taining pre-mRNA. Although unconventional ideas can predicate important discoveries, the facts remain that translation, at least as of this writing, provides the only known means of recognizing nonsense codons and is confined to the cytoplasm. And, it is difficult to envision how the translational machinery, or some variation thereof, could completely assess a reading frame that is interrupted by introns, let alone in a way that could also influence the choice of splice sites. An alternative view of NAS is that nonsense codons alter exonic splicing enhancers (ESEs) in addition to the translational reading frame[6]. There are a few reports[10,11] that missense or silent mutations can cause NAS presumably, but not assuredly, by affecting either ESEs or their functional antitheses, exonic splicing silencers (ESSs). These led Liu *et al.*[1] to investigate the idea that disease-associated NAS of *BRCA1* RNA could reflect the ability of single point mutations to weaken ESEs.

ESE does it

ESEs are discrete but degenerate sequence motifs of approximately 6–8 nucleotides. They reside within exons that are both constitutively as well as alternatively spliced, and are required for efficient splicing of these exons. They are not necessarily purine rich, as was originally thought[12]. Each type of ESE binds a different SR protein (a member of a family of modular splicing factors that harbors one or more RNA recognition motifs and an arginine/serine-rich region). Of the four types of ESEs examined by Liu *et al.*, exon 18 of *BRCA1* was found to contain nine sequence matches, and the G→T nonsense mutation was found to disrupt the 5′ most of these[1]. The ability of Liu *et al.* to reproduce NAS using a mini-*BRCA1* gene and HeLa-cell nuclear extract active in splicing enabled them to analyze novel point mutations, including a missense mutation predicted to disrupt ESE function and a nonsense mutation that was not predicted to disrupt ESE function. They found that exon skipping correlates with ESE disruption, regardless of the nature of the point mutation. Replacement of the 5′-most ESE with a heptamer predicted to have ESE function or a heptamer predicted to lack ESE function led to the same conclusion, which argues against a mechanism for exon skipping that involves disrupting the secondary structure. Driving home the effect of ESEs on NAS, in particular, and splice-site usage in general, is the finding that more than half of 50 nonsense, missense and silent mutations that are known to cause exon-skipping of transcripts other than *BRCA1* reduced or eliminated at least one predicted ESE (ref. 6).

Pointing to the nucleus

As 15–20% of a randomized 18- or 20-bp sequence can substitute functionally for an ESE (ref. 12), the number of different types and prevalence of ESEs characterized so far are surely underestimates. And, although ESSs are less understood[13], they, together with ESEs, help to explain the observation that a significant fraction of the nearly 15% of point mutations known to cause human genetic diseases and result in defective splicing are located within exons but are distinct from splice junctions and are not known to generate cryptic splice sites[6,14]. So, it no longer makes sense to predict how a point mutation associated with a disease affects gene expression based solely on the amino-acid change conferred by the affected codon. Because many ESEs and ESSs coincide with sequences that encode protein, it becomes important to analyze each point mutation for its effect on RNA processing.

What of the idea that NAS is a consequence of altered translational reading frame? Despite the fact that Liu *et al.* find no support for this possibility, the scientific community will undoubtedly remain abuzz with thoughts of nonsense codon recognition in the nucleus—if not for NAS, then for nucleus-associated NMD, which has been proposed to take place either during mRNA export using the cytoplasmic translational machine (there is some evidence that cytoplasmic ribosomes associate with the 5′ end of mRNA during export[3–5,7]) or in the nucleoplasm by a translation-like mechanism. □

1. Liu, H.-X., Cartegni, L., Zhang, M.Q. & Krainer, A.R. *Nature Genet.* **27**, 55–58 (2001).
2. Mazoyer, S. *et al. Am. J. Hum. Genet.* **62**, 713–715 (1998).
3. Maquat, L.E. *Am. J. Hum. Genet.* **59**, 279–286 (1996).
4. Hentze, M.W. & Kulozik, A.E. *Cell* **96**, 307–310 (1998).
5. Li, S. & Wilkinson, M.F. *Immunity* **8**, 135–141 (1998).
6. Valentine, C.R. *Mut. Res.* **411**, 87–117 (1998).
7. Maquat, L.E. in *Translational Control of Gene Expression* (eds. Sonenberg, N., Hershey, J.W.B. & Mathews, M.B.) (Cold Spring Harbor Press, Cold Spring Harbor, 2000).
8. Dietz, H.C. & Kendzior, R.J., Jr. *Nature Genet.* **8**, 183–188 (1994).
9. Gerasppe, A., Burger, L. & Pintel, D.J. *J. Biol. Chem.* **274**, 22452–22458 (1999).
10. D'Souza, I. *et al. Proc. Natl. Acad. Sci. USA* **96**, 5598–5603 (1999).
11. Vuillaumier-Barrot, S. *et al. Human Mut.* **14**, 543–544 (1999).
12. Blencowe, B.J. *Trends in Biol. Sci.* **25**, 106–110 (2000).
13. Fairbrother, W.G. & Chasin, L.A. *Mol. Cell. Biol.* **20**, 6816–6825 (2000).
14. Cooper, T.A. & Mattox, W. *Am. J. Hum. Genet.* **61**, 259–266 (1997).
15. Kan, J.L.C. & Green, M.R. *Genes Dev.* **13**, 462–471 (1999).

