Molecular Computation of Solutions to Combinatorial Problems

Leonard M. Adleman


Stable URL:
http://links.jstor.org/sici?sici=0036-8075%2819941111%293%3A266%3A5187%3C1021%3AMCOSTC%3E2.0.CO;2-B

Science is currently published by American Association for the Advancement of Science.

Your use of the JSTOR archive indicates your acceptance of JSTOR’s Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR’s Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/aaas.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.
eral minutes to obtain near-field fluorescence spectra with good signal-to-noise ratios. Furthermore, the recent work of Xie and Dunn (33) (J. L. Lumicks et al. (34) (J. L. Lumicks et al.) showed that the metal-coated probe tip can significantly perturb the electronic properties of the molecule being detected. In contrast, the far-field confocal fluorescence approach provides unlimited laser throughput and a three-dimensional sectioning capability and is truly noninvasive, although its resolution is diffusion limited. These features are expected to allow important applications such as enhanced Raman spectroscopy at the single-molecule level and on-line fluorescence identification and sorting of individual molecules and quantum-confined nanomaterials. The extraordinary sensitivity achieved in this work allows the direct, real-time study of the dynamics of a single molecule and the chemical and biochemical reactions that such a molecule may undergo in solution.

REFERENCES AND NOTES

11. In fluorescence correlation spectroscopy, the intensity recorded at time $t$ is multiplied by that recorded at $t + 1$, and the product is integrated over a finite period of time; see D. E. Koppel, Phys. Rev. A 10, 1958 (1974).
21. Laser excitation at 488.0 and 514.5 nm was provided by an argon ion laser (Lexel, Lasers, Fremont, CA). The laser beam entered the microscope through a back port and was directed to an oil-immersion objective (x100, NA = 1.3, Nikon Instrument Group, Melville, NY) by a dichroic beam splitter (550DRLP02 or 540CRLP02, Omega Optical Inc., Brattleboro, VT). The laser beam was focused to a diffraction-limited spot by the high NA objective in our study, which was verified qualitatively by comparing the laser focal size and 1-μm polystyrene microparticles (Duke Scientific, Palo Alto, CA). Fluorescence was collected by the same objective, passed the same dichroic beam splitter, and was directed to a side port by a reflective mirror. Efficient rejection of out-of-focus signals was achieved by placing a pinhole (50 to 100 μm diameter, Newport Corp., Irvine, CA) in the primary image plane. A single interference band-pass filter (Omega Optical Inc., Brattleboro, VT) used to reject the laser light and the Rayleigh and Raman scattered photons. The fluorescence signal was then focused on a photon-counting Si avalanche photodiode (quantum efficiency, 65% at 630 nm, and dark noise, 7 counts per second) (Model SPCM-A200, EG&G Can- ada, Vaudreuil, Quebec). Time-dependent data were acquired by using a multichannel scalar (EG&G OR- TEC, Oak Ridge, TN) run on a personal computer (IBM PC-AT). Fluorescent dyes and other materials were purchased from Molecular Probes, Inc. (Eugene, OR), Eastman Chemicals (Kingsport, TN), and Sigma Chemical Corp. (St Louis, MO).
26. A complicating factor is photobleaching, which con- verts the molecule being detected into a nonfluores- cent state and prevents its further detection. The mul- tiple detection and similar fluorescence intensity ob- served for molecules of greatly different photodetec- tion efficiencies (that is, R6G and fluorescein) indicate however that photobleaching is not signifi- cant in this study.
27. This calculation is based on the diffusion equation $\tau_p = a^2/2D$, where $\tau_p$ is the diffusion time, $a$ is the diffusion distance in one dimension, and $D$ is the diffusion coefficient ($2.8 \times 10^{-6}$ cm$^2$ s$^{-1}$ for hoda- mine 6G in water/ethanol).
33. X. S. Xie and R. C. Dunn, ibid. 265, 361 (1994).
35. S. N. acknowledges the Whitaker Foundation for a young investigator award, D.T.C. is a Beckman Cell Science Scholar of Stanford University. This work was supported by Beckman-Instuments, Inc.
36. July 1994; accepted 19 September 1994

Molecular Computation of Solutions to Combinatorial Problems

Leonard M. Adleman

The tools of molecular biology were used to solve an instance of the directed Hamiltonian path problem. A small graph was encoded in molecules of DNA, and the “operations” of the computation were performed with standard protocols and enzymes. This experiment demonstrates the feasibility of carrying out computations at the molecular level.

