Applications to recombinant fermentations for β-galactosidase, glyceraldehyde-3-phosphate dehydrogenase and streptokinase have shown significant improvements in performance, better than even well-mixed noise-free (ideal) operation. These results suggest that the non-ideal features of large bioreactors should be usefully harnessed rather than being eliminated.

References

1 Mayr, B. *et al.* (1992) Engineering approach to mixing quantification in bioreactors. *Bioproc. Eng.* 8, 137–143
2 Pedersen, A.G. *et al.* (1994) Characterisation of mixing in stirred tank bioreactors equipped with Rushton turbines. *Biotechnol. Bioeng.* 44, 1013–1017
3 Larsson, G. *et al.* (1996) Substrate gradients in bioreactors: origin and consequences. *Bioproc. Eng.* 14, 281–289
4 Royce, P.N. (1993) A discussion of recent developments in fermentation monitoring and control from a practical perspective. *Crit. Revs Biotechnol.* 13, 117–149
5 Patnaik, P.R. (1997) A recurrent neural network for a fed-batch fermentation with recombinant *Escherichia coli* subject to inflow disturbances. *Process Biochem.* 32, 391–400
6 Tanner, R.D. *et al.* (1985) The effect of imperfect mixing on an idealised kinetic fermentation model. *Chem. Eng. Sci.* 40, 1213–1219
7 Patnaik, P.R. (1998) Dispersion-induced behaviour in sub-critical operation of a recombinant fed-batch fermentation with run-away plasmids. *Bioproc. Eng.* 18, 219–226
8 Simutis, R. *et al.* (1997) Exploratory analysis of bioprocesses using artificial neural network based methods. *Biotechnol. Prog.* 13, 479–487
9 Montague, G.A. and Morris, A.J. (1994) Neural-network contributions in biotechnology. *Trends Biotechnol.* 12, 312–324
10 Luyben W.L. (1990) *Process Modeling, Simulation and Control.* Mc Graw–Hill, New York
11 Moser, A. (1988) *Bioprocess Technology. Kinetics and Reactors.* Springer-Verlag, New York
12 Patnaik, P.R. (2001) Further enhancement of fed-batch streptokinase yield in the presence of inflow noise by coupled neural networks. *Process Biochem.* 37, 145–151
13 Patnaik, P.R. (2001) A simulation study of dynamic neural filtering and control of a fed-batch bioreactor under nonideal conditions. *Chem. Eng. J.* 84, 533–541
14 Patnaik, P.R. (1996) Incomplete mixing in large bioreactors – a study of its role in the fermentative production of streptokinase. *Bioproc. Eng.* 14, 91–96
15 Patnaik, P.R. (1999) Improvement of the microbial production of streptokinase by controlled filtering of process noise. *Process Biochem.* 35, 309–315
16 Patnaik, P.R. (2001) Enhancement of protein activity in a recombinant fermentation by optimising fluid dispersion and initial plasmid copy number dispersion. *Biochem. Eng. J.* 9, 111–118
17 Patnaik, P.R. (1999) Coupling of a neural filter and a neural controller for improvement of fermentation performance. *Biotechnol. Techniques* 13, 735–738
18 Patnaik, P.R. (2000) Optimising initial plasmid copy number distribution for improved protein activity in a recombinant fermentation. *Biochem. Eng. J.* 5, 101–107
19 Lamotte, D. *et al.* (1994) Transitory plasmid instability during the exponential phase of growth of a recombinant *Escherichia coli*. *Biotechnol. Lett.* 16, 1113–1118
20 Straight, J.V. and Ramkrishna, D. (1994) Cybernetic modelling and regulation of metabolic pathways. Growth on complementary nutrients. *Biotechnol. Prog.* 10, 574–587
21 Jones, K.D. and Kompala, D.S. (1999) Cybernetic model of the growth dynamics of *Saccharomyces cerevisiae* in batch and continuous cultures. *J. Biotechnol.* 71, 105–131
22 Patnaik, P.R. (2000) On the improvement of bacterial growth on complementary substrates by partial segregation in the broth. *J. Chem. Technol. Biotechnol.* 75, 229–236

