however, such as natural products derived from slow-growing plants and having complex, undefined biosynthetic pathways, plant cell culture will remain the most viable option.

The work of Lee *et al.*[3] marks an important departure from traditional plant cell culture. Instead of culturing heterogeneous mixtures of dedifferentiated cells, they isolated cells derived from vascular cambium and propagated them in solution. When exposed to the appropriate ratio of growth regulators, explants from most plant organs can be induced to dedifferentiate to form so-called callus cultures, which can be transferred to liquid media and disaggregated into single cells (**Fig. 1**). The resulting suspension cultures are amenable to bioprocessing technologies used for large-scale mammalian and microbial cultures. However, the starting cell population from the plant organ is a mixture of specialized cell types, which vary in cell cycle participation[8] and other properties, and the cell types that remain after extended culture in the dedifferentiated state are also probably heterogeneous, contributing to the instability of many culture properties.

In contrast, Lee *et al.*[3] bypass the dedifferentiation step by isolating undifferentiated CMCs from a variety of species, including *T. cuspidata*, *Panax ginseng* and *Ginkgo biloba*. The procedure (**Fig. 1**) is simple and rapid, with CMCs isolated from explants within a month. The authors focus on characterizing the potential of a cell line derived from *T. cuspidata* cambium cells for producing paclitaxel and related taxoids. The *T. cuspidata* CMC cultures are distinct from DDC cultures derived from needles and embryos on the basis of morphological characteristics, increased ability to differentiate, hypersensitivity to radiation and comparison of molecular signatures. A Gene Ontology analysis of the >500 differentially regulated genes confirms the upregulation of several genes known to be overexpressed in cambial cells. The cultured CMCs can thus be considered innately undifferentiated and inherently distinct from the dedifferentiated cells typically used for plant cell tissue culture.

The performance of CMCs as an applied bioprocessing technology is shown to be far superior to that of DDC cultures established at the same time. Growth rates and paclitaxel production are improved at both the laboratory scale (125 ml) and pilot scale (up to 20 liters), aggregate size is reduced, and variability in growth measured over repeated subculture cycles is markedly decreased. The authors also demonstrate the feasibility of using both stirred-tank and air-lift bioreactor designs for promoting growth and paclitaxel synthesis with CMCs, suggesting flexibility when designing a large-scale process. Although the paclitaxel yields from CMCs are similar to some values reported elsewhere for dedifferentiated cells, they are lower than the maximum values reported[9]. However, Lee *et al.*[3] did not attempt extensive process engineering optimization, and higher yields seem possible. The utility of this new approach is also demonstrated for the production of ginsenosides from *P. ginseng* cultures.

Some secondary metabolites accumulate to much higher extents in differentiated organ cultures (e.g., shoot or root). The most notable examples are the vinca alkaloids, produced by means of *Catharanthus roseus* hairy root cultures[10]. The benefits of using CMC cultures for the synthesis of such products will have to be evaluated.

A reliable, cost-effective supply of natural products for use as pharmaceuticals, fragrances, dyes and insecticides remains a major challenge for many systems. Plant cell tissue culture has been limited by inconsistent performance and the economic constraints associated with slow growth and low product yield. Compared with mammalian cell culture, plant cell culture has required batch times of months rather than weeks and has reached product titers of mg/l rather than g/l. Additional problems include cell aggregation, susceptibility to shearing, variability in growth and profusion of necrotic cells. CMCs appear to enhance cell culture performance in all of these areas, and notably do not require selection of specific cells and aggregates for consistent growth over repeated subcultures, thereby minimizing maintenance requirements. These improvements yield cell cultures that are substantially closer to mammalian cell cultures with regard to large-scale process considerations. CMC-based strategies should therefore facilitate the development of economically viable plant cell tissue culture processes for many natural products.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Mann, J. *Nat. Rev. Cancer* **2**, 143–148 (2002).
2. Koehn, F.E. & Carter, G.T. *Nat. Rev. Drug Discov.* **4**, 206–220 (2005).
3. Lee, E.-K. *et al. Nat. Biotechnol.* **28**, 1213–1217 (2010).
4. Roberts, S.C. *Nat. Chem. Biol.* **3**, 387–395 (2007).
5. Kolewe, M.E., Gaurav, V. & Roberts, S.C. *Mol. Pharm.* **5**, 243–256 (2008).
6. Ajikumar, P.K. *et al. Science* **330**, 70–74 (2010).
7. Keasling, J.D. *ACS Chem. Biol.* **3**, 64–76 (2008).
8. Naill, M.C. & Roberts, S.C. *Biotechnol. Bioeng.* **90**, 491–500 (2005).
9. Bringi, V., Kadkade, P.G., Prince, C.L. & Roach, B.L. US patent 7264951 (2007).
10. Leonard, E., Runguphan, W., O'Connor, S. & Prather, K.J. *Nat. Chem. Biol.* **5**, 292–300 (2009).

