

---

Molecular Computation by DNA Hairpin Formation

Author(s): Kensaku Sakamoto, Hidetaka Gouzu, Ken Komiya, Daisuke Kiga, Shigeyuki Yokoyama, Takashi Yokomori and Masami Hagiya

Source: *Science*, May 19, 2000, New Series, Vol. 288, No. 5469 (May 19, 2000), pp. 1223-1226

Published by: American Association for the Advancement of Science

Stable URL: <https://www.jstor.org/stable/3075259>

## REFERENCES

Linked references are available on JSTOR for this article:

[https://www.jstor.org/stable/3075259?seq=1&cid=pdf-reference#references\\_tab\\_contents](https://www.jstor.org/stable/3075259?seq=1&cid=pdf-reference#references_tab_contents)

You may need to log in to JSTOR to access the linked references.

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



American Association for the Advancement of Science is collaborating with JSTOR to digitize, preserve and extend access to *Science*

JSTOR

34. T. Irifune and A. E. Ringwood, *Earth Planet. Sci. Lett.* **117**, 101 (1993).
35. S. E. Kesson and A. E. Ringwood, *Chem. Geol.* **78**, 97 (1989).
36. S. Heinemann, T. G. Sharp, F. Seifert, D. C. Rubie, *Phys. Chem. Miner.* **24**, 206 (1997).
37. T. Kato, A. E. Ringwood, T. Irifune, *Earth Planet. Sci. Lett.* **89**, 123 (1988).
38. M. Akaogi, Y. Hamada, T. Suzuki, M. Kobayashi, M. Okada, *Phys. Earth Planet. Inter.* **115**, 67 (1999); T. Irifune, *Nature* **370**, 131 (1994); L.-G. Liu, *Earth Planet. Sci. Lett.* **36**, 237 (1977).
39. A. E. Ringwood and J. F. Lovering, *Earth Planet. Sci. Lett.* **7**, 371 (1970).
40. W. A. Bassett and G. E. Brown Jr., *Annu. Rev. Earth Planet. Sci.* **18**, 387 (1990).
41. The infrared reflectance measurements were conducted using a Bruker IFS-66v interferometer with attached microscope; the apparatus was equipped with a global source, KBr beamsplitter and liquid-nitrogen cooled mid-range MCT (mercury-cadmium-telluride) detector. Spectra are reported with a resolution of  $4\text{ cm}^{-1}$  and were collected from spot sizes ranging between 40 and 90  $\mu\text{m}$  in diameter; typical collection times were 3 to 10 min. Spectra reported are representative examples taken from zones of differing chemistries within the thin sections.
42. K. Omori, *Am. Mineral.* **56**, 1607 (1971).
43. M. Madon and G. D. Price, *J. Geophys. Res.* **94**, 15687 (1989).
44. J. W. Salisbury, L. S. Walter, N. Vergo, D. M. D'Aria, *Infrared (2.1–25 Micron) Spectra of Minerals* (Johns Hopkins Univ. Press, Baltimore, MD, 1991).
45. T. G. Sharp, C. M. Lingemann, C. Dupas, D. Stoffler, *Science* **277**, 352 (1997).
46. D. J. Durben and G. H. Wolf, *Am. Mineral.* **77**, 890 (1992); E. Knittle and R. Jeanloz, in *High Pressure Research in Mineral Physics*, M. H. Manghni and Y. Syono, Eds. (American Geophysical Union, Washington, DC, 1987), pp. 243–250.
47. J. R. Brucato, L. Colangeli, V. Mennella, P. Palumbo, E. Bussolotti, *Astron. Astrophys.* **348**, 1012 (1999).
48. Geobarometers were obtained by linear regressions of pressure versus composition data (Fig. 6A) from experiments by (32, 37, 57). Aware of the possible effect of bulk composition on the barometers, data for garnets used for the barometer ranged from 2 to 11.13 wt% FeO: namely, 8.04 to 11.13 wt% and 6.2 to 6.4 wt% (32); 2.0 to 3.2 wt% (37); 6.09 to 6.92 wt%; and 3.7 to 6.52 wt% (57). This range is similar to that of the Malaita garnet: namely, 2.6 to 15.4 wt% FeO. An independent check with majorite containing 3.12 wt% FeO (58) from experiments at 10 GPa yields  $P$  estimates of 9.37 and 9.74 GPa by the Si and Al+Cr barometers, respectively.
49. A. Suzuki, E. Ohtani, T. Kato, *Science* **269**, 216 (1995).
50. S. E. Haggerty, *Earth Planet. Sci. Lett.* **122**, 57 (1994).
51. P. H. Nixon and R. H. Mitchell, *Mineral. Mag.* **43**, 587 (1980).
52. C. R. Neal and P. H. Nixon, *Trans. Geol. Soc. S. Afr.* **88**, 347 (1985); C. R. Neal, *J. Petrol.* **29**, 149 (1988).
53. H. Tsai, H. O. A. Meyer, J. Moreau, H. J. Milledge, *Mineral Inclusions in Diamond: Premier, Jagersfontein and Finsch Kimberlites, South Africa, and Williamson Mine, Tanzania*, 2nd International Kimberlite Conference, October 1977, Sante Fe, NM, F. R. Boyd and H. O. A. Meyer, Eds. (American Geophysical Union, Washington, DC, 1979), pp. 16–26.
54. A. M. Dzierwowski and D. L. Anderson, *Phys. Earth Planet. Inter.* **25**, 297 (1981).
55. S. P. Grand and D. V. Helmberger, *Geophys. J. R. Astron. Soc.* **76**, 399 (1984).
56. F. J. Simmons, A. R. Zielhuis, R. van der Hilst, *Lithos* **48**, 17 (1999).
57. C. Herzberg and J. Z. Zhang, *J. Geophys. Res.* **101**, 8271 (1996); M. J. Walter, *J. Petrol.* **39**, 29 (1998).
58. S. E. Kesson, A. E. Ringwood, W. O. Hibberson, *Earth Planet. Sci. Lett.* **121**, 261 (1994).
59. R. O. Moore, M. L. Otter, R. S. Rickard, J. W. Harris, J. J. Gurney, *Geol. Soc. Aust. Abstr.* **16**, 409 (1986); T. Stachel and J. W. Harris, *Contrib. Mineral. Petrol.* **127**, 336 (1997); T. Stachel, J. W. Harris, G. P. Brey, *Contrib. Mineral. Petrol.* **132**, 34 (1998); M. L. Otter and J. J. Gurney, *Geol. Soc. Aust. Spec. Publ.* **14**, 1042 (1989); N. V. Sobolev et al., *Russ. Geol. Geophys.* **38**, 379 (1997).
60. Back-scattered electron (BSE) and secondary electron (SE) images were obtained at the CMM at The University of Queensland by JEOL Super Probe and JEOL 6400F field emission scanning electron microscope, respectively.
61. A. M. Hofmeister, T. J. Fagan, K. M. Campbell, R. B. Schall, *Am. Mineral.* **81**, 418 (1996).
62. We thank A. Ewart, Y. Niu, S. Kesson, and H. O'Neill for helpful comments during the course of the research and preparation of the manuscript. We also thank J. Nailon from the CMM at The University of Queensland for patience and help during acquisition of the SEM images. P. Fredricks of the CIDC at Queensland University of Technology kindly provided the Raman spectroscopic confirmation of carbon in Fig. 4A. The manuscript was substantially improved with helpful reviews by H. Green and two anonymous reviewers. The paper is published with permission of the Directors of Solsearch Ltd.