Pratap R. Patnaik
Institute of Microbial Technology,
Sector 39-A, Chandigarh-160 036, India.
e-mail: pratap@imtech.res.in

Gene expression profiling diagnosis through DNA molecular computation

Allen P. Mills, Jr

Gene expression profiling is the characterization of cells based on the level of gene activity represented by concentrations of complementary DNA reverse transcribed from messenger RNA. The spectrum of cDNA concentrations, the expression profile, is determined using a DNA microarray. Although this approach is valuable for research, a simpler scheme that would give answers on a shorter time-scale for clinical applications is needed. An Adleman DNA self-assembly computer that would use cDNA as input might be ideal for clinical cell discrimination and a neural network architecture would be appropriate for making the necessary classifications. Preliminary experimental results suggest that expression profiling should be feasible using a DNA neural network that acts directly on cDNA.

Many people are working to develop diagnostic techniques for classifying tumor cells based on gene activity [1–4]. Gene expression profiling is cell characterization based on the level of gene activity represented by the concentrations of messenger RNA (mRNA). RNA is not very stable – it is usually converted, by reverse transcriptase, into a collection of complementary DNA (cDNA) molecules, which are more robust. The profile of active genes represented by their cDNA concentrations is measured using a DNA microarray. The cDNA is stained with a fluorescent dye and allowed to hybridize with an array of tens of thousands of DNA oligomers representing many genes. The array is then exposed to light that excites the dye. The fluorescent intensities of the various cDNA oligomers are measured and compared with the intensities observed from a library of known cells. An identification of cell type to assist in diagnosis is made based on the comparison.

This approach is especially valuable in the differentiation of pathological strains that have indistinguishable phenotypes, which can be essential for determining the best therapy. However, once the work of understanding and classifying expression profiles has been done, a routine application could use a simpler scheme, for example involving a much smaller array that would require less apparatus and that would give an immediate discrimination amongst a limited set of possibilities. This suggests that a diagnosis could possibly use only molecular beacons [5] or a limited number of probes.

Ever since Adleman's pioneering experiment in 1994 [6] it has been clear that, in principle, molecular-scale computations can be performed exploiting the self-assembly of DNA. A DNA computer using cDNA as input might be ideal for clinical cell discrimination. In particular, a neural network version of the Adleman DNA computing scheme [7] would be appropriate for making the necessary classifications. In this article we explore the possibility that the current paradigm based on microarray technology could be supplemented by an apparatus using direct computation at the