Making antibodies from scratch

J Christopher Love

Synthesis and screening of a small library of antibody fragments yields promising hits.

Screening large combinatorial libraries is standard procedure in small-molecule drug discovery but is not generally used to discover therapeutic monoclonal antibodies. The main obstacle is the prohibitive cost of library synthesis, as it has been believed that upward of 100 million unique antibodies would be needed. In this issue, Mao *et al.*[1] overturn this conventional wisdom by showing that a library of only ~10,000 antibody fragments, designed *in silico* and synthesized *de novo*, can yield new leads for several targets. Their results suggest that rational approaches for determining the relationships between the sequences and activities of a library of molecules can aid the search for new therapeutic antibodies.

Monoclonal antibodies are now widely used to treat a range of diseases, particularly cancers and autoimmune disorders. More than 20 are approved by the US Food and Drug Administration, generating over $15 billion in revenue annually, and hundreds are in clinical trials[2]. The insatiable demand for new therapeutic antibodies has spurred efforts to improve the efficiency of the processes used to discover them. The most common method used in industry relies on creating libraries of hybridomas from immunized (transgenic) mice, screening them by limiting serial dilution, and applying molecular engineering to humanize the leads or to improve their affinities and specificities. Advances in automated screening and liquid handling have improved this technique, but it remains arduous and costly.

J. Christopher Love is in the Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. e-mail: clove@mit.edu

The ability to clone antibody genes by PCR has led to a second approach in which *in vitro* libraries of recombinant antibodies are displayed on the surfaces of organisms such as phage, yeast or bacteria, and the highest-affinity binders are isolated by selection[3]. Such libraries of antibodies (or fragments of their binding regions) are usually designed to maximize their size and thus their diversity. This strategy has been motivated by (i) the theoretical diversity of natural antibody repertoires, estimated at $>10^{10}$ based on the number of unique combinations of gene segments possible and subsequent additions and deletions[4], (ii) the breadth of diversity feasible for libraries of genes in suitable hosts (10^7–10^{12}) (ref. 3) and (iii) the general unimportance of library size in selection-based strategies. Although displayed recombinant libraries are now widely used for discovery, selections typically rely on rounds of enrichment based only on affinities. It is not clear, however, that it is always desirable to identify antibodies with the highest affinities for a target[5,6].

Natural antibody repertoires may in fact be smaller than the predicted diversity of 10^{10}. Mice whose antibody repertoires are constrained by a single gene encoding the variable region of the heavy chain are still able to produce specific antibodies to most antigens[7]. Moreover, high-throughput sequencing of the variable genes expressed by zebrafish B cells has shown that only a relatively small percentage of the possible combinations of genes account for a large fraction of the repertoire in an individual[8]. Similar analysis of the diversity in circulating human B cells has indicated that the numbers of unique clones may be as low as 10^6 (ref. 9). These findings, among others, suggest that relatively small libraries could yield interesting monoclonal antibodies.

Mao *et al.*[1] have now shown this idea to be true experimentally. They computationally designed a combinatorial library of heavy and light chain antibody germline genes that represented a diverse structural repertoire and then *de novo* synthesized DNA encoding the genes (**Fig. 1**). Antibody fragments were constructed by transforming *Escherichia coli* with heavy and light chain genes, and then purifying the expressed protein in individual wells of 96-well microtiter plates. This process yielded a defined, spatially addressable library of ~10,000 members that could be screened using conventional high-throughput screening tools—that is, the sequence of every candidate molecule is known beforehand, unlike in hybridoma- or display-based methods. Remarkably, the small library yielded 85 hits for 7 out of 9 antigens screened in parallel, with affinities ranging from ~0.5 μM to 100 μM.

