dominant core constituent. For that material, the Berkeley and Mainz data are in partial agreement to 60 GPa but diverge at higher pressure. For iron, the Berkeley data (*11*) give a very high melting temperature, whereas the Mainz data (*12*) suggest only a modest increase in melting temperature with pressure. Thus, melting is not always measured at lower temperatures in Mainz, and differences between the laboratories are not systematically in the same direction.

Zerr and Boehler's data imply a dramatic decrease in the homologous temperature in the lower mantle. Indeed, if they are right, one could ask whether earthquakes might occur in the deep mantle. None has yet been observed. As noted by the authors, high melting temperatures for perovskite would preclude large-scale melting and chemical differentiation in the lower mantle and would allow very high temperature gradients near the core-mantle boundary. On the other hand, the Sweeney and Heinz data (8) suggest that the homologous temperature of the lower mantle approaches or exceeds 1. Such a high value also seems at odds with conventional wisdom for the dynamics of the lower mantle. Geophysical determinations (*3*) give lower mantle viscosities 30 times higher than the viscosities in the upper mantle.

In his incorrect estimates for Earth's age, Kelvin took bounds for the melting of the mantle between 1300 and 4300 K—a range nearly as wide as results of all recent experimental data discussed here. Our ability to accept a greater age for the Earth has not been based on a refinement of measured melting temperatures but rather is based on a change in the theoretical framework of interpretation. A lesson can be drawn from Kelvin's overconfidence. It is quite conceivable that some important bit of physics has been overlooked in one or more of the recent high-pressure experiments. Despite best intentions, melting may not have been accurately determined. But more importantly, efforts to understand dynamics in the mantle are probably as limited by inadequate theory as by possibly inaccurate experiments. If either Zerr and Boehler or (on the opposite extreme) Sweeney and Heinz are right, one or more of our current working concepts may require modification. Homologous temperature or conventional views of deep-mantle convection may not be correct theoretical concepts. Both improved experiments (as exemplified by the work of Zerr and Boehler) and better theories are needed to improve our understanding. A consensus on high-pressure melting may require further efforts to replicate these very difficult experiments.

References

1. Lord Kelvin (W. Thomson), *Mathematical and Physical Papers V* (Cambridge Press, London, 1911), pp. 205–230
2. A. Zerr and R. Boehler, *Science* **261**, 553 (1993).
3. G. Davies and M. A Richards, *J. Geol.* **100**, 151 (1992).
4. K. C. Creager and T. M Boyd, *J Geophys Res.* **96**, 2293 (1991).
5. E Knittle and R. Jeanloz, *Science* **235**, 668 (1987).
6 D. L. Heinz and R Jeanloz, *J Geophys. Res.* **92**, 11437 (1987).
7 E. Knittle and R Jeanloz, *Geophys. Res. Lett.* **16**, 421 (1989).
8 J. S. Sweeney and D. L Heinz, *ibid.* **20**, 855 (1993).
9 E. Ito and T. Katsura, in *High-Pressure Research: Application to Earth and Planetary Sciences*, Y. Syono and M. H. Manghnani, Eds (American Geophysical Union, Washington, D.C , 1992)
10. T. Kato and M. Kumazawa, *Nature* **316**, 803 (1985).
11. Q Williams, R. Jeanloz, J. Bass, B. Svendson, T. J. Ahrens, *Science* **236**, 181 (1987).
12 R. Boehler, *Nature* **363**, 534 (1993)

The Future of DNA Sequencing

Lloyd M. Smith

The demand for improved DNA sequencing methodologies posed by the Human Genome Project has spurred the development of both conventional and unconventional approaches (*1*). The conventional approaches employ the same strategy as the method developed by Sanger and co-workers in the mid-1970s; nested sets of DNA fragments are produced by enzymatic termination reactions and are separated by size with denaturing polyacrylamide gel electrophoresis. The overall sequencing process with this strategy is complex and multifaceted. For high-volume use, every aspect must be streamlined and automated, and the individual steps need to be integrated to provide a seamless whole. Solution of these problems comprises much of the sequencing technology development supported by the Human Genome Project, and the resultant systems for large-scale sequence analysis are the most likely near-term candidates for performance of the sequencing.