  PII: S0167-7799(01)01915-1

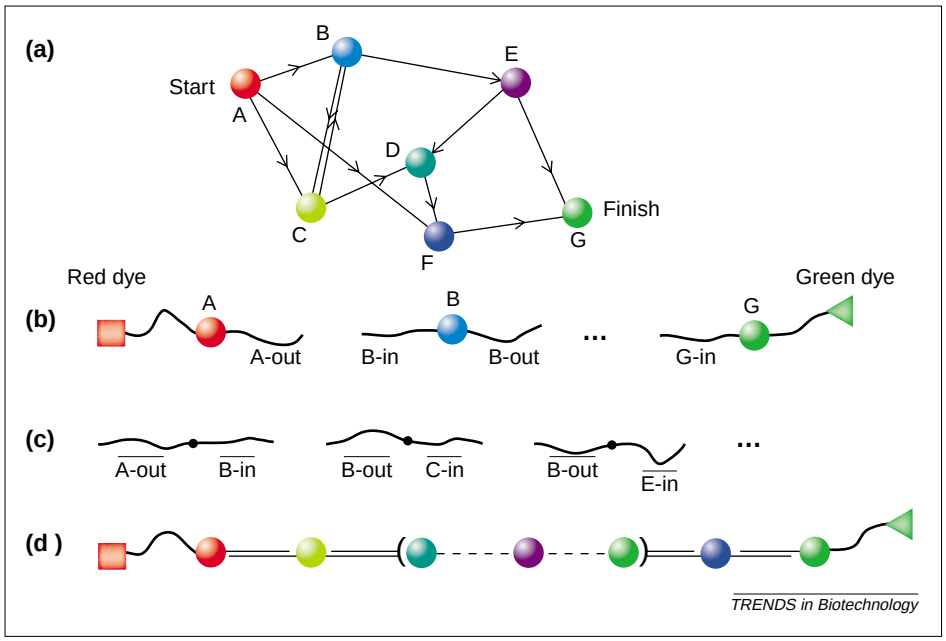

Fig. 1. Adleman's DNA solution to the directed Hamiltonian path problem. (a) A problem with seven cities and various one-way streets connecting them. (b) The cities (colored circles) and their one-way streets represented by various DNA oligomers. (c) Collection of the permitted one-way street connections represented by complementary DNA oligomers. (d) A valid solution with the correct length (seven cities) and containing the start city represented by a red dye molecule and the finish city represented by a green dye molecule.

molecular level, using self-assembly of artificial DNA oligomers in solution. Here, I outline the concepts of DNA computing and present experimental results indicating that expression profiling should be feasible using a DNA neural network acting directly on cDNA.

DNA computation

The field of molecular computing using the self-assembly of DNA began in 1994 with Adleman's *in vitro* DNA solution of the directed Hamiltonian path problem [6]: given a certain number of cities that are connected only by certain one-way streets, is there is a way to visit each city once and only once with a continuous path that begins and ends at particular preassigned cities? The amount of time to solve this problem increases exponentially with the number of cities – the difficulty increases as a nonpolynomial (NP) function of its size. The problem is of such generality that its solution would imply the solubility of all problems of its class, including the large number factoring problem. The directed Hamiltonian path problem is therefore dubbed NP-complete, and its solution would yield a way to solve many important optimization problems. Adleman, well known as a co-inventor of the RSA public key encryption scheme, which is based on the impracticality of factoring large numbers, represented the directed paths in his problem by a liquid solution containing connected pairs of short single-stranded DNA oligomers representing seven cities, as shown in Figure 1. To this he added complementary oligomer pairs representing the permitted in-going and out-going gates of the various city pairs. The oligomers were allowed to hybridize with each other, forming random self-assembled chains of double-stranded DNA that stand for various possible attempted solutions.

The molecular solution to the directed Hamiltonian path problem is any double-stranded DNA molecule containing one and only one representation of each of the seven cities, and beginning at 'start' and ending with 'finish'. The solution was found among all the incorrect possibilities by a clever series of filtering steps that selected sequentially: (1) molecules containing seven and only seven city representations; (2) molecules containing the start city; (3) molecules containing the finish city; (4) molecules containing the second city; and so on. Unfortunately, the Adleman scheme is not useful for problems larger than ~30 cities, for which buckets of DNA would be needed! Also, a problem of this size can be solved using silicon computers, so it might seem that there would be no need for such a DNA computation. However, as is routinely done in silicon computers, problems of this type could be divided into millions of smaller problems. The use of Adleman or

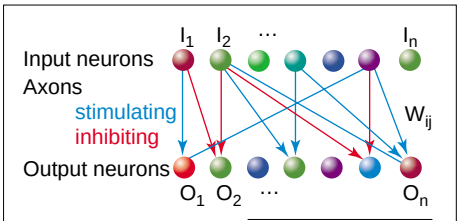

Fig. 2. A perceptron, an artificial neural network with no hidden layers. The activity at a collection of input neurons {I} (i.e. the whole set of object I) is communicated to a set of output neurons {O} (i.e. the whole set of object O) via stimulating or inhibiting axon connections with strengths W_{ij}. The output of a neuron is proportional to the sum of its inputs with upper and lower limits beyond which the output saturates.