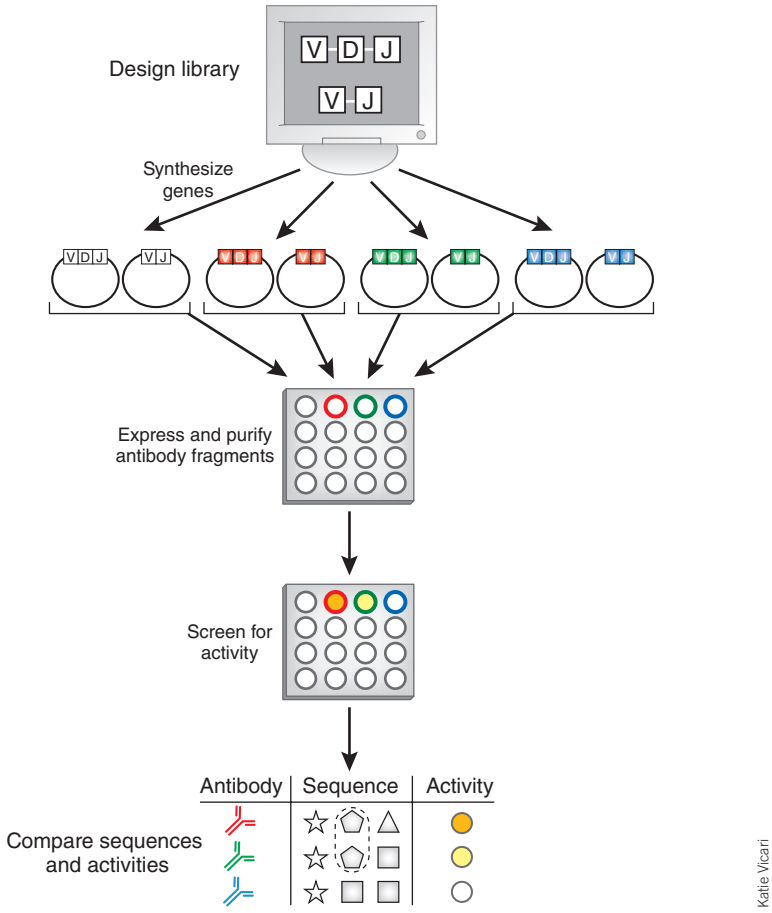

Figure 1 Strategy for the design, synthesis and screening of a spatially addressable library of antibody fragments. Members of the library are designed *in silico* by defining specific combinations of germline gene segments for the variable regions of the heavy and light chains of an antibody. Each pair of genes encoding a unique combination of heavy and light chains is synthesized *de novo* and used to transform *E. coli* in a known location in a microtiter plate. After parallel expression and purification, the antibody fragments are screened for specificity to nine unique antigens. Information about binding for each member in the library can then be related to their sequences to highlight key positions in the sequences that confer specificity. These sites become targets for subsequent modifications to alter the antibody's specificity, affinity or biological activity.

Because the sequence of each member in the library was known *a priori*, these hits could be compared to other members in the library having similar sequences but different affinities or specificities. This correlation allowed direct assessments of the 'sequence-activity' relationships in a manner analogous to relating structures and activities for small-molecule drugs. Essential residues highlighted by this analysis were then targeted for subsequent maturation to improve affinities and tune biological activity. Mao *et al.*[1] describe one matured, full-length antibody that partially antagonized Notch-1 signaling by blocking delta-like ligand 4 (DLL-4) and another antibody that exhibited more potent inhibition. These results show that a single screen could yield several lead structures with distinct biological activities.

Some targets will present challenges for the approach of Mao *et al.*[1] Antibodies expressed from germline genes have low affinities naturally, and the immune system relies on avidity to stabilize weak interactions with antigens. Naive B cells express many copies of their unique B-cell receptors on their surface, and natural decavalent IgM circulates to bind new antigens into immune complexes. In some cases, the affinities of monovalent antibody fragments may simply be too low to score. Generating libraries of multivalent species may help address this issue. Nonetheless, certain targets or functional activities may still pose difficulties. For example, all broadly neutralizing antibodies against HIV-1 identified to date have unusual structural elements (e.g., domain swaps and large loops in the complementarity-determining regions) arising from significant maturation *in vivo*[10]. It is not obvious that simple 'sequence-activity' analysis for germline genes would reveal the

best candidates for such antibodies or the subsequent routes to maturation.

Further technical developments of the method described by Mao *et al.*[1] should accommodate larger libraries than those used here while maintaining specific knowledge about each member. Declining costs of gene synthesis and advancing systems for miniaturized expression, purification and screening of multiple recombinant proteins in parallel should also help. Such advances would begin to bring the vast experience available in high-throughput screening of combinatorial libraries of small molecules to bear on the problem of identifying new therapeutic monoclonal antibodies.

COMPETING FINANCIAL INTERESTS
The author declares no competing financial interests.

1. Mao, H. *et al. Nat. Biotechnol.* **28**, 1195–1202 (2010).
2. Nelson, A.L., Dhimolea, E. & Reichert, J.M. *Nat. Rev. Drug Discov.* **9**, 767–774 (2010).
3. Hoogenboom, H.R. *Nat. Biotechnol.* **23**, 1105–1116 (2005).
4. Kindt, T.J., Osborne, B.A. & Goldsby, R.A. *Kuby Immunology*, edn. 6 (W.H. Freeman, 2006).
5. Suntharalingam, G. *et al. N. Engl. J. Med.* **355**, 1018–1028 (2006).
6. Weiner, L.M. & Carter, P. *Nat. Biotechnol.* **23**, 556–557 (2005).
7. Xu, J.L. & Davis, M.M. *Immunity* **13**, 37–45 (2000).
8. Weinstein, J.A., Jiang, N., White, R.A., Fisher, D.S. & Quake, S.R. *Science* **324**, 807–810 (2009).
9. Boyd, S.D. *et al. Sci. Transl. Med.* **1**, 12ra23 (2009).
10. Kwong, P.D. & Wilson, I.A. *Nat. Immunol.* **10**, 573–578 (2009).