14 February 2000; accepted 2 March 2000

## REPORTS

# Molecular Computation by DNA Hairpin Formation

Kensaku Sakamoto,<sup>1\*</sup> Hidetaka Gouzu,<sup>1</sup> Ken Komiya,<sup>1</sup>  
Daisuke Kiga,<sup>1†</sup> Shigeyuki Yokoyama,<sup>1</sup> Takashi Yokomori,<sup>3</sup>  
Masami Hagiya<sup>2\*</sup>

Hairpin formation by single-stranded DNA molecules was exploited in a DNA-based computation in order to explore the feasibility of autonomous molecular computing. An instance of the satisfiability problem, a famous hard combinatorial problem, was solved by using molecular biology techniques. The satisfiability of a given Boolean formula was examined autonomously, on the basis of hairpin formation by the molecules that represent the formula. This computation algorithm can test several clauses in the given formula simultaneously, which could reduce the number of laboratory steps required for computation.

the number of required laboratory steps. It has been claimed that the self-assembly and the potential to form secondary structures of the molecules are useful for the embodiment of such computers (15–18), but no actual computation of a hard combinatorial problem had been performed in the autonomous manner. Here, we describe a DNA-based solution of the satisfiability (SAT) problem, where the main logic of computation was implemented on the basis of hairpin formation by single-stranded DNA (ssDNA) molecules.

The SAT problem is to find Boolean-value assignments that satisfy the given formula. Each variable is assigned the Boolean value (either 0 or 1); a set of the values (a value assignment) for the variables satisfies the formula if the value of the formula becomes 1. In a subclass of the SAT problem, called conjunctive normal form (CNF)–SAT, Boolean formulas are restricted to the form of  $C_1 \wedge C_2 \wedge \dots \wedge C_n$ , where each  $C_i$  is a “clause” and “ $\wedge$ ” is the logical AND operation. A clause is of the form  $L_1 \vee L_2 \vee \dots \vee L_m$ , where each  $L_j$  is a “literal” and “ $\vee$ ” is the logical OR operation; a literal is either a variable or its negation (if variable  $x$  takes 1 or 0, then its negation, denoted by  $\neg x$ , takes 0 or 1, respectively).