Boolean DNA computation might thus be extended into a practical domain if microfluidic techniques were used to break up the problem into more practical sized pieces.

If applied to large problems the Boolean DNA computing paradigm would also suffer from the accumulation of mishybridization errors, even if these occur with small probability. An error-correcting scheme, perhaps similar to the codes used in ordinary computers, is a possibility. Another approach to DNA computing, suggested by Platzman in 1998, is to create an inherently fault tolerant neural network version of Adleman's computer. A neural network is a highly interconnected collection of neurons talking to each other via stimulating or inhibiting axon connections. The neurons can be thought of as little amplifiers that have one or more inputs and outputs. A neuron acts as a decision-making device because its outputs saturate when the sum of the inputs is greater than or less than certain threshold values. By appropriate choice of the weights or strengths of the connections in a neural network, any function can be approximated [8]. Neural networks are useful for classifying, generalizing and predicting based on a limited data set, and thus might be helpful for making diagnoses once the rules have been established by careful laboratory and clinical studies [9]. The simplest form of neural network, shown in Fig. 2, and known as a perceptron [10,11], consists of a single input layer of neurons connected to a single output layer. More general architectures having hidden layers of neurons are needed to design networks with the maximum utility.

DNA neural networks

A neural network can be represented using certain DNA oligomers to represent neurons

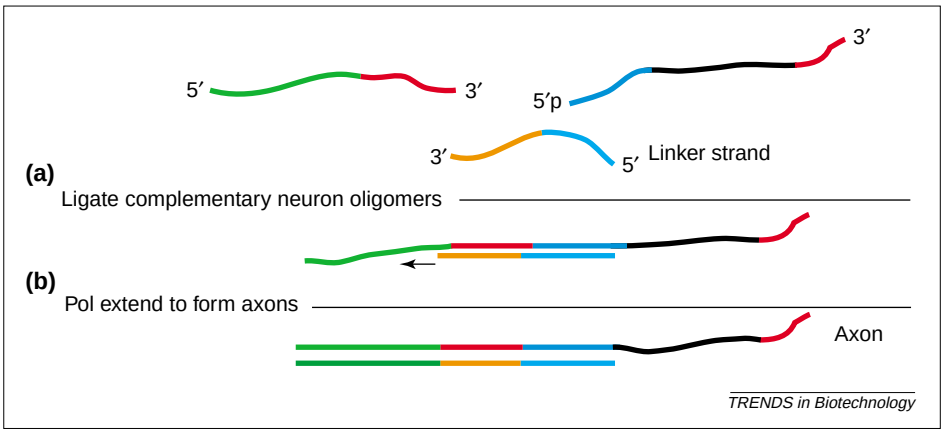

Fig. 3. Steps in forming an axon. (a) DNA oligomers complementary to the DNA oligomers representing the desired input and output neurons are attached to each other using a linker oligomer and T4 DNA ligase. (b) The output end of the linked oligomers is protected from hybridization to input neuron oligomers, by polymerase extension of the linker oligomer along the output oligomer.

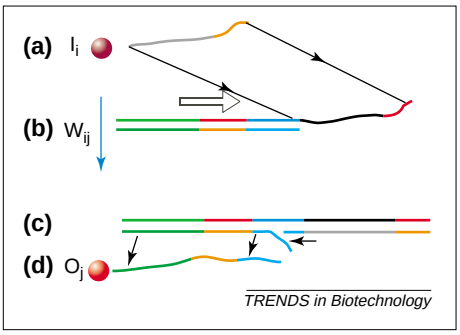

Fig. 4. Single layer DNA neural network represented by DNA. (a) The activities of the network's input neurons are represented by the concentrations of DNA oligomers I_i. (b) The strength of the axon connection between an input I_i and an output O_j is represented by the concentration of axons W_{ij} that have single-stranded segments complementary to I_i and double-stranded segments composed of O_j and its complement. (c) To operate the network, the collections of input neuron oligomers and axons are mixed together and exposed to the action of exonuclease-free Klenow fragment DNA polymerase. (d) If an axon has been primed by an oligomers I_i, it will be extended by the polymerase, thus causing the release of the output oligomer O_j.