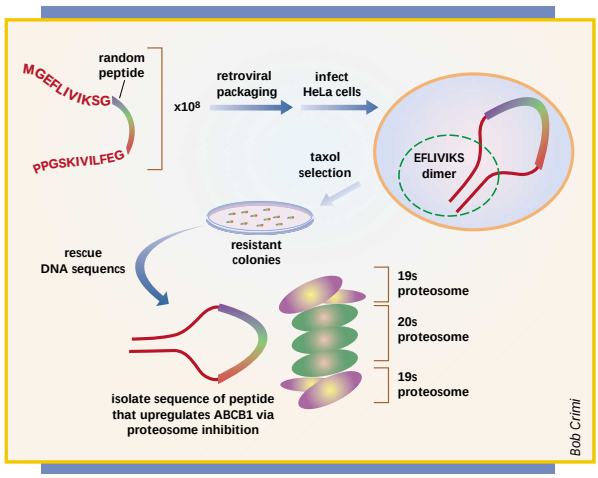
A peptide needle in a signaling haystack

Paul D. Roepe

Department of Chemistry and Program in Tumor Biology, Lombardi Cancer Center, Georgetown University, Washington, DC, USA.
e-mail: roepep@georgetown.edu

Expressing random peptides within cells has been hampered by technical problems, and so a simple method to express random free peptide libraries within mammalian cells comes as a welcome advance. It offers a powerful new way to approach the analysis of complex signal transduction cascades, as is illustrated in dissecting the emergence of tumor cell resistance to the chemotherapeutic drug taxol.

Professor Gobind Khorana once told me that a particularly fruitful way to approach the interface of chemistry and biology was to "develop clever ways of making and analyzing many different biomolecular structures." But sound though this advice may be, it is more difficult to heed as collective knowledge accumulates. Some biopolymers can be randomly synthesized and expressed within the cell with relative ease; others cannot. Twenty years ago, the technology for such investigations was relatively new, whereas today, construction and expression of random mutant proteins (and nucleic acid polymers) is relatively routine. One notable exception, however, is the construction of randomly generated peptide sequences, which historically have been technically difficult to produce. More versatile and reliable manipulation of intracellular expression of some naturally occurring peptides (and mutants of these)



Peptide production. DNA sequences encoding random 18-residue peptide sequences are flanked by codons for the SKIVILF dimerization sequence. After packaging of the sequences into retroviral particles, cells were infected. The presence of the dimerization domains presumably leads to more stable peptide expression, through the formation of protease-resistant tertiary structures such as that depicted. Higher-order structures, however, are also possible (that is, dimers of the peptide). After selecting for infected cells (selected in this instance, for resistance to taxol), the introduced sequence expressed is rescued and subcloned into an inducible expression system for further study in transfectants. By means of this method, Xu *et al.*[1] isolate an extremely interesting peptide sequence that is found to both increase the level of human *ABCB1* mRNA and inhibit the 20S proteosome complex, indicating the peptide upregulates ABCB1 through inhibiting the degradation of a transcription factor, whose identity is not yet known.

would be extremely useful. A powerful new tool for the intracellular expression of randomly generated peptide sequences is reported on page 23 of this issue by Xu *et al.*[1]. They describe the first successful screening of a random free peptide library generated inside living mammalian cells. And, as an interesting demonstration of the value of this new tool, they identify a role not previously ascribed to specific proteosome subunits in the drug-induced overexpression of the human multidrug resistance gene (*ABCB1*, also known as *MDR1*), which is associated with laboratory models of chemotherapeutic drug resistance.

Panning for gold

Random peptide libraries have been available for some time, and several approaches exist for teasing out a specific sequence of interest through its expression on the cell surface or through the use of phages (in the case of 'peptide display'). These approaches typically involve fusion of randomly generated DNA that encodes a random peptide sequence to a gene that encodes a cell surface (or phage surface) protein, followed by panning for the displayed fusion protein by immunological or chemical methods.