In 1959, Richard Feynman gave a visionary talk describing the possibility of building computers that were "sub-microscopic" (1). Despite remarkable progress in computer miniaturization, this goal has yet to be achieved. Here, the possibility of computing directly with molecules is explored.

A directed graph $G$ with designated vertices $v_i$ and $v_o$ is said to have a Hamiltonian path (2) if and only if there exists a sequence of compatible "one-way" edges $e_1, e_2, \ldots, e_k$ (that is, a path) that begins at $v_i$ and ends at $v_o$ and enters every other vertex exactly once. Figure 1 shows a graph that for $v_i = 0$ and $v_o = 6$ has a Hamiltonian path, given by the edges $0 \rightarrow 1, 1 \rightarrow 2, 2 \rightarrow 3, 3 \rightarrow 4, 4 \rightarrow 5, 5 \rightarrow 6$. If the edge 2→3 were removed from the graph, then the resulting graph with the same designated vertices would not have a Hamiltonian path.

Similarly, if the designated vertices were changed to $v_i = 3$ and $v_o = 5$ there would be no Hamiltonian path (because, for example, there are no edges entering vertex 0).

There are well-known algorithms for deciding whether an arbitrary directed graph with designated vertices has a Hamiltonian path or not. However, all known algorithms for this problem have exponential worst-case complexity, and hence there are instances of modest size for which these algorithms require an impractical amount of computer time to render a decision. Because the directed Hamiltonian path problem has been proven to be NP-complete, it seems likely that no efficient (that is, polynomial time) algorithm exists for solving it (2, 3).

The following (nondeterministic) algorithm solves the directed Hamiltonian path problem:

Step 1: Generate random paths through the graph.
Step 2: Keep only those paths that begin with $v_i$ and end with $v_o$.
Step 3: If the graph has $n$ vertices, then keep only those paths that enter exactly $n$ vertices.
Step 4: Keep only those paths that enter all of...
the vertices of the graph at least once.

Step 5: If any paths remain, say "Yes"; otherwise, say "No."

The graph shown in Fig. 1 with designated vertices \( v_{in} = 0 \) and \( v_{out} = 6 \) was solved with the algorithm above implemented at the molecular level. Note that the labeling of the vertices in such a way that the (unique) Hamiltonian path enters the vertices in sequential order is only for convenience in this exposition and provides no advantage in the computation. The graph is small enough that the Hamiltonian path can be found by visual inspection, however, it is large enough to demonstrate the feasibility of this approach. It seems clear that the methods described here could be scaled-up to accommodate much larger graphs.

To implement Step 1 of the algorithm, each vertex \( i \) in the graph was associated with a random 20-mer sequence of DNA denoted \( O_i \). For each edge \( i \rightarrow j \) in the graph, an oligonucleotide \( O_{i\rightarrow j} \) was created that was the 3'-10-mer of \( O_i \) (unless \( i = 0 \), in which case it was all of \( O_i \)) followed by the 5'-10-mer of \( O_j \) (unless \( j = 6 \), in which case it was all of \( O_j \)). Notice that this construction preserves edge orientation. For example, \( O_{2\rightarrow 3} \) will not be the same as \( O_{3\rightarrow 2} \). The 20-mer oligonucleotide with the sequence that is Watson-Crick complementary to \( O_i \) was denoted \( O_i^* \) (Fig. 2).

For each vertex \( i \) in the graph (except \( i = 0 \) and \( i = 6 \)) and for each edge \( i \rightarrow j \) in the graph, 50 pmol of \( O_i \) and 50 pmol of \( O_{i\rightarrow j} \), respectively, were mixed together in a single ligation reaction (4). The \( O_i \) oligonucleotides served as splints to bring oligonucleotides associated with compatible edges together for ligation (Fig. 2). Hence the ligation reaction resulted in the formation of DNA molecules encoding random paths through the graph.

The scale of this ligation reaction far exceeded what was necessary for the graph under consideration. For each edge in the graph, approximately \( 3 \times 10^{13} \) copies of the associated oligonucleotide were added to the ligation reaction. Hence it is likely that many DNA molecules encoding the Hamiltonian path were created. In theory, the creation of a single such molecule would be sufficient. As a result, for this graph quantities of oligonucleotides less than an attomole would probably have been sufficient. Alternatively, a much larger graph could have been processed with the picomole quantities used here.

