Molecular computing revisited: a Moore’s Law?

Michael S. Livstone, Danny van Noort and Laura F. Landweber

Dept of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA

Moore’s Law states that the processing power of microchips doubles every one to two years. This observation might apply to the nascent field of molecular computing, in which biomolecules carry out logical operations. Incorporation of new technologies that improve sensitivity and throughput has increased the complexity of problems that can be addressed. It is an ultimate goal for molecular computers to use the full potential of massive parallelism.

In 1994, Leonard Adleman performed the first successful experiment that solved a mathematical problem using DNA [1]. In doing so, he provided a concrete demonstration of what many biologists had long believed: that the components that encode the building blocks of life are, at their core, a complex computer that can process information in many different ways. Since then, the field of ‘molecular computing’ including DNA, RNA and other information-bearing biopolymers, has been the subject of numerous scientific articles, books, international conferences and even a posthumously published novel by Robert Ludlum [2]. Two factors drive progress in molecular computing: the use of the inherent massive parallelism of nucleic acid interactions to perform many computations simultaneously and the incorporation of new technologies into nascent ‘computers’ to improve automation and sensitivity. Here, we examine the nature of this parallelism and the technologies that have been applied to molecular computing (see Box 1).

Representative molecular computing problems

Important considerations in molecular computation include the methods of encoding information, generating potential solutions and selecting and identifying correct solutions. As a framework, we discuss two NP-complete problems that have been approached by researchers, Hamiltonian path (HP) and Satisfiability (SAT).

Adleman’s original problem was an instance of HP, which asks: ‘Given a set of cities, a starting point, an ending point, and the (one-way) routes connecting them, is it possible to visit every city exactly once?’ One algorithm generates all possible orderings of cities. The difficulty increases exponentially with the number of cities, so that even with only seven, there are potentially 120 ways to traverse the five intermediates. Adleman’s approach [1] (reviewed in [5]) was to encode the cities as unique 20-nucleotide (nt) sequences of single-stranded DNA and the paths between them as 20-nt sequences complementary to half the sequence of each of the cities they connect. He generated all potential solutions by combining the sequences representing cities and routes in a test tube and isolated the unique correct solution using a three-step process. First, routes that begin and end in the correct cities were selected by PCR with specific primers. Next, routes of the correct length, 140 base pairs, were gel-isolated. Finally, he sequen­tially purified those molecules that hybridized with each of the five intervening ‘cities’. Including verification, the process took approximately one week of bench time.

The most difficult problem solved on a molecular computer to date [6] is a 20-variable, 24-clause instance
Box 1. Molecules and massive parallelism

Mathematicians classify problems according to how hard they are to solve (i.e. how much time they take). Specifically, problems are classified according to the power of a computer on which they can be solved in a reasonable, or ‘polynomial’ number of steps, versus an unreasonable, or ‘exponential’ number of steps (for reviews, see [3,4]). Problems in the group ‘P’ are (polynomially) solvable on a ‘deterministic’ computer; those in ‘NP’ (nondeterministically polynomial), generally believed to be more difficult than P, are solvable on a non-deterministic computer. The hardest of these are known as ‘NP-complete’. (Note that we do not state that NP problems require a non-deterministic computer. This is because it is one of the great unresolved mysteries of mathematics whether P and NP are actually different, and anyone who proves the conjecture either way can claim a million dollar prize from the Clay Mathematics Institute at http://www.claymath.org/prizeproblems).

Deterministic algorithms are those in which the next step is fully defined. For example, the question ‘is x a multiple of y?’ can be solved by continually subtracting y from x. The next step is always either ‘subtract y’ (if x is positive), answer ‘yes’ (if x is zero), or answer ‘no’ (if x is negative). By contrast, non-deterministic algorithms, such as the one used by Addleman to address the Hamiltonian Path problem, have steps that branch in multiple directions. If routes from Atlanta lead either to Boston or Chicago, to explore the routes the non-deterministic computer either has to ‘guess’ which way to go or ‘split’ and go both ways. Repeated splitting leads to the rapid accumulation of potential paths the computer might take.