Lipton proposed a DNA-based solution of CNF–SAT (2), which has been demonstrated in experimental studies (5–8). We took a different approach, because autonomous com-

In 1994, Adleman experimentally demonstrated that DNA molecules and common molecular biology techniques could be used to solve hard

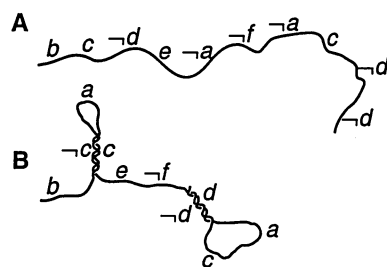
combinatorial problems (1), especially problems involving large searches. The data carried on a number of molecules are processed simultaneously by such techniques, and highly data-parallel computation is achieved.

Adleman's work was later generalized by Lipton (2), whose study encouraged further experimental work, based on Adleman and Lipton's paradigm (3–8). A number of theoretical studies have also emerged in the past 5 years (9, 10), including the idea of autonomous DNA computers (11–14). In these systems, the logic of computation is implemented without external control or interference (other than the regulation of temperature), which could drastically reduce

<sup>1</sup>Department of Biophysics and Biochemistry, <sup>2</sup>Department of Information Science, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. <sup>3</sup>Department of Mathematics, School of Education, Waseda University, 1-6-1 Nishi-waseda, Shinjuku-ku, Tokyo 169-8050, Japan.

\*To whom correspondence should be addressed.

†Present address: Yokoyama Cytologic Project, Exploratory Research for Advanced Technology, Japan Science and Technology Corporation, c/o RIKEN, Hirosawa, Wako-shi, Saitama 351-0198, Japan.



**Fig. 1.** Illustration of the computation based on DNA hairpin formation. (A) The ssDNA representing a satisfying literal string,  $b-c-d-e-a-f-a-c-d$ , stays in the nonhairpin form. (B) A literal string,  $b-c-a-c-e-f-d-c-a-d$ , has pairs of the complementary literals ( $c-c$  and  $-d-d$ ), and the ssDNA representing it forms hairpins.

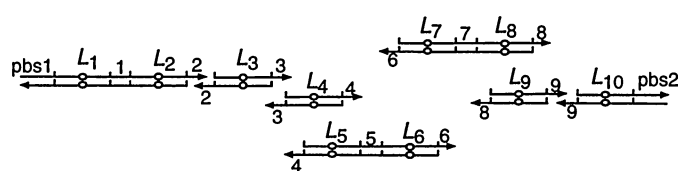
puting must encode the logical constraints into DNA sequences themselves. "Literal strings" were introduced to encode the given formula; they are conjunctions of the literals selected from each clause (one literal per clause). For example, the formula  $(a \vee b) \wedge (\neg a \vee \neg c)$  can be represented by the set of four literal strings,  $a-\neg a$ ,  $a-\neg c$ ,  $b-\neg a$ , and  $b-\neg c$ . A formula is satisfiable if there is such a literal string that does not involve any variable together with its negation (we refer to such strings as "satisfying"), because the literal string gives a value assignment that makes every clause take the value 1 simultaneously, so that the value of the entire formula will become 1. If each variable is encoded with a sequence complementary to that encoding its negation, then the literal strings containing one or more pairs of complementary literals can form hairpins, and the satisfying strings stay in the form without such hairpins ("nonhairpin" form) (Fig. 1).

Thus, the following algorithm solves the CNF-SAT problem. (i) Generate the literal strings according to the given formula. This step is implemented by a ligation reaction, which concatenates the literals. (ii) Allow ssDNA molecules, each representing a literal string, to form hairpins. This step performs the main logic of computation only by regulating the temperature. Even enzymes are not necessary. (iii) Remove the hairpin-forming molecules. The remaining molecules represent the satisfying literal strings, which can be identified with the solutions (value assignments) to the problem. Molecular biology techniques for this step were developed in the present study.

The first two steps are implemented autonomously, and each step processes several clauses simultaneously; the third step may be more laborious. The major difference between our algorithm and the previous molecular algorithms for solving the SAT problem is that our algorithm involves no sequence-specific operations that must be repeated more times for larger problems.

To test the feasibility of the above algorithm, we addressed a six-variable 10-clause

**Fig. 2.** Illustration of "literal units" in the course of the assembly into a 10-clause or full-length literal string. The arrows represent ssDNA molecules; their direction is from the 5' to 3' ends.  $L_k$  indicates the literal from clause  $k$  ( $1 \leq k \leq 10$ ), and the numbers at both ends of each literal indicate the linker numbers. The circles in the middle of the literals represent the Bst NI sites, and pbs1 and pbs2 indicate the primer binding sites for PCR.



instance of 3-SAT, where each clause contains (at most) three literals. Even the 3-SAT problem belongs to NP-complete, a class of computational problems that are hard to solve (19).

