Solution of a 20-Variable 3-SAT Problem on a DNA Computer

Ravinderjit S. Braich,1 Nickolas Chelyapov,1 Cliff Johnson,1 Paul W. K. Rothemund,2 Leonard Adleman1*

A 20-variable instance of the NP-complete three-satisfiability (3-SAT) problem
was solved on a simple DNA computer. The unique answer was found after
an exhaustive search of more than 1 million (220) possibilities. This computational
problem may be the largest yet solved by nonelectronic means. Problems of this
size appear to be beyond the normal range of unaided human computation.

The vast parallelism, exceptional energy ef-
icity, and extraordinary information den-
sity inherent in molecular computation have
raised the possibility that molecular comput-
ing have been proposed (1, 2, 7–17), and of these, several have been explored ex-
perimentally and found to be feasible (1, 6, 8, 13, 14, 16–19).

The 3-SAT problem is an NP-complete computational problem (20) for which the
fastest known sequential algorithms require ex-
ponential time. The problem became the bench-
mark for testing the performance of DNA com-
puters, after Lipton (2) demonstrated that it was
well suited to take advantage of the parallelism
afforded by molecular computation. A group
led by Smith (14) used surface-based chemistry
to solve a four-variable (16 possible truth as-
signments) instance of the problem. Yoshida
and Suyama (18) also solved a four-variable
instance using a DNA program implementing
a breadth first search. Sakamoto et al. (13)
solved a six-variable (64 possible truth as-
signments) problem using hairpin DNA. A
group led by Landeber (16) used RNA to
solve an instance of a nine-variable (512
possible truth assignments) satisfiability
problem related to the “Knights Problem” in
chess. Here, a 20-variable (1,048,576 pos-
sible truth assignments) instance of the 3-SAT
problem is solved.

In the present study, the architecture em-
ployed is related to the Sticker Model described
by Roweis et al. (9). The Sticker Model uses
two basic operations for computation: separa-
tion based on subsequence and application of
stickers. Only separations are used in the
current study. Separations are carried out using
oligonucleotide probes immobilized in poly-
acylamide gel–filled glass modules. Infor-
mation-carrying DNA strands are moved
through the modules by electrophoresis. Strands with subsequences complementary to
those of the immobilized probes hybridize and
are retained in the module; strands with-
out complementary subsequences pass
through the module relatively unhindered
(21). Captured strands are released from the
probes by running electrophoresis at a tem-
perature higher than the melting temperature
of the probe/complement duplex. Released
strands may be transported via electrophore-
sis to new modules for further separations.

Using electrophoresis to transport DNA
strands between gel-filled glass modules
results in a computer that is “dry” and
potentially automatable. Because covalent
bonds are neither made nor broken during
separations, DNA strands and glass mod-
ules are potentially reusable for multiple
computations.

The Boolean formula. The input to the
computation was a 20-variable 24-clause
3-conjunctive normal form (3-CNF) formula,
Φ (Fig. 1A). To make the computation as
challenging as possible, Φ was designed to
have a unique satisfying truth assignment
(Fig. 1B). This design has endowed Φ with
an iterative syntactical structure that can be
seen on close inspection. It is important to
note, however, that the DNA computation
undertaken here made no use of this structure,
and instead, exhaustively searched all 220
(1,048,576) possible truth assignments in the
process of finding the unique satisfying as-
signment. Hence, it is reasonable to assume
that any 20-variable 24-clause 3-CNF for-
mla would have been just as readily solved.

The library. To represent all possible
truth assignments, a Lipton encoding (2) was
used. For each of the 20 variables xk (k = 1 . . .
20), two distinct 15 base “value sequenc-
es” were designed: one representing true
(1), 6, 8, 13, 14, 16–19).

Fig. 1. The computational problem. (A) 20-variable 3-CNF Boolean formula Φ. The symbol “¬” indicates “not.” (B) The unique truth assignment satisfying Φ.
gies (IDT), Skokie, IL] and used as probes during separation operations.