Because they rely on the rules of chemistry, molecular computers mimic both the ‘guessing’ and ‘splitting’ aspects of non-deterministic computers. Interactions between DNA molecules in solution are governed by the probabilistic nature of intermolecular collisions. Furthermore, a very small amount, by lab standards, of DNA corresponds to a large number of molecules; one picomole of a synthetic oligonucleotide contains nearly $10^{12}$ molecules and costs less than one tenth of a cent. Therefore, a very small and inexpensive amount of DNA can randomly construct the numerous pathways characteristic of a non-deterministic algorithm.

![Diagram of molecular library](http://tibtec.trends.com)

Fig. 1. (a) Combinatorial library of molecular representations of Boolean systems. Each bit has a unique sequence encoding either the TRUE (blue) or FALSE (red) state of the variable. The optional 5’ prefix and 3’ suffix (green) each have a unique constant sequence to permit PCR or other procedures. Bit sequences shown are from [7]. (b) Detection by flap endonucleases (FEN’s). Target DNA molecules (blue) hybridize to effector oligonucleotides (red) and probes (green). Treatment with FEN’s cleaves the probe at the complex junction (yellow arrowhead). The 5’ arm of the probe hybridizes to the hairpin FRET probe bearing fluorophore (red) and quenching agent (black). Treatment with FEN’s liberates the fluorophore, permitting fluorescence. Freed effector oligonucleotide and cleaved probe can participate in multiple-turnover reactions.

http://tibtec.trends.com
sequences, should be sufficiently large that no sequence can hybridize either with another sequence or with the complement of another sequence. In the original HP problem, it was sufficient to design the oligonucleotides by eye because only seven were needed [1].

Several computer programs currently exist for generating the larger numbers of unique sequences required by more complex problems. They take into account such factors as Hamming distance [6–8] and predicted melting temperature [6,7]. The largest set of sequences generated is of 108 16-nt sequences with eight variable positions [8]. By forbidding the inclusion of G, Faulhammer et al. [7] generated a set of oligonucleotides (twenty, 15-nt bits and nine 5-nt spacers) with little or no predicted secondary structure.

It is also important to choose an appropriate application for a set of molecules. A set of short unique sequences is the ideal choice for encoding distinct points in an HP problem because each ‘city’ could have a distinct representation that requires a minimal mass of starting material. However, to perform bit-wise analyses, it is more useful to assign unique sequences to the ‘TRUE’ and ‘FALSE’ states of each bit (Fig. 1a). This representation requires a larger mass of starting material, but permits analysis of individual bits in a single operation.

**Surface-based computing**

All molecular computation protocols must at some point separate correct solutions from incorrect solutions. Although many computational operations can be performed purely in the liquid phase [1,7,9], it is often faster to perform some stage of the computation on a fixed substrate, such as glass coverslips [10], gold films [10,11], or complementary molecules covalently bound within a polyacrylamide gel [6]. This is functionally equivalent to batch-adsorption chromatography.

Liu et al. [10] attached molecules representing solutions to a SAT problem to a gold film and repeatedly hybridized them to molecules representing Boolean clauses, destroying the unhybridized single-stranded molecules after each step. This ‘mark-destroy-unmark’ strategy allows automation of queries and obviates the need for ethanol precipitation or other forms of concentration.

**Other new technologies**

The explosion of new technologies used for genomic-scale investigations is readily applicable to molecular computing. In an elaboration of their earlier work, Wang et al. [11] used two of these tools, flap endonucleases and addressed fluorescence readout, to solve a small SAT problem.

After attaching molecules representing all potential solutions to a surface and destroying those that failed to satisfy five Boolean clauses, complements of the remaining molecules were collected and divided among the wells of a microtiter plate. In each well, these molecules were used to initiate a chain reaction utilizing flap endonucleases, archaeal enzymes that cleave dangling single-stranded molecules at complex junctions (Fig. 1b), thereby liberating a fluorophore from its quenching agent. Rapid turnover allows ~1000-fold amplification at each step, which, overall, can substitute for 20 rounds of efficient PCR. Each well hosts a reaction detecting a different molecule, allowing multiple parallel detections. This process increases the throughput of the reaction and is faster than PCR, cloning, or other forms of recovery.

**The future of DNA computing**

Incorporating other new technologies might help automate and accelerate molecular computers in order to address even more complex problems. We and others would like to use microfluidics [12,13] to sort potential solutions. It might also be possible to supplant the paradigm of enzymes added exogenously to DNA or RNA by using ribozymes as 2-in-1 information-bearing and information-processing molecules, so that molecules can both represent solutions and sort themselves [14].