and connected pairs of DNA oligomers complementary to two different neuron DNA oligomers to represent the axons. The concentrations of the neuron DNA oligomers represent their states of activity. The concentrations of the axon DNA oligomers govern the rate at which input neuron DNA concentrations lead to the production of an associated concentration of output neuron oligomers. The manner by which this is accomplished is a matter of choice, given the array of enzymes and reactions provided both by Nature and man's ingenuity.

One way to represent a neural network in DNA is to use axons that are protected on one end to insure that hybridization with the input neuron oligomers only occurs on the desired input end of the axon. The axons are constructed as shown in Fig. 3: two oligomers complementary to the desired input and output neurons are temporarily connected via a linker oligomer that attaches itself to sticky ends on the two oligomers complementary to the neurons. The two oligomers are subsequently joined permanently using T4 DNA ligase. The axon is then protected (prevented from hybridization with neuron oligomers) on its output end by exonuclease-free Klenow fragment DNA polymerase extension of the 3′ end of the linker oligomer.

To effect the operation of the single-layer neural network, a collection of input neuron oligomers is allowed to hybridize with axons and acted on by exonuclease-free Klenow fragment DNA polymerase in the suitable reaction buffer, as indicated in Figure 4. Output neuron oligomers are released by the polymerase from those axons that have been primed on their single-stranded ends by complementary neuron oligomers.

Experiment

To test this scheme, a set of random 50% CG neuron oligomers and their complements 26 bases long was selected to have a minimal tendency to stick to themselves forming hairpins or other entanglements. The neuron DNA oligomers and their complements are labeled EP or EM, for positive (stimulating) and negative (inhibiting) amplitudes respectively. A linker DNA oligomer of 24 bases was chosen similarly to the neuron DNA oligomers. Oligomers complementary to either the 5′ or 3′ half of the linker oligomer are attached to the 5′ ends of an input neuron oligomer and the 3′ ends of an output neuron oligomers respectively for forming an axon using ligase. Fluorescently-labeled probe oligomers, consisting of the neuron oligomers or their complements joined at their 5′ ends to a DNA dimer and the dye TET, are used to effect the semiquantitative detection of various oligomers in a solution using gel electrophoresis. The gels are read by exciting the dye with 514 nm light and recording only the TET fluorescence light using a 530 nm filter and a charged coupled device (CCD) camera. Contrast in the CCD images is enhanced (see Fig. 5) using a mod(n) intensity display.

First, axons were prepared from input DNA oligomers EM16 and output DNA oligomers EP13 as indicated in Fig. 5a, and examined using gel electrophoresis. The four columns in Fig. 5a are gel lanes that were each loaded with the same solution of unprotected axons formed from the two neuron oligomers EP13 and EP16, plus a quantity of one of the four fluorescent probes, TagEP13, TagEM13, TagEP16 and TagEM16 respectively. Note that TagEP13 is the probe for EP13, and therefore contains the single-stranded oligomer EM13 that is complementary to EP13. We are testing for the presence of the two neuron oligomers that are not supposed to be present in our unprotected axon as a control for the next part of the experiment (Fig. 5b), in which a complementary oligomer EM13 will be displaced from the protected axon. The gel lanes display short fluorescent DNA fragments at the bottoms of the photos and long ones at the tops. In Fig. 5a and 5b we see the presence of free dye molecules and excess probes in all four lanes. In Fig. 5a only the expected neuron oligomers EP13 and EP16 are visible, both as the desired ligated unprotected axons and as unligated neuron oligomers. The latter are present owing to the less-than-perfect efficiency of the ligase reaction.