To reduce errors in computation, sequences were designed to discourage intra- and interlibrary strand hybridization and unintended probe-library strand hybridization. To achieve these goals, sequences were generated to satisfy previously reported constraints (19). In particular, G’s do not occur in value sequences.

Synthesis of long molecules on automated DNA synthesizers can be inefficient. To avoid such inefficiencies, creation of the full library began with the synthesis of two “half-libraries,” one for variables $x_1$ through $x_{10}$ (the left half-library) and one for variables $x_1$ through $x_{20}$ (the right half-library). Half-libraries were synthesized on a dual column ABI 392 DNA/RNA Synthesizer (Applied Biosystems, Foster City, CA) at a 0.2-μmol scale using polyethylene-based solid support. A mix-and-split combinatorial synthesis technique was used (16). Briefly, for the left half-library, oligonucleotides with sequences $X_{10}^t$ and $X_{10}^r$ were synthesized in separate columns. The columns were removed from the synthesizer and opened. The solid-support beads were mixed together and divided into halves, which were then put into separate columns. The columns were closed and synthesis restarted with sequences $X_{10}^t$ and $X_{10}^r$ in separate columns. This process was repeated until all 10 variables had been processed. A similar process was used for the right half-library.

To assess the degeneracy of the half-libraries and the efficacy of capturing half-libraries with Acrydite-modified probes, a gel capture experiment was performed [Web fig. 1 (23)]. For each of the 40 sequences $X_i^t$, $k = 1 \ldots 20$, $Z = T$ or $F$, a “capture layer” was created by adding the corresponding Acrydite-modified probe to polycrylamide gel. An aliquot of the appropriate 5'-[32P]labeled half-library was run through the capture layer via electrophoresis. As expected, for each of the 40 probes, approximately half of the strands of the half-library were captured whereas approximately half passed through. This suggested that probes stayed in the gels and captured half-library strands. This also suggested that half-library strands had subsequences complementary to those of the probes and that, for each variable, the number of half-library strands representing true and the number representing false were approximately equal.

To further test the half-libraries, polymerase chain reaction (PCR) amplifications were run with primer sets: $<X_{10}^t, X_{10}^r>$, $<X_{10}^t, X_{10}^r>$, under standard conditions for 35 cycles using 400 fmol of right half-library as template. Gel analysis showed that products of expected lengths were obtained in all cases (25). This confirmed that subsequences $X_{10}^t$, $X_{10}^r$, $X_{10}^t$, $X_{10}^r$ were present at the expected positions in the left half-library, and that subsequences $X_{10}^t$, $X_{10}^r$, $X_{10}^r$, $X_{10}^t$ were present at the expected positions in the right half-library.

The 300-oligomer (300-mer) full library was created from the two half-libraries using a polymerase extension method similar to that described in (26). Subsequently, two stages of PCR amplification were performed to produce the quantity of full library required for the computation. Gel analysis of the final product showed only the presence of a band corresponding to 300 base pairs (bp) (Fig. 2). Spectrophotometric analysis of the full library showed a total of approximately 750 pmol of 300-bp DNA. Note that the full library consisted of library strands duplexed with Acrydite-modified complements. Because all library strands had the same length (300 bases), they were expected to run at expected rates during electrophoresis.

To test the full library, PCR amplifications were run with primer sets: $<X_{10}^t, X_{10}^r>$, $<X_{10}^t, X_{10}^r>$, $<X_{10}^t, X_{10}^r>$, for various $k$. Gel analysis of the resulting products showed bands of the expected lengths (Fig. 3).

The computer and the computational protocol. The computer consisted of an electrophoresis box with a hot chamber and a cold chamber, a glass “library module” filled with polycrylamide gel containing covalently bound full library, and for each of the 24 clauses of $\Phi$, a glass “clause module” filled with polycrylamide gel containing covalently bound probes and designed to capture only library strands encoding truth assignments satisfying that clause (Fig. 4). The computational protocol was as follows:

Step 1: Insert the library module into the hot chamber of the electrophoresis box and the first clause module into the cold chamber of the box. Begin electrophoresis. In theory, during Step 1, library strands melt off their Acrydite-modified complements in the library module and migrate to the first clause module. Library strands encoding truth assignments satisfying the first clause are captured in the capture layer, while library strands encoding nonsatisfying assignments...
run through the capture layer and continue into the buffer reservoir. In particular, those library strands with sequences \(X^3_t\) or \(X^6_{16}\) or \(X^8_{18}\) are retained in the capture layer, whereas those with sequences \(X^1_t\) and \(X^4_{16}\) and \(X^7_{18}\) run through.