To implement Step 2 of the algorithm, the product of Step 1 was amplified by polymerase chain reaction (PCR) using primers \( O_0 \) and \( O_6 \) (5). Thus, only those molecules encoding paths that begin with vertex 0 and end with vertex 6 were amplified. To implement Step 3 of the algorithm, the product of Step 2 was run on an agarose gel, and the 140-base pair (bp) band (corresponding to double-stranded DNA encoding paths entering exactly seven vertices) was excised and soaked in doubly distilled H\(_2\)O (ddH\(_2\)O) to extract DNA (6). This product was PCR-amplified and gel-purified several times to enhance its purity. To implement Step 4 of the algorithm, the product of Step 3 was affinity-purified with a bioin-avidin magnetic beads system. This was accomplished by first generating single-stranded DNA from the double-stranded DNA product of Step 3 and then incubating the single-stranded DNA with \( O_0 \) conjugated to magnetic beads (7). Only those single-stranded DNA molecules that contained the sequence \( O_1 \) (and hence encoded paths that entered vertex 1 at least once) annealed to the bound \( O_0 \) and were retained. This process was repeated successively with \( O_2, O_3, O_4, \) and \( O_5 \). To implement Step 5, the product of Step 4 was amplified by PCR and run on a gel.

Figure 3 shows the results of these procedures. In Fig. 3A, lane 1 is the result of the ligation reaction in Step 1. The smear with striations is consistent with the construction of molecules encoding random paths through the graph (8). Lanes 2 through 5 show the results of the PCR reaction in Step 2. The dominant bands correspond to the amplification of molecules encoding paths that begin at vertex 0 and end at vertex 6.

Figure 3B shows the results of a "graduated PCR" performed on the single-stranded DNA molecules generated from the band excised in Step 3. Graduated PCR is a method for "printing" results and is performed by running six different PCR reactions with the use of \( O_0 \) as the right primer and \( O_6 \) as the left primer in the ith tube. For example, on the molecules encoding the Hamiltonian path 0 → 1, 1 → 2, 2 → 3, 3 → 4, 4 → 5, 5 → 6, graduated PCR will produce bands of 40, 60, 80, 100, 120, and 140 bp in successive lanes. On the molecules encoding the path 0 → 1, 1 → 3, 3 → 4, 4 → 5, 5 → 6, graduated PCR will produce bands of 40, 60, 80, 100, and 120 bp in successive lanes, where \( x \) denotes the presence of a band in lane 2 (corresponding to the omission of vertex 2 along this path). On molecules encoding the path 0 → 3, 3 → 2, 2 → 3, 3 → 4, 4 → 5, 5 → 6, graduated PCR will produce bands of 60, 80, 100, 120, and 140 bp in successive lanes, where 80-100 denotes that both a 40-bp and an 80-bp band will be produced in lane 3 (corresponding to the double passage of vertex 3 along this path). The most prominent bands in Fig. 3B appear to be those that would arise from the superimposition of the bands predicted for the three paths described above. The bands corresponding to path 0 → 1, 1 → 3, 3 → 4, 4 → 5, 5 → 6 were not expected and suggest that the band excised in Step 3 contained contamination from 120-bp molecules. However, such low weight contamination is not a problem because it does not persist through Step 4. Figure 3C shows the results of graduated PCR applied to the molecules in the final product of Step 4. These bands demonstrate that these molecules encode the Hamiltonian path 0 → 1, 1 → 2, 2 → 3, 3 → 4, 4 → 5, 5 → 6 (9).

This computation required approximately 7 days of lab work. Step 4 (magnetic bead separation) was the most labor-intensive, requiring a full day at the bench. In general, with the use of the algorithm above the number of procedures required should grow linearly with the number of vertices in the graph. The labor required for large graphs might be reduced with use of alternative procedures, automation, or less labor-intensive molecular algorithms.

The number of different oligonucleotides required should grow linearly with the number of edges. The quantity of each oligonucleotide
Fig. 3. Agarose gel electrophoresis of various products of the experiment. (A) Product of the ligation reaction (lane 1), PCR amplification of the product of the ligation reaction (lanes 2 through 8), and molecular weight marker in base pairs (lane 6). (B) Graduated PCR of the product from Step 3 (lanes 1 through 6); the molecular weight marker is in lane 7. (C) Graduated PCR of the final product of the experiment, revealing the Hamiltonian path (lanes 1 through 6); the molecular weight marker is in lane 7.

cleotide needed is a rather subtle graph theoretic question (8). Roughly, the quantity used should be just sufficient to insure that during the ligation step (Step 1) a molecule encoding a Hamiltonian path will be formed with high probability if such a path exists in the graph. This quantity should grow exponentially with the number of vertices in the graph. The molecular algorithm used here was rather naive and inefficient, and as with classical computation, finding improved algorithms will extend the applicability of the method.