The majority of the large-scale sequencing being done today employs a random selection (shotgun) strategy and fluorescence-based automated sequencing instruments. Large-scale sequencing is approaching a throughput of 1 megabase (Mb) per year of finished sequence at a cost in the vicinity of a dollar per base (*2*). A further four- to fivefold drop in cost will bring the price into a range commensurate with the budget and goals of the Human Genome Project. The shortfall in throughput capability is more serious: An aggregate capability of sequencing at least 500 Mb per year is needed by the year 2000 for the minimal three billion bases of the genome to be in hand by 2006, the nominal target date for the project. This is at least two orders of magnitude from our present ability.

Two recent developments have the potential to significantly increase the throughput of electrophoresis-based sequencing instruments: ultrathin gel electrophoresis and replaceable "gels." Performing electrophoresis in ultrathin (50 μm) gels increases heat transfer efficiency; this in turn permits higher electric fields to be applied without deleterious thermal effects. The higher electric fields give correspondingly more rapid separations; an increase in separation speed of about an order of magnitude is readily achieved. This gain is not achieved without cost, however. Spacing between the DNA fragments decreases at higher electric field strengths, potentially decreasing resolution and read length (*3*). Nonetheless, in a properly designed system greatly increased separation speeds, and hence throughput, may be obtained. Work is in progress on such systems, both with arrays of capillaries (*4*) and ultrathin slabs (*5*).

The significant increases in speed now attainable in gel electrophoresis highlights another limitation of today's sequencing systems: gel preparation. The cross-linked polyacrylamide gels used in sequencing are not generally reusable nor are they amenable to commercial production due to their chemical instability. Thus the burden of gel preparation continues to rest with the user. This requires significant labor as well as being an undesirable source of irreproducibility in the sequencing process. An electrophoretic system requiring only an hour for the separation would require twenty-four gels for continuous around-the-clock operation. Either manual gel replacement or a cumbersome automated system would be needed, both undesirable features. A possible alternative solution is DNA separation in entangled polymer networks (*6*). These are aqueous solutions of hydrophilic polymers, typically linear polyacrylamide or modified celluloses. The polymer solutions form interpenetrating networks similar to those of cross-linked polyacrylamide gels and permit comparable sieving type separations.

The big advantage is that these non–

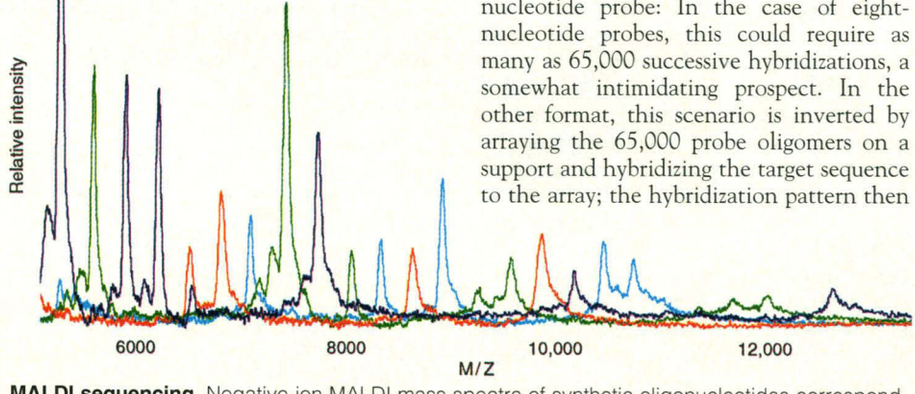
The author is in the Department of Chemistry, University of Wisconsin, Madison, WI 53706

cross-linked solutions are liquid, albeit viscous, and thus may in principle be pumped in and out of the electrophoresis cell as desired. They offer for the first time the possibility of a closed sequencing instrument capable of unattended operation. The input to such instruments would be sets of DNA sequencing reactions, and the output streams of raw sequence data; as automation of the other aspects of sequencing progresses, modules could be interfaced to the instrument to yield eventually a fully automated system.

Although shotgun methods dominate large-scale sequencing efforts, other strategies without the intrinsic redundancy and gap closure problems of shotgun sequencing continue to tantalize sequencing devotees. Multiple short (five- to six-nucleotide) oligonucleotides juxtaposed to form an effective primer (*7*) open the possibility of utilizing presynthesized libraries of short oligomers, from which appropriate sets may be chosen to form primers in "walking" steps along a sequencing target of interest. Current efforts are directed toward determining the best conditions and the rules for primer selection and to developing compatibility with fluorescence-based instrumentation (*8*). Another approach to minimize redundancy is to pre-order the clones to be sequenced; deletion libraries continue to be useful in small-scale sequencing projects, and transposon-based mutagenesis for the insertion of primer sequences, followed by clone ordering, is being actively pursued as a strategy in several larger scale projects (*9*).