However, expression of random free peptide libraries within mammalian cells has been notoriously difficult to achieve because of promiscuous and robust intracellular protease activities and other problems. For example, the yeast two-hybrid technique can be used to search for protein–protein interactions within an intracellular environment, but modifying it with the aim of identifying small peptide motifs that bind to a given protein target typically results in high background. In addition, this method does not pivot on the physiology mediated by the protein–protein interaction at endogenous levels to document the interaction, but rather on induction of a reporter gene. A few previously successful attempts at expressing random peptides within bacterial cells have involved fusion of the library to a carrier protein[2] that provides a scaffold that stabilizes it and protects it from proteases. Although this method shows promise in certain applications, such as defining peptide inhibitors of essential soluble enzymes, the presence of a carrier protein (these are typically large) may make it difficult to screen for effects dependent on spatially restricted 'docking' in some cases. For example, docking of PDZ domains to peptide sequences at the carboxy-terminal end of proteins requires that the peptide be freely accessible and not shielded or buried within other protein domains.

Support from the flanks

To avoid problems of proteolysis and the carrier protein, Xu *et al.*[1] express a library that includes the sequence EFLIVIKS before and after the random peptide sequence. In the mid-1980s, Bodenmuller *et al.*[3] reported that the short peptide sequence, SKVILF, forms dimers that are stable in solution. So, flanking random peptides with this sequence presumably should protect and stabilize the expressed peptide. And, if so, a phenotype mediated by the peptide can be screened for. To achieve this goal, DNA encoding the protected random peptide sequences is first cloned into a retroviral vector (see figure), and then packaged into a retrovirus. The pool of viral particles (each particle presumably carrying nucleic acid encoding a unique random peptide sequence) are collected and used to infect the cells, thereby introducing the random sequence. After selecting the infected cells for a particular phenotype, the peptide sequence is rescued and packaged within a tetracycline-inducible system. This step facilitates convenient confirmation of the phenotype in a subsequent round of transfections, followed by detailed *in vitro* investigation of the biochemistry mediated by the peptide motif. The method is elegant in its simplicity.

Resisting taxol

Many examples of the usefulness of the approach could have been explored in its initial application, but Xu *et al.*[1] chose the important problem of resistance to taxol, a drug used in the treatment of breast tumors and other cancers. Deciphering the molecular basis of chemotherapeutic drug resistance, and using that information to design better second-line therapies remains a primary goal of the cancer pharmacology community. The discovery of *ABCB1* (ref. 4) served as a lightning rod for intensive research in this area; a deluge of reports on its function and hypothesized role in antitumor drug resistance followed over the next 10 years. And yet today, the idea that the overexpression of ABCB1 in and of itself has any clinical relevance to the resistance to chemotherapeutic drugs remains highly questionable. Nonetheless, the biochemistry mediated by this protein, as well as the mechanism that controls its overexpression, remain important issues with many implications for the cancer clinic.

Using the peptide library approach, Xu *et al.*[1] have generated data that may help resolve the controversial proposal that overexpression of ABCB1 protein is linked to taxol resistance. They also provide evidence that degradation of a transcription factor mediated by the proteosome can have a significant role in modulating the expression of *ABCB1*. This was not previously recognized as a major contribution to *ABCB1* overexpression, and the result may help to explain why so many diverse signals (for example, see ref. 5) induce substantial overexpression of ABCB1 in many different cell types.

The major source of confusion about ABCB1—including its role in the clinic, its molecular mechanism of function, and regulation of its transcription upon diverse stimuli—stems from the variety of ways in which drug-resistant, laboratory cell lines overexpressing ABCB1 have been generated. For example, if overexpression of ABCB1 in a cell selected for resistance to the anthracycline doxorubicin is associated with resistance to taxol, how is it possible to discern whether the association is direct or

a consequence of other events (for example, altered lipid metabolism or apoptotic signaling) caused by selection with the toxic substance? Few studies have stably overexpressed ABCB1 to high levels without any exposure to chemotherapeutics, and then analyzed the tumor cell transfectants for drug resistance so that contributions of ABCB1 to complex drug resistance phenotypes could be isolated[6].