As the computation is scaled up, the possibility of errors will need to be looked at carefully. During Step 1, the occasional ligation of incompatible oligonucleotides may result in the formation of molecules encoding "pseudo-paths" that do not actually occur in the graph. Although such molecules may be amplified during Step 2 and persist through Step 3, they seem unlikely to survive the separation in Step 4. Nonetheless, at the completion of a computation, it would be prudent to confirm that a putative Hamiltonian path actually occurs in the graph. During the separation step, molecules encoding Hamiltonian paths may fail to bind adequately and be lost, whereas molecules encoding non-Hamiltonian paths may bind nonspecifically and be retained. The latter problem might be mitigated by more stringent or repeated separation procedures. One might deal with the former problem by periodically applying PCR with primers designed to amplify Hamiltonian paths (in the example above, primers O₁ and O₂). The balanced use of these techniques may be adequate to control such errors.

The choice of random 20-mer oligonucleotides for encoding the graph was based on the following rationale. First, because 420 20-mer oligonucleotides exist, choosing randomly made it unlikely that oligonucleotides associated with different vertices would share long common subsequences that might result in "unintended" binding during the ligation step (Step 1). Second, it was guessed that with high probability potentially deleterious (and presumably rare) features such as severe hairpin loops would not be likely to arise. Finally, choosing 20-mers assured that binding between "split" and "edge" oligonucleotides would involve 10 nucleotide pairs and would consequently be stable at room temperature. This approach was successful for the small graph considered above; however, how to best proceed for larger graphs may require additional research.

What is the power of this method of computation? It is premature to give definitive answers; however, some remarks seem in order. A typical desktop computer can execute approximately 1010 operations per second. The fastest supercomputers currently available can execute approximately 1012 operations per second. If the ligation (concatenation) of two DNA molecules is considered as a single operation and if it is assumed that about half of the approximately 4 \times 10^{14} edge oligonucleotides in Step 1 were ligated, then during Step 1 approximately 10^{14} operations were executed. Clearly, this step could be scaled-up considerably, and 10^{25} or more operations seems entirely plausible (for example, by using micromole rather than picomole quantities).

At this scale, the number of operations per second during the ligation step would exceed that of current supercomputers by more than a thousandfold. Furthermore, hydrolysis of a single molecule of adenosine triphosphate to adenosine monophosphate plus pyrophosphate provides the Gibbs free energy (ΔG = −8 kcal mol⁻¹) for ligation operation (10, 11); hence in principle 1 is sufficient for approximately 2 \times 10^{19} such operations.

This is remarkable energy efficiency, considering that the second law of thermodynamics dictates a theoretical maximum of 34 \times 10^{19} (irreversible) operations per joule at (300 K) (12, 13). Existing supercomputers are far less energy-efficient, executing at most 109 operations per joule. The energy consumed during other parts of the molecular computation, such as oligonucleotide synthesis and PCR, should also be small in comparison to that consumed by current supercomputers. Finally, storing information in molecules of DNA allows for an information density of approximately 1 bit per cubic nanometer, a dramatic improvement over existing storage media such as videotapes, which store information at a density of approximately 1 bit per 10^{12} nm².

Thus, the potential of molecular computation is impressive. What is not clear is whether such massive numbers of inexpensive operations can be productively used to solve real computational problems. One major advantage of electronic computers is the variety of operations they provide and the flexibility with which these operations can be applied. Whereas two 100-digit integers can be multiplied quite efficiently on an electronic computer, it would be a daunting task to do such a calculation on a molecular computer using currently available protocols and enzymes (14).

Nonetheless, for certain intrinsically complex problems, such as the directed Hamiltonian path problem where existing electronic computers are very inefficient and where massively parallel searches can be organized to take advantage of the operations that molecular biology currently provides, it is conceivable that molecular computation might compete with electronic computation in the short term. It is a research problem of considerable interest to elucidate the kinds of algorithms that are possible with the use of molecular methods and the kinds of problems that these algorithms can efficiently solve (12, 15, 16).