Unconventional approaches to sequencing vary in their degree of radicalism. Mass spectrometric approaches range from simple replacement of the fluorescence detector in gel electrophoresis with a mass spectrometric detector to ambitious approaches for sequence determination on a single large DNA molecule in an ion trap. An intermediate approach being pursued in several laboratories is the replacement of the gel electrophoretic separation of Sanger fragment sets with a mass spectral separation and detection. This possibility arises because of the relatively new technique of matrix-assisted laser desorption and ionization (MALDI), which permits singly charged ions from proteins as large as 300,000 daltons to be produced and mass analyzed (*10*). Most MALDI work to date has focused on the analysis of proteins; the mass range and generality for proteins are not yet available in the MALDI analysis of nucleic acids. Whereas homopolymers of deoxythymidine are readily analyzed, other homopolymers and mixed sequence oligomers are more recalcitrant (*11*). With the best matrix for mixed sequence oligomers found to date, 3-hydroxypicolinic acid, the largest oligomer analyzed was 67 bases, with about 10 pmol of each component analyzed (*12*) (see figure). The key to the problem appears to lie in as yet mysterious interactions of the analyte with matrix crystals that form during sample preparation. Development in this area continues to rely primarily upon an empirical search for new matrices, reflecting our limited understanding of the underlying processes. The technique is at least an order of magnitude away from practicality in both mass range and sensitivity. The optimist will note, however, the rapid rate of progress in this field over the last 3 to 4 years and project successful further development. If this proceeds as hoped, separations of Sanger mixtures may be performed in seconds, comparing well to the hour or so required even in the ultrafast gel electrophoretic formats.

Another up-and-coming new technology is sequencing by hybridization (SBH) (*13*). Here sequence information is obtained by hybridization of small (for example, eight nucleotides) probes to a target to be sequenced. Two formats have been proposed for the method. In one, many targets are arranged on a support and hybridized successively with every possible oligonucleotide probe: In the case of eight-nucleotide probes, this could require as many as 65,000 successive hybridizations, a somewhat intimidating prospect. In the other format, this scenario is inverted by arraying the 65,000 probe oligomers on a support and hybridizing the target sequence to the array; the hybridization pattern then determines the sequence. Several technical issues arise in practice: A particularly thorny issue is the effect of repetitive DNA upon sequence reconstruction. Because of this, most practitioners of SBH agree that it is unlikely to serve as a primary sequencing tool for complex genomes. Much of the appeal of SBH derives from its clear utility in other areas. For example, probe arrays can be custom designed for specific DNA diagnostic applications, such as mutation detection, HLA typing, and genetic identification, or for repetitive sequencing of defined regions or detection of genetic variations within these regions. These arrays could be very powerful for the sequence analysis of short nonrepetitive DNA fragments and used in conjunction with primary sequence data derived by other methods to provide a rapid means of confirming and correcting sequence data. Support-bound oligonucleotide arrays of limited complexity can be made manually by simply arranging suitably derivatized oligonucleotides on a surface; preparation of large combinatorial arrays requires more sophisticated approaches involving photolithography and microfabrication (*14*). Such arrays will be available for use in the not-too-distant future.



MALDI sequencing. Negative-ion MALDI mass spectra of synthetic oligonucleotides corresponding to mock A, C, G, and T sequencing reactions. The order of the peaks corresponds to the sequence (A, purple; C, blue; G, green; T, red). [Adapted from M. C. Fitzgerald, L. Zhu, L. M. Smith, *Rapid Commun. Mass Spectrom.*, in press.]

In this short perspective one can only give a taste of the interesting and innovative approaches to improved sequencing underway around the world. The sequence data in GenBank have grown exponentially over the last decade (it presently contains 150 Mb of sequence data). If this rate of growth can be maintained over the next decade, our capabilities will be sufficient to meet the mandate of the Human Genome Project, and the resultant information and technological capability will radically change much of the nature of biological and medical research. The next 5 years are crucial as we watch to see how many (and which) of the seeds that have been sown give rise to blooms.