However, there might be fundamental limits that restrict the utility of molecular computers. As problems grow in complexity beyond the toy problems attempted so far, they require larger amounts of starting material in larger volumes of solution. Hartmanis [15,16] calculated that solving a 200-city HP problem in the manner of Adleman would require an amount of DNA more massive than the Earth. Advances in sensitivity and detection can ameliorate that burden, but the exponential growth of the number of solutions remains.

At the Eighth International Meeting on DNA-Based Computers, Akira Suyama described the development of what he termed a ‘programmable DNA computer’ [17]. The device combines technology originally developed for high-throughput screening, including a computer-controlled robotic pipettor and magnetic bead separator, to manipulate reactions in multiwell plates. Using a basic instruction set (for example, ‘CLEAVE’, ‘AMPLIFY’, ‘DETECT’), his group solved a 10-variable, 43-clause 3-SAT problem and independently performed gene expression profiling using the same set-up. The former used a ‘breadth-first’ search, in which each path is first extended by one step and then all paths are evaluated and either accepted or rejected before proceeding. This type of algorithm requires all paths to be retained in memory, which is costly for an electronic computer but takes advantage of the massive parallelism of a molecular computer. It also provides a hint for how to overcome requirements for massive amounts of starting material, since correct solutions to a problem are constructed from building blocks rather than selected from a pool of all possible solutions.

Moore’s law states that computing power doubles approximately every two years [18]. In 2000, our laboratory published a solution to a nine-variable SAT problem [7]. 2002 brought a solution to a 20-bit problem [6]. We optimistically hope to see a solution to a 40-bit problem in the next two years. The road map exists: recent technological advancements make it more practical to harness massive parallelism to perform non-deterministic algorithms on molecular computers, and well-chosen algorithms are pushing back the boundaries imposed by physical constraints.

**References**

From human genes to stem cells: new challenges for patent law?

Timothy A. Caulfield

Health Law Institute, 4th Floor Law Centre, University of Alberta, Edmonton, Alberta, Canada T6G 2H6

The social controversies that have surrounded human cloning and the use of embryos for research purposes might create unique patent issues for stem cell researchers. Policy makers should learn from the legal and ethical concerns associated with human gene patents and develop coherent patent policies that recognize and clearly address emerging social controversies.

Over the past few years, the idea of patenting human genes, plants and other higher life forms has stirred much social controversy [1–3]. Despite a large amount of public debate many of the issues remain, including concern that human gene patents harm the research environment, lead to an inappropriate commodification of life and adversely affect public access to useful health care procedures [2,4].

As policy makers continue to struggle with the issues associated with gene patents, a new area of biotechnology innovation seem poised to introduce a myriad of new patent issues. Few scientific breakthroughs have created as much social dialogue as stem cell research. To date, much of the debate has focused on the ethical issues associated with research involving human embryos and the related cloning technology. However, as stem cell research moves forward, the role of patents in this controversial area seems destined to receive more attention. Indeed, it has been noted that the number of patents applications in the area has, over the past few years, increased 300% [5]. This article is a brief review of some of the relevant social concerns.

Patentability of stem cells

Cell lines and genetically modified single cell organisms have long been considered patentable subject matter in many countries [6]. In fact, some of the earliest ‘biotech’ patents cases – most notably US Supreme Court decision in Diamond v. Chakrabarty – involve the patenting of cells [7].

To date, >2000 patents applications have been filed worldwide wherein inventions involving human and non-human stem cells are claimed. Of these, >500 applications refer to embryonic stem cells [8]. Several stem cell patents have already been issued [9]. Some of the first and most well known are held by the Wisconsin Alumni Research Fund (WARF), an entity affiliated with the University of Wisconsin that first reported the isolation and differentiation of stem cells. For example, US Patent No. 6,200,806 entitled ‘Primate Embryonic Stem Cells’ claims a purified preparation of pluripotent human embryonic stem cells (James A. Thomson is named as the inventor and WARF the assignee. The patent was granted by the US Patent and Trademark Office on 13 March 2001) [10]. Given the existence of these patents and related US jurisprudence involving other kinds of cell lines there seems little doubt that human embryonic stem cells are patentable, at least from the perspective of the US Patent and Trademark Office.

However, despite the technical patentability of stem cell