For the long term, one can only speculate about the prospects for molecular computation. It seems likely that a single molecule of DNA can be used to encode the "instantaneous description" of a Turing machine (17) and that currently available protocols and enzymes could (at least under idealized conditions) be used to induce successive sequence modifications, which would correspond to the execution of the machine. In the future, research in molecular biology may provide improved techniques for manipulating macromolecules. Research in chemistry may allow for the development of synthetic designer enzymes. One can imagine the eventual emergence of a general purpose computer consisting of nothing more than a single macromolecule conjugated to a ribosomelike collection of enzymes that act on it.

REFERENCES AND NOTES
4. Each oligonucleotide (50 pmol) with 5′-terminal phosphate residue, 5 units of T4 DNA ligase (Boehringer-Mannheim, Germany), ligase buffer, and dATP to a total volume of 100 μl was incubated for 4 hours at room temperature.
5. All PCR amplifications were performed on a Perkin-Elmer (Norwalk, CT) 9600 thermal cycler. For amplification in Step 2, 50 pmol of each primer and 5 units of Taq DNA polymerase (Gibco-BRL, Grand Island, NY) in PCR buffer to a total volume of 50 μl were processed for 35 cycles at 94°C for 1 min and at 30°C for
Efficient Neutralization of Primary Isolates of HIV-1 by a Recombinant Human Monoclonal Antibody


The ability of antibodies to neutralize diverse primary isolates of human immunodeficiency virus–type 1 in vitro has been questioned, with implications for the likely efficacy of vaccines. A recombinant human antibody to envelope glycoprotein gp120 was generated and used to show that primary isolates are not refractory to antibody neutralization. The recombinant antibody neutralized more than 75 percent of the primary isolates tested at concentrations that could be achieved by passive immunization, for example, to interrupt maternal-fetal transmission of virus. The broad specificity and efficacy of the antibody implies the conservation of a structural feature on gp120, which could be important in vaccine design.

Protection from viral disease has traditionally been associated with the preexistence in serum of antibodies capable of neutralizing virus in vitro. Indeed, vaccines are frequently assessed on the ability to elicit neutralizing antibody responses. In the case of human immunodeficiency virus–type 1 (HIV-1), there was initial optimism about the likely efficacy of subunit vaccines given that vaccine sera from several trials were capable of neutralizing laboratory isolates of virus in vitro (1, 2). The grounds for optimism were shaken when it was found that the vaccinee sera were largely ineffective against primary isolates of HIV-1 (2). Some discussion subsequently centered around the validity of standard HIV-1 neutralization assays when applied to primary isolates (2–5). If the assays are meaningful, then they call into question the ability of antibody to effectively neutralize a spectrum of primary isolates. Hyperimmune pooled human plasma preparations are capable of neutralizing a number of primary isolates (3–5), but they represent a combination of specificities that might be difficult to elicit by all except the most complex vaccines (6). A single antibody able to effectively neutralize a broad spectrum of primary isolates would validate the vaccine approach and would provide a template for vaccine design. Furthermore, it would constitute a reagent for passive immunotherapy such as in the interruption of maternal-fetal transmission. We describe here such a human antibody derived by recombinant methods (7).

The generation of the antibody Fab fragment b12 from a combinatorial phage display library has been described previously (8). Fab b12 is directed to the CD4 binding site of gp120 and is a potent neutralizer of the HIV-1 laboratory strains IIIB and MN (9–11). Selection for potency and strain cross-reactivity was achieved through experimental design. The library donor was a long-term asymptomatic U.S. male, presumably infected with a clade B strain of HIV-1; the antigen for affinity selection was gp120 from the atypical IIIB strain, thereby favoring selection of cross-reactive antibodies. A large number of bacterial supernatants containing antibody Fab fragments to gp120 (anti-gp120) at low initial concentrations were directly screened for neutralizing ability to find the most potent Fabs. Although Fab b12 is capable of neutralizing some primary isolates (12), the corresponding whole antibody molecule is likely to be more effective. Therefore, Fab b12 was converted to a whole immunoglobulin G1 (IgG1) molecule by cloning the variable region of Ig heavy chain (VH) and light chain genes into a vector created for high-level mammalian expression (13). The whole antibody IgG1

* To whom correspondence should be addressed.

27 May 1994; accepted 19 September 1994