* **85,** 3407–3411 (1988).
3. Urry, D. W. *et al. Phil. Trans. R. Soc. Lond. B* **357,** 169–184 (2002).
4. Cappello, J. *et al. Biotechnol. Prog.* **6,** 198–202 (1990).
5. Rabotyagova, O. S., Cebe, P. & Kaplan, D. L. *Biomacromolecules* **10,** 229–236 (2009).
6. Elvin, C. M. *et al. Nature* **437,** 999–1002 (2005).
7. Kim, W. & Conticello, V. P. *J. Macromol. Sci. C* **47,** 93–119 (2007).
8. Cao, Y. & Li, H. *Nature Mater.* **6,** 109–114 (2007).

Competing financial interests: declared (see online article for details).

GENE REGULATION

Breaking the second genetic code

J. Ramón Tejedor and Juan Valcárcel

Diverse messenger RNAs, and thus proteins, can be generated from a single piece of DNA. A computational approach is helping to uncover complex combinatorial rules by which specific gene instructions are selected.

At face value, it all sounds simple: DNA makes RNA, which then makes protein. But the reality is much more complex. For instance, depending on what further processing the transcribed messenger RNA sequence undergoes before being translated into a protein, it could code for different proteins. In this issue, Barash *et al.*[1] (page 53) identify the combination of RNA features that determine the sites of mRNA processing — the splicing code.

Classical experiments in the 1960s deciphered the code by which triplets of nucleotides — the units that make up DNA and RNA — are translated into amino acids, the building blocks of proteins[2]. This code, initially deciphered in bacteria, is shared by all known forms of life. But the genetic instructions of complex organisms exhibit a counter-intuitive feature not shared by simpler genomes: nucleotide sequences coding for a protein (exons) are interrupted by other nucleotide regions that seem to hold no information (introns). This bizarre organization of genetic messages forces cells to remove introns from the precursor mRNA (pre-mRNA) and then splice together the exons to generate translatable instructions.

An advantage of this mechanism is that it allows different cells to choose alternative means of pre-mRNA splicing and thus generates diverse messages from a single gene. The variant mRNAs can then encode different proteins with distinct functions[3] (Fig. 1a, overleaf). Knowledge of how cells produce alternative mRNAs is essential to understanding the output of our genome and its regulation. However, predicting the alternatively spliced products of a gene in different tissues or under varying physiological conditions has proved difficult. Barash *et al.*[1] undertake an epic assault on this problem, providing considerable hope that the 'splicing code' is indeed breakable.

One difficulty with understanding alternative pre-mRNA splicing is that the selection of particular exons in mature mRNAs is determined not only by intron sequences adjacent to the exon boundaries, but also by a multitude of other sequence elements present in both exons and introns[4]. These auxiliary sequences are recognized by regulatory factors that assist or prevent the function of the spliceosome — the molecular machinery in charge of intron removal[5].

A second difficulty is that the effects of a particular sequence or factor can vary depending on its location relative to the intron–exon boundaries or other regulatory motifs[4]. The

NATURE|Vol 465|6 May 2010

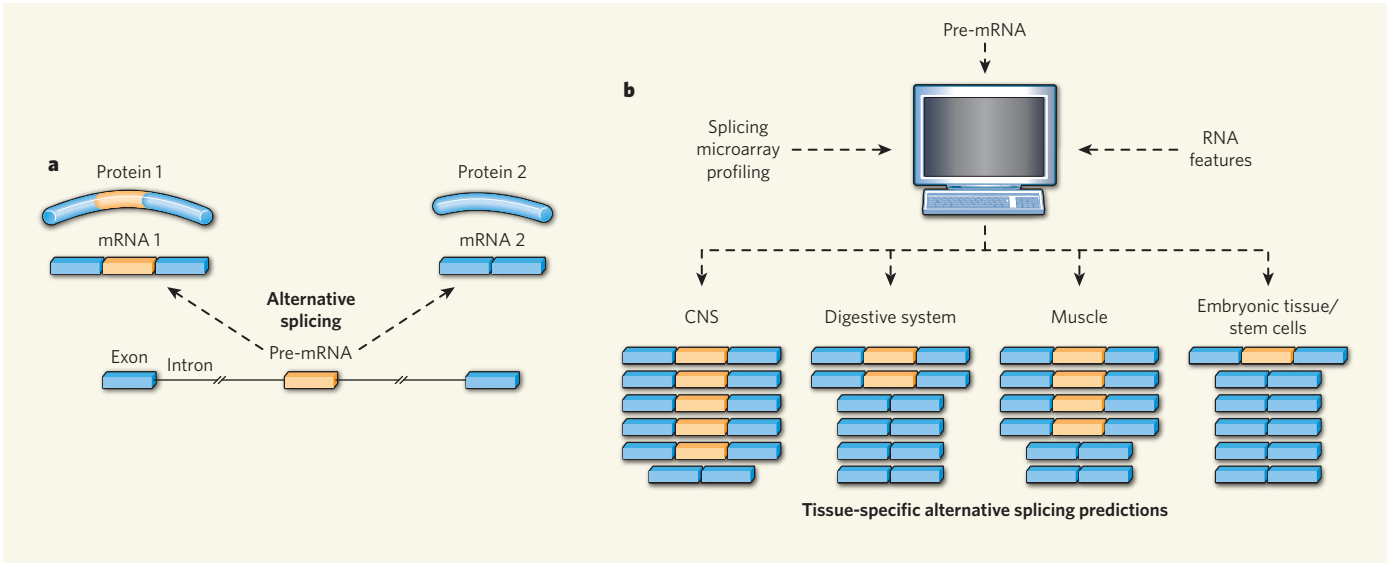

Figure 1 | Predicting pre-mRNA fate. a, Genomic DNA sequences are transcribed as messenger RNA precursors (pre-mRNA) containing exons and introns that can be processed by alternative pathways to generate different mRNAs encoding distinct proteins. **b**, Using data on alternative splicing obtained by microarray profiling of mRNAs from different tissues, as well as a compendium of regulatory sequences (RNA features), Barash *et al.*[1] identify combinations of features that can predict, for a given pre-mRNA, the ratio of alternatively spliced mRNAs in four different tissue types: central nervous system (CNS), digestive system, muscle and embryonic tissue/stem cells.