The axons were then protected from hybridizing to any input neuron oligomers on their single-stranded 5′ ends using exonuclease-free Klenow fragment DNA polymerase extension to cause their 5′ ends to become double-stranded. The complement to the input neuron single-stranded oligomer was added as a primer, and the solutions were purified with a resin to remove all single-stranded DNA fragments. At this point, nothing but free probes can be observed in an

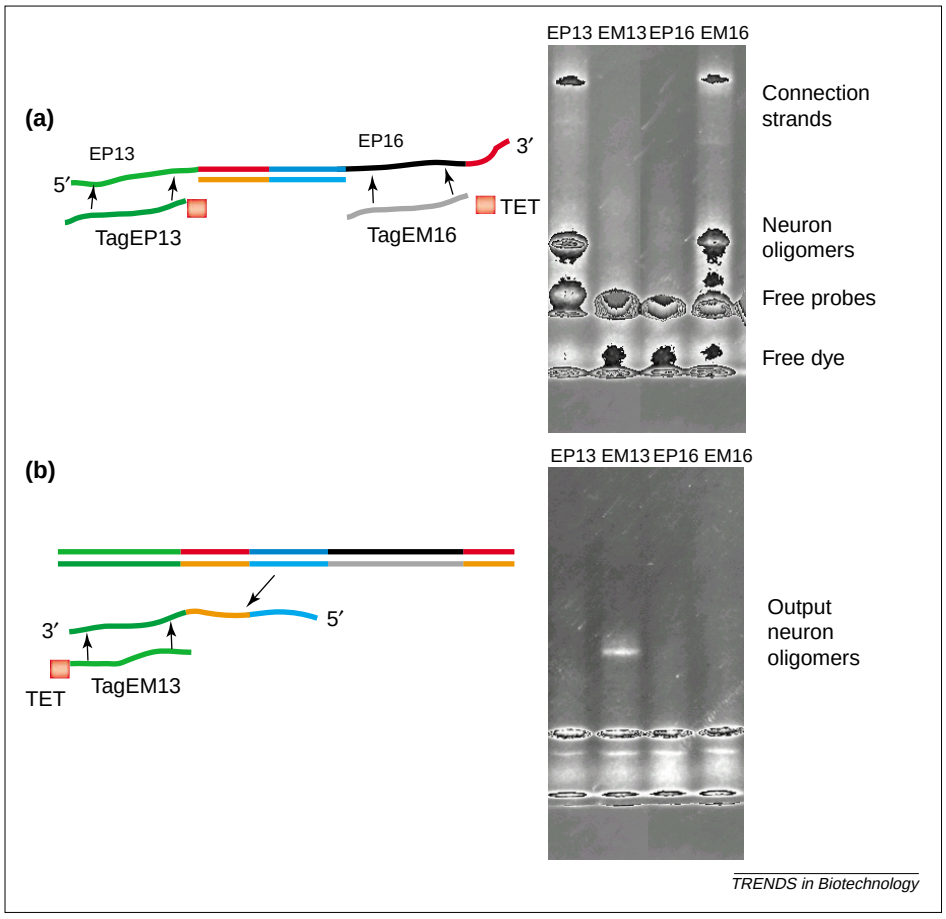

Fig. 5. Test of a single node suitable for making a DNA neural network. (a) Axon connecting an input neuron EP16 to an output neuron EM13. The axon has been tagged successively in the four lanes of the photo by fluorescent dyes complementary to the four oligomers EP13, EM13, EP16 and EM16. The gel photo shows that the axon is of double length and composed of neuron oligomers EP13 and EM16. (b) The axon has been made double-stranded on its 5′ end by polymerase extension and then primed by an input oligomer EP16 so that it is completely double-stranded. The output neuron oligomer EM13 is released by the further action of polymerase, and is the only oligomer visible in the gel photo.

electrophoresis gel because the fluorescent probes do not bind to double-stranded DNA. The single-stranded output neuron oligomer was then released by the strand displacement that occurs when exonuclease-free Klenow fragment DNA polymerase extended the input neuron oligomer along the double-stranded portion of the axon. The output oligomer hybridized to its complementary probe is clearly identified as EM13 in Figure 5b. Because the DNA was not denatured during processing, there is no way that these oligomers could have appeared other than by polymerase extension. This indicates that the scheme for implementing a small-scale neural network work is possible, and I am now working on scaling up to a problem containing 16 neurons and of order 256 axons.