Step 2: Remove both modules from the box. Discard the module from the hot chamber. Wash the box and add new buffer. Insert the module from the cold chamber into the hot chamber and migrate to the clause module in the cold chamber. Begin electrophoresis. In theory, during Step 2, library strands melt off their Acryldite-modified probes in the clause module located in the hot chamber and migrate to the clause module in the cold chamber. Library strands encoding truth assignments satisfying the clause associated with the module in the cold chamber will be captured, while library strands encoding non-satisfying assignments will run through the capture layer and continue into the buffer reservoir.

Step 3: Repeat Step 2 for each of the remaining 22 clauses. In theory, at the end of Step 3, the final (24th) clause module will contain only those library strands which have been captured in all 24 clause modules and hence encode truth assignments satisfying each clause of \(\Phi\) and therefore \(\Phi\) itself.

Step 4: Extract the answer strands from the final clause module, PCR-amplify, and "read" the answer.

**Determination of the answer.** Gel was extruded from the final (24th) clause module and soaked in 1 ml of water at 65°C overnight to extract the library strands it contained. The library strands were lyophilized, reconstituted in 200 \(\mu\)l of water, desalted, and recovered in 45 \(\mu\)l of water. This became the answer stock.

For assigning truth values to variables \(x_1\) and \(x_{20}\), 1- \(\mu\)l aliquots of 10-, 20-, 30-, 40-, 50-, 60-, and 100-fold dilutions of the answer stock were PCR-amplified with primer sets: \(<X^1_t, X^3_{20}>\), \(<X^1_t, X^6_{20}>\), \(<X^4_{16}, X^8_{20}>\), \(<X^7_{18}, X^2_{20}>\). Gel analysis of the PCR products for 10-, 20-, 30-, 40-, 50-fold dilutions showed no bands except for primer set: \(<X^1_t, X^3_{20}>\). These primer sets gave only a band corresponding to 300 bp. Based on this, \(x_1\) was assigned to be \(F\) and \(x_{20}\) was assigned to be \(T\). Analysis of the PCR products for the 60- and 100-fold dilutions showed no bands for any primer set (25).

For assigning truth values to the variables \(x_2, x_3, \ldots, x_{19}\), and as a redundant test for the truth value of \(x_{20}\), a \(1 \mu\)l aliquot of the 50-fold dilution of the answer stock was PCR-amplified with primer sets: \(<X^1_t, X^3_{18}>\), \(<X^1_t, X^6_{18}>\), \(<X^4_{16}, X^8_{18}>\), \(<X^7_{18}, X^2_{18}>\), where \(k = 2, 3, \ldots, 20\). Gel analysis showed (Fig. 5) that in each case only one combination of primers gave a band and this band was of the expected length (compare with Fig. 3). On this basis, truth-values were assigned to each variable. These experimentally derived truth-values corresponded to the unique satisfying truth assignment for \(\Phi\) (Fig. 1B).

**Capture-release efficiency.** The following analysis for correct strands, those encoding the unique satisfying truth assignment, was made. Because PCR of 1 \(\mu\)l of a 50-fold dilution of the 45 \(\mu\)l answer stock revealed only the correct truth assignment, it is probable that at a minimum, 50 \(\times\) 45 = 2250 correct strands were in the answer stock. Since the computation began with approximately 500 pmol (3 \(\times 10^{14}\) molecules) of full library, and since approximately 1 in 220 were correct strands, it follows that the probability of a correct strand surviving the entire computation was at least 2250 \(\times (220)^{-1}\) \(= 7.5 \times 10^{-6}\). Because the computation had 24 capture-release steps, it follows that on average, the probability of a correct strand surviving the final capture release step was at least \((7.5 \times 10^{-6})^{(1/24)} = 0.61\). This analysis assumes that PCR was...
We did not implement stickers and, hence, our library strands behaved like fixed memories. With stickers as originally conceived, library strands would act as more powerful write-once memories. Recent research (27) suggests that DNA “strand invasion” might provide a means for the specific removal of stickers from library strands. This could give rise to library strands that act as very powerful read-write memories. Further investigation of this possibility seems worthwhile.