challenge, therefore, is to compute the algebra of a myriad of sequence motifs, and the mutual relationships between the regulatory factors that recognize them, to predict tissue-specific splicing.

To achieve this, Barash *et al.*[1] provided a computer with two types of information (Fig. 1b). The researchers gathered microarray data evaluating the ratio between inclusion and skipping of more than 3,000 alternatively spliced exons in four types of mouse tissue. They also took advantage of the collective knowledge generated by the splicing research community to compile thousands of RNA sequence features corresponding to known binding sites for regulatory factors, as well as sequence motifs enriched around alternatively spliced RNA regions, even if their cognate regulatory factors remain unknown. Moreover, they considered characteristics of the exon/intron organization, their evolutionary conservation, the folded structure of the RNA chain and the relationships among all these elements. The computer was then asked to identify the combination of features that could best explain the experimentally determined tissue-specific selection of exons.

Considering the complexity of the system, the approach achieved notable successes. It correctly identified alternative exons, and predicted their differential regulation between pairs of tissue types with considerable accuracy. The code identified features whose distribution and frequent co-occurrence with other sequence elements is associated with tissue-specific regulation. This allows reinterpretation of the function of previously defined regulatory motifs and suggests previously unknown properties of known regulators as well as unexpected functional links between them. For instance, the code inferred that the inclusion of exons that lead to truncated proteins is a common mechanism of gene-expression control during the transition between embryonic and adult tissues.

Despite these successes, however, Barash and colleagues' work[1] may be better seen as revealing the first piece of a much larger Rosetta Stone required to interpret the alternative messages of our genomes. The expected wave of massive data sets generated by high-throughput technologies[6] should soon provide further inputs for improving the code. These include identification *in vivo* of binding sites for regulatory proteins by techniques such as cross-linking/immunoprecipitation (CLIP), extensive description of mRNA variants by high-throughput sequencing, and functional characterization of regulators by RNA interference screens.

The code is likely to work in a cell-autonomous manner and, consequently, may need to account for more than 200 cell types in mammals. It will also have to deal with the extensive diversity of alternative-splicing patterns beyond simple decisions of single exon inclusion or skipping. The limited evolutionary conservation of alternative-splicing regulation (estimated to be around 20% between humans and mice) opens up the question of species-specific codes. Moreover, coupling between RNA processing and gene transcription influences alternative splicing, and recent data[7,8] implicate the packing of DNA with histone proteins and histone covalent modifications — the epigenetic code — in the regulation of splicing. The interplay between the histone and the splicing codes will therefore need to be accurately formulated in future approaches. The same applies to the still poorly understood influence of complex RNA structures on alternative splicing.

Deciphering the genetic code allowed the identification of protein-coding genes and thus provided a key conceptual framework for understanding genome organization. An important measurement of the value of Barash and co-workers' paper[1] will be its usefulness in interpreting the output of genes in genome-sequencing projects and in rationalizing changes in alternative splicing caused by natural sequence variation or underlying pathological conditions. Another key assessor of this[1] and other codes of post-transcriptional regulation[9] will be their amenability to converting large data sets of intriguing relationships between sequence motifs into testable hypotheses that will help to unravel the underlying mechanisms, the code's molecular fabric. ■

J. Ramón Tejedor and Juan Valcárcel are at the Centre de Regulació Genòmica and Universitat Pompeu Fabra, Barcelona, Spain. Juan Valcárcel is also at the Institució Catalana de Recerca i Estudis Avançats, Dr. Aiguader 88, 08003 Barcelona, Spain.
e-mail: juan.valcarcel@crg.es

1. Barash, Y. *et al. Nature* **465,** 53–59 (2010).
2. Nirenberg, M. *Trends Biochem. Sci.* **29,** 46–54 (2004).
3. Nilsen, T. W. & Graveley, B. R. *Nature* **463,** 457–463 (2010).
4. Chasin, L. A. *Adv. Exp. Med. Biol.* **623,** 85–106 (2007).
5. Chen, M. & Manley, J. L. *Nature Rev. Mol. Cell Biol.* **10,** 741–754 (2009).
6. Blencowe, B. J., Ahmad, S. & Lee, L. J. *Genes Dev.* **23,** 1379–1386 (2009).
7. Luco, R. F. *et al. Science* **327,** 996–1000 (2010).
8. Fox-Walsh, K. & Fu, X.-D. *Dev. Cell* **18,** 336–338 (2010).
9. Piqué, M., López, J. M., Foissac, S., Guigó, R. & Méndez, R. *Cell* **132,** 434–448 (2008).

Addendum
Gökhan S. Hotamisligil, a co-author of the News & Views article 'Metabolism: Host and microbes in a pickle' (*Nature* **464,** 1287–1288; 2010) declared a competing financial interest that was not noted in the print or PDF versions of the article. The declaration can be found in the full-text online version.