Toward global RNA structure analysis

David M Mauger & Kevin M Weeks

Deep sequencing provides a first view of the RNA structures in a eukaryotic transcriptome.

The structures of RNA molecules, even in complex environments, have been interrogated for many years by studying how individual nucleotides react with enzymatic or chemical probes[1]. However, these methods have largely been limited to studying single RNAs or small fragments of large RNAs. Writing in *Nature*, Kertesz *et al.*[2] have succeeded in analyzing much of the yeast transcriptome by melding a traditional biochemical method for probing RNA structure with the power of highly parallel DNA sequencing in an approach called parallel analysis of RNA structure (PARS). Their results provide new insight into the role of mRNA structure in gene expression, confirming, in particular, that RNA structure regulates protein synthesis, most probably by controlling the accessibility of mRNAs to the translational machinery.

Kertesz *et al.*[2] extracted total RNA from *Saccharomyces cerevisiae*, enriched it for mRNA, refolded the RNA, and treated it separately with S1 nuclease and RNase V1, nucleases specific for single- and double-stranded RNA, respectively (**Fig. 1**). The resulting structure-selective cleavage fragments were converted to double-stranded DNAs and sequenced using the SOLiD system, generating tens of millions of reads. Sequences within less abundant transcripts and those that were more difficult to convert from RNA to double-stranded DNA were largely undetected. The authors then determined individual RNase cleavage sites from the sequencing reads and compared the digestion frequencies of the two RNases at each nucleotide. This comparison, termed a PARS score, provides a measure of the single-stranded or double-stranded character of each nucleotide in each RNA with sufficient read coverage. By characterizing thousands of yeast transcripts, the authors generated structural data for approximately half of the *S. cerevisiae* transcriptome.

These data are especially valuable when analyzed at a global level (**Fig. 1**). In aggregate, they lend strong support to the general hypothesis that the information encoded in large mRNAs is regulated and organized by RNA structure. In several cases, the authors were able to confirm and extend models for the roles of RNA structure previously developed through detailed analyses of individual RNAs. The data revealed that the prototypical yeast mRNA consists of regions with distinct characteristics: the open reading frame is more structured than either the 5′ or 3′ untranslated regions (UTRs). The start and stop codons, which define the protein open reading frame, tend to occur in relatively unstructured regions, a feature that may facilitate access by the translation machinery. The relative lack of secondary structure in the 3′ UTR may ensure that factors that regulate mRNA translation, stability and localization have access to their binding sites.

The finding that open reading frames are more structured than the flanking regions is consistent with a study from our laboratory[3] showing that the structured regions of the HIV-1 genome occur preferentially downstream of sequences encoding individual protein domains, perhaps facilitating protein folding during translation. Also in accordance with earlier results[4], Kertesz *et al.*[2] found that highly translated messages appear to have less defined structure near the translational start site than those with lower translation levels, suggesting that mRNA structure modulates ribosome activity.

Characterizing the extent of double- and single-stranded structure in the yeast transcriptome represents an important first step toward the goal of determining the complete three-dimensional structures of these RNAs. However, generating accurate nucleotide-resolution structures from RNase cleavage data is very challenging[5]. Moreover, *in vitro* folding, as used in PARS, is not likely to fully recapitulate folding in the cellular environment. The ability of any approach to guide RNA structure prediction is best assessed by comparing predicted models to RNA structures known to be in a functional conformation. Although very few yeast RNAs have been studied over significant lengths, the secondary structure of the 18S ribosomal RNA from the small (40S) ribosomal subunit is well characterized[6]. We found that the PARS-assisted, secondary-structure prediction for the 18S RNA[2] contains only half of the accepted base pairs for this RNA and thus does not resemble the biologically functional state in many regions. In addition, the PARS-assisted secondary structure for 9 of 14 tRNAs with read depths greater than 1 were predicted incorrectly. Thus, at this juncture, PARS results do not reproduce these physiologically relevant RNA structures and should be interpreted cautiously.

In contrast, previous, lower-throughput approaches have yielded very high-resolution structural information under defined, controlled conditions in which the RNAs retain most of their native structure. In the mid-1980s, Harry Noller's laboratory demonstrated that primer extension could identify RNA nucleotides modified by small, structure-selective, organic chemicals. Using these reagents,

David M. Mauger and Kevin M. Weeks are in the Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina, USA.
e-mail: weeks@unc.edu