Despite our successes, and those of others, in the absence of technical breakthroughs, optimism regarding the creation of a molecular computer capable of competing with electronic computers on classical computational problems is not warranted. However, molecular computers can be considered in a broader context. They may be useful in specialized environments where, for example, extreme energy efficiency or extraordinary information density is required. They may provide a much-needed means for controlling chemical/biological systems in the same way that electronic computers have provided a means for controlling electrical/mechanical systems. They do provide a focus for the integration of ideas from biology and computation and this can lead to spin-offs, such as the promising work on DNA self-assembly (JJ). They enlighten us about alternatives to electronic computers and studying them may ultimately lead us to the true “computer of the future.”

Most importantly, DNA computers, such as the one presented here, illustrate that biological molecules can be used for distinctly nonbiological purposes. For such purposes, these molecules represent an untapped legacy of 3 billion years of evolution, and there is great potential in their further exploration.

References and Notes
21. Value sequences X(T) and X(T), (k = 1...20). All sequences written 5’ to 3’. X(T) = TTA CAA TCT CTT, X(T) = ATT TCT AAT CTC, X(T) = TTA CAA TCT CTT, X(T) = GCC TAT TAA TAA TAA TAA, (T= TAA TTT CAA TCT, X(T) = TTA CAA CAC TCT, X(T) = TTA CAA TCT CTT.
22. Value sequences X(T) and X(T), (k = 1...20). All sequences written 5’ to 3’. X(T) = TTA CAA TCT CTT, X(T) = ATT TCT AAT CTC, X(T) = TTA CAA TCT CTT, X(T) = GCC TAT TAA TAA TAA TAA, TTA CAA CAC TCT, X(T) = TTA CAA TCT CTT, X(T) = TTA CAA TCT CTT.
23. Web figure 1 is available on Science Online at www.sciencemag.org/cgi/content/full/1069528/DC1.
24. Standard conditions: 50 μl total volume containing 50 mM KCl, 1.5 mM MgCl2, 10 mM Tris (pH 8.8), 200 mM of each dNTP, 10 μl of each primer, and 1 unit of Taq DNA Polymerase, on a GeneAmp PCR System 9700 (Applied Biosystems) with the following temperature profile: 95°C for 15 s, 40°C for 45 s, 72°C for 60 s.
25. R. S. Braich et al., data not shown.
28. Modules were inserted into the electrochemistry box in pairs. The exterior of each module was sheathed with tape at the end furthest from the capture layer. A sleeve was created from the top 6 mm of a 20 μl pipette tip (RT-10, Rainin). The modules were inserted taped end first into opposite ends of the sleeve. To ensure good contact, a 5- to 10-μl aliquot of 5% polyacrylamide solution was placed between them. The modules were manually adjusted to ensure that they touched and were in correct alignment. For a 1 saccharide module, a 1.5-cm-diameter hole was drilled through the plexiglass divider of the electrochemistry box and fitted with a PE Quickdisconnect tube connector (for tubing of inner diameter 1/8 inches to 1/4 inches, Aldrich) from which the nozzles were cut off.
29. We gratefully acknowledge the contributions of M. Goodman, A. Solitari, D. Hwang, K. N. Chelyapov Jr., and the members of our laboratory for molecular DNA. Support was granted by NASA/Jet Propulsion Laboratory, the Defense Advanced Research Projects Agency, the Office of Naval Research, and NSF.
3 January 2002; accepted 5 March 2002
30. Published online 14 March 2002; 10.1126/science.1069528
Include this information when citing this paper.