By identifying a peptide motif that can up regulate *ABCB1* mRNA levels by inhibiting proteosome degradation of a relevant transcription factor, Xu *et al.*[1] have serendipitously discovered an additional tool for exploring ABCB1 regulation. But the strategy they use should prove extremely valuable in addressing many additional questions. Questions that even Professor Khorana might find challenging. □

1. Xu, X. *et al*. *Nature Genet.* **27**, 23–29 (2001).
2. Blum, J.H., Dove, S.L., Hochschild, A. & Mekalanos, J.J. *Proc. Natl. Acad. Sci. USA* **97**, 2241–2246 (2000).
3. Bodenmuller, H., Schilling, E., Zachmann, B. & Schaller, H. *EMBO J.* **5**, 1825–1829 (1986).
4. Chen, C.-j. *et al. Cell* **47**, 381–389 (1986).
5. Wei, L.Y. & Roepe, P.D. *Biochemistry* **33**, 7229–7238 (1994).
6. Hoffman, M.M., Wei, L.-Y. & Roepe, P.D. *J. Gen. Physiol.* **108**, 295–313 (1996).

Fish × 3

Geoff Duyk & Karin Schmitt

Exelixis Inc., 170 Harbor Way, South San Francisco, California 94083, USA. e-mail: duyk@exelixis.com

Researchers will be able to gain access to the complete zebrafish genome sequence within less than two years, accelerating gene identification of existing mutants and opening up new possibilities for genome-wide reverse genetics tools. Here, we discuss the context and promise of the zebrafish genome project.

It has now been decided that the Sanger Centre, with support from the Wellcome Trust, will lead the effort to establish the complete genomic sequence of the zebrafish (*Danio rerio*). At the end of October, representatives from the zebrafish community gathered at the Wellcome Trust Genome Campus* to discuss the sequencing strategy, review available resources and, most importantly, embrace the era of the genome.

Why the zebrafish? First, it represents an excellent model of vertebrate gene function, as has been demonstrated by numerous papers indicating similar or identical function of orthologous genes. Second, it is easy to handle and inexpensive to house. And third, the availability of well-established forward and reverse genetic tools and an impressive extant mutant resource render it an attractive candidate. It is little short of extraordinary how the zebrafish community has grown since the first experiments carried out by George Streisinger at the University of Oregon. Despite its allure, it should be noted that *Danio rerio* is not the only fish in the genomics sea; the Joint Genome Institute (Walnut Creek, CA) plans to deliver a draft of the puffer fish *Fugu rubripes* genome by 2001 and efforts are ongoing at Genoscope (France) to obtain partial sequence information of a related puffer fish, *Tetraodon nigroviridis*.

Sequencing of the zebrafish genome is to begin immediately, with the finished sequence expected by the end of 2002, although many details of the project have

Photo courtesy of John Postlethwait.

yet to be worked out. To begin with, the project will rely on a medium-density, whole-genome shotgun coverage of the genome, based on a mixture of small-insert clone libraries. This will be followed by directed finishing using bacterial artificial chromosomes. The strategy remains untested for larger genomes, but is similar to ongoing efforts dedicated towards the completion of the mouse genome (which is 2.9×10^9 bp; that of the zebrafish is 1.7×10^9 bp). Based on this experience, one may safely predict that one of the greatest challenges will be to establish methods—assembly algorithms—that facilitate ordering of sequencing contigs. The suggestion[1] of a comparatively recent duplication of the zebrafish genome may receive an additional boost (or kick) as the structure of the genome emerges. It should be noted, however, that such a duplication would represent an additional challenge to sequence assembly and annotation. The availability of sequence from other vertebrates may assist in this task and result in an improved initial product accomplished with a lower density of sequence coverage.

Indeed, the zebrafish genome project offers many opportunities for testing approaches to determine optimal genome coverage, integration of available genetic, physical and radiation-hybrid mapping data and the generation of new single-nucleotide polymorphisms (SNPs) by sample sequencing different zebrafish 'strains'—although the acquisition of SNPs is not a part of the initial project. The impetus for its addition is high; whereas zebrafish researchers enjoy a reasonably dense microsatellite map, recent experience from the human genetics community and advances in technology indicate that SNP genotyping is easier to automate and ultimately less expensive than typing with microsatellites. There is no reason to expect that the strategies used to complete the sequence of the human or the mouse genome should remain static; as these evolve, so too will the zebrafish project.

Also discussed at the meeting was the recently reported method of selectively repressing gene function by injecting zebrafish embryos with modified oligonucleotides (morpholinos), which hybridize the 5′ UTR sequence of the targeted gene[2]. These initial successes open up many new possibilities for carrying out systematic "reverse genetics" experiments in zebrafish. And, although all of us will be searching the sequence data as it is being released, a careful annotation of the nearly assembled sequence is necessary to provide information on contiguous gene space. How this will be done is not yet known, but collective approaches in the mouse and fly community have generated a lot of positive feed-

**Zebrafish Workshop, The Sanger Centre, Hinxton, UK; 26–27 October 2000.*