References

1. T. Hunkapiller, R. J. Kaiser, B. F. Koop, L. Hood, *Science* **254**, 59 (1991).
2. J. Sulston *et al.*, *Nature* **356**, 37 (1992).
3. J. A. Luckey, T. B. Norris, L. M. Smith, *J. Phys. Chem.* **97**, 3067 (1993); J. A. Luckey and L. M. Smith, *Electrophoresis* **14**, 492 (1993).
4. R. A. Mathies and X. C. Huang, *Nature* **359**, 167 (1992).
5. A. J. Kostichka *et al.*, *Bio/Technology* **10**, 78 (1992).
6. P. D. Grossman, in *Capillary Electrophoresis: Theory & Practice*, P. D. Grossman and J. C. Colburn, Eds. (Academic Press, San Diego, 1992), pp. 215–233; D. Tietz and A. Chrambach, *Electrophoresis* **14**, 185 (1993); M. C. Ruiz-Martinez *et al.*, *Anal. Chem.*, **65**, 2851 (1993).
7. J. Kieleczawa, J. J. Dunn, F. W. Studier, *Science* **258**, 1787 (1992); L. E. Kotler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4241 (1993).
8. W. Hou and L. M. Smith, *Nucleic Acids Res.* **21**, 3331 (1993).
9. M. P. Strathmann *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 1247 (1991); C. M. Berg *et al.*, *Gene* **113**, 9 (1992).
10. F. Hillenkamp, M. Karas, R. C. Beavis, B. T. Chait, *Anal. Chem.* **63**, 1193A (1991).

11. G. R. Parr, M. C. Fitzgerald, L. M. Smith, *Rapid Commun. Mass Spectrom.* **6**, 369 (1992).
12. K. J. Wu *et al.*, *ibid.* **7**, 142 (1993).
13. R. Drmanac and R. Crkvenjakov, *Int. J. Genome Res.* **1**, 1 (1992); R. Drmanac *et al.*, *Science* **260**, 1649 (1993).
14. S. P. A. Fodor *et al.*, *Science* **251**, 767 (1991); *Nature* **364**, (1993).

The Two-Component Pathway Comes to Eukaryotes

Daniel E. Koshland Jr.

Two simplifying principles of biology are what might be called "the principle of redundancy" and "the principle of diversity." Mother Nature follows the principle of redundancy by selecting a simple mechanism or module as a building block for a complex system and then using that module over and over again in other systems. The principle of diversity utilizes the concept that there are many ways of achieving the same goal, for example, creating a living organism or generating motility. Just as very different chemicals can be created by different permutations of protons, neutrons, and electrons, so can different biological species be constructed from similar receptors, enzyme pathways, and membranes.

These principles are admirably illustrated in this issue of *Science* in which two papers report that a signal transduction system, the "two-component" pathway, known to be widespread in bacteria, also occurs in eukaryotes (*1*, *2*). This pathway, as understood from bacterial systems, is shown in the figure: The stimulus (S)—which can be a nutrient that activates a chemotaxis pathway, a condition such as osmotic pressure, or low nitrogen concentrations—binds to its receptor (R), inducing a conformational change. This change causes the receptor to interact with a kinase and, consequently, the kinase autophosphorylates on a histidine residue near its carboxyl-terminal end. That phosphate is then transferred to a carboxyl side chain of aspartate in the amino-terminal end of the response regulator. The response regulator binds to an output protein, such as a flagellar motor or a transcription factor.

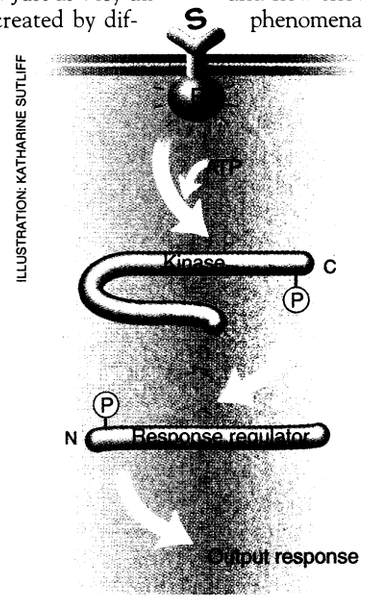


ILLUSTRATION: KATHARINE SUTLIFF

The two-component pathway.