The neural network needed to solve an expression profiling classification would probably require only a few thousand different neurons and might be configured as a one or two layer neural network. Such a relatively small DNA neural network would have relatively rapid cycle times and need only small quantities of neuron and axon solutions. The well-known error-tolerance of neural networks suggests that networks of the required size should perform well despite the presence of a background of undesirable reactions.

It is thus possible that we will be able to make a simple expression profiling neural network that could be used as a routine diagnostic tool, as suggested in Figure 6. Much work needs to be done after the hardware solution is perfected to find a set of appropriate axon concentrations that will do an acceptable job of classifying clinical cell samples. The longer-term prospects include the possibility that the same technology could be scaled up to sizes that would provide massive computing capabilities not available otherwise.

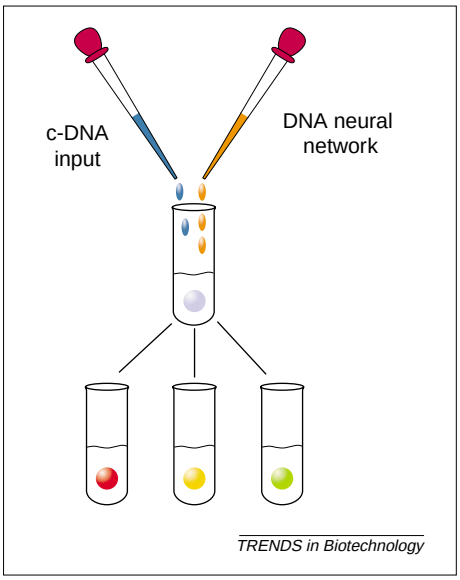

Fig. 6. Possibility for an expression profiling kit to assist in rapid diagnosis. Complementary DNA representative of the messenger RNA gene activity of a sample of cells is combined with a DNA neural network solution. After a suitable interaction time, an indicator on the sample vial exhibits a color corresponding to the network's evaluation of the cell type from a limited set of possibilities.

References

1 Eisen, M.B. and Brown, P.O. (1999) DNA arrays for analysis of gene expression. *Methods Enzymol.* 303, 179–205
2 Duggan, D.J. *et al.* (1999) Expression profiling using cDNA microarrays. *Nat. Genet.* 21, 10–14
3 Ross, D.T. *et al.* (2000) Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* 24, 227–235
4 Perou, C.M. *et al.* (2001) Molecular portraits of human breast tumours. *Nature* 406, 747–752
5 Li, Q.G. *et al.* (2000) Molecular beacon-based homogeneous fluorescence PCR assay for the diagnosis of infectious diseases. *Anal. Sci.* 16, 245–248
6 Adleman, L.M. (1994) Molecular computation of solutions to combinatorial problems. *Science* 266, 1021–1024
7 Mills, A.P. Jr. *et al.* (2000) DNA analog vector algebra and physical constraints on large-scale DNA-based neural network computation. In *DNA Based Computers V*, (Winfree, E. and Gifford, D.K., eds), pp. 65–73, American Mathematical Society
8 Funahashi, K. (1989) On the approximate realization of continuous mappings by neural networks. *Neural Netw.* 2, 183–192
9 Penny, W. and Frost, D. (1996) Neural networks in clinical medicine. *Med. Decis. Making* 16, 386–398
10 McCulloch, W.S. and Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5, 115–133
11 Rosenblatt, F. (1958) The perceptron: a probalistic model for information storage and organization in the brain. *Psychol. Rev.* 65, 386–408

Allen P. Mills, Jr.
Physics Dept, University of California, Riverside, CA 92521-0413, USA.
e-mail: apmjr@citrus.ucr.edu