The response regulator is a molecule whose concentration controls an output response. It was originally postulated to explain memory during chemotaxis in bacteria (*3*) and has now been found to operate in mammalian cells as well (*4*). That the response regulator is a part of a two-component phosphorylation cascade (the kinase being the first component) was initially described for the nitrogen fixation pathway (*5*, *6*), then found also in chemotaxis (*7*), and now shown to operate in such varied phenomena as sporulation, virulence, and transformation competence (*8*). Phosphorylation cascades are also common in eukaryotes, but in eukaryotes they generally employ serine, threonine, and tyrosine phosphorylations (*9*). In contrast, bacterial phosphorylation cascades usually use an internal phosphotransfer reaction of the kinase to phosphorylate a histidine, and the response regulators are phosphorylated on a carboxyl group (*7*, *8*, *10*, *11*). Indeed the sequences around the histidine (~100 amino acid residues) and around the carboxyl residue (a three-dimensional array of carboxyl groups) are characteristic and occur in the proteins of each cascade (*8*, *12*).

These characteristic sequence similarities led Chang and co-workers from the Meyerowitz group to identify the ethylene response system in plants (which controls growth and ripening) and Ota and Varshavsky to identify the *SLN1* system in yeast (also critical for growth) as two-component systems. Indeed, the most important technique in making these discoveries was computerized sequence comparison, now a pervasive tool in biological sciences. The number of gene sequences now accumulated in the data banks ensures that the identification of a new gene sequence can in many cases immediately lead to an assignment of function by analogy. In this week's *Science*, the sequences of new genes in a plant and yeast led to the deduction of the identity of the pathways.

The eukaryotic two-component systems have elaborated upon the most common bacterial mechanism in that both the kinase domain and the response regulator domain are parts of the primary structures of the *ETR1* gene product of the ethylene system (*1*) and the SLN1 protein of the yeast growth response (*2*). Both ETR1 and SLN1 have characteristic hydrophobic transmembrane sequences, and so it appears that they are receptors in which the two-component system has been built into the carboxyl-terminal cytoplasmic domain. Bacteria had begun this consolidation process: In the chemotaxis and nitrogen fixation systems, the kinases are separate proteins, activated or deactivated by a receptor with no covalent ties to the kinase. But in osmolarity regulation, for example, the kinase is a domain of the receptor protein primary structure, a part of the receptor itself. And in a few bacterial systems all three functions—receptor, kinase, and response regulator—are combined in one polypeptide. The new reports expand our horizons. It now seems possible that all of the prokaryotic permutations may be found in eukaryotes.

The principle of redundancy leads to similar transduction schemes in markedly different organisms, and the principle of diversity suggests that the new schemes are tailored to the needs of the organism in which they reside. Illuminating the details will be tasks for the future. For instance, in bacteria the response regulators generally have labile carboxyl phosphates due to the presence of phosphatases, and it will be interesting to see whether this specific feature is retained in the eukaryotic two-component pathways. The findings that such diverse systems as plants and bacteria share similar pathways should not be considered denigrating to eukaryotes (even humans) nor uplifting to bacteria. Rather, they are further unifying examples of the general principles of redundancy and diversity.

The author is in the Department of Molecular and Cell Biology at the University of California at Berkeley, Berkeley, CA 94720.

References

1. C. Chang, S. F. Kwok, A. B. Blecker, E. M. Meyerowitz, *Science* **262**, 539 (1993).
2. I. M. Ota and A. Varshavsky, *ibid.*, p. 566.
3. D. E. Koshland Jr., *ibid.* **196**, 1055 (1977).
4. B. H. Morimoto and D. E. Koshland Jr., *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10835 (1991).
5. A. J. Ninfa and B. Magasanik, *ibid.* **83**, 5905 (1986).
6. ———, *ibid.*, p. 5909.
7. J. F. Hess *et al.*, *Nature* **336**, 139 (1988).
8. J. B. Stock *et al.*, *ibid.* **344**, 395 (1990).
9. E. G. Krebs and E. H. Fisher, *Biochem. Biophys. Acta* **20**, 150 (1956).
10. J. F. Hess *et al.*, *Cell* **53**, 79 (1988).
11. D. A. Sanders *et al.*, *J. Biol. Chem.* **264**, 21770 (1989).
12. J. S. Parkinson, *Cell* **73**, 857 (1993).