The Boolean formula is  $F = (a \vee b \vee \neg c) \wedge (a \vee c \vee d) \wedge (a \vee \neg c \vee \neg d) \wedge (\neg a \vee \neg c \vee d) \wedge (a \vee \neg c \vee e) \wedge (a \vee d \vee \neg f) \wedge (\neg a \vee c \vee d) \wedge (a \vee c \vee \neg d) \wedge (\neg a \vee \neg c \vee \neg d) \wedge (\neg a \vee c \vee \neg d)$ . There exists the unique solution  $(a, b, c, d, e, f) = (0, 1, 1, 0, 1, 0)$ , which is identified with 24 satisfying literal strings out of  $3^{10}$  (59,049) possible literal strings.

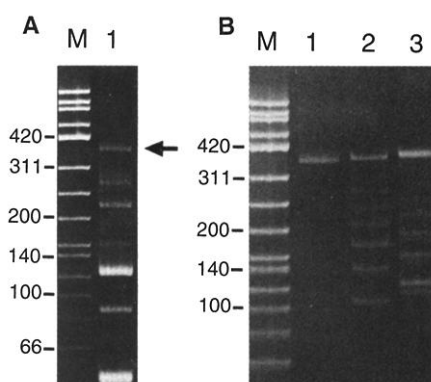
For generating a pool of these possible strings (20), the DNA for each literal in clause  $i$  has linker  $i-1$  (or a primer-binding site, pbs1) on the left and linker  $i$  (or pbs2) on the right (Fig. 2). Thus, a literal string is a linear assembly of 10 literal DNA molecules, where 30-base literal sequences are concatenated, through the 4-base linkers, exclusively with those from the neighboring clauses. The literal DNA was prepared, in the form of double-stranded DNA (dsDNA) with the protruding 5' ends as the linkers, by digesting longer dsDNA molecules, including three to four literals, into each literal DNA. The enzyme that we used for this digestion, Bst XI, recognizes 5'-CCAWNNNN-WTGG-3', where W denotes either A or T and N denotes any base. NNNN covers the linker, and thus each literal sequence is of the form 5'-WTGG...CCAW-3'. The Bst NI site, CCAGG, is also contained for the hairpin-removing step (as described below). The remaining 17 nucleotides were designed to meet the criteria commonly used for DNA computers, and the linker sequences are not palindromic for avoiding self-concatenation.

These DNA molecules for 30 literals from the 10 clauses (3 literals per clause) were mixed in a test tube and concatenated with DNA ligase, and the ligation products were separated by gel electrophoresis (Fig. 3A). The final yield (1.5 to 3 ng) of the full-length molecules was enough to contain  $>10^4$  molecules for every literal string. The quality of this pool of the literal strings (pool 0) was examined by sequence analysis (Table 1). The literals occurred roughly at random, except for the apparent bias found in clauses 5 and 7; the claim of unbiased concatenations for our ligation scheme requires further verification. "Erroneous" literals, that is, literals absent in the corresponding clauses of the formula, were not found; this is reinforced

by the additional data collected during computations (some bias may appear in these data because of the implemented operations).

The logic of our computation, embodied as hairpin formation, was performed by regulating the temperature of a DNA solution in which the literal-string DNA molecules are dissolved at low concentration, allowing the dsDNA to melt and refold intramolecularly. In the absence of available means for this hairpin-removing process, we developed the following two techniques. One is destructive: enzymatic digestion is used to remove hairpin DNA. The other is similar to the polymerase chain reaction (PCR) and increases the population of non-hairpin molecules in a given DNA pool.

The literal sequences have a restriction enzyme site for Bst NI in the middle of the sequence, so that the double-stranded regions of a hairpin molecule become susceptible to the enzyme (20). The selectivity for hairpin DNA requires that the restriction sites alone do not form stable base pairs with each other, regardless of the complementarity between the literals. To check this point, we performed a preliminary experiment with the following ssDNA molecules: molecule A, consisting of the sequences for literal  $a$ , linker 1, and literal  $b$  in



**Fig. 3.** Electrophoresis of products after each operation. (A) The concatemers of the literal units (lane 1). The band corresponding to the full-length literal strings is indicated by the arrow. Lane M is a size marker. (B) The full-length literal strings, biotinylated at one end (pool 0) (lane 1); the products after the Bst NI digestion (pool 1) (lane 2); and the products after further processing by ePCR (pool 3) (lane 3). Lane M is a size marker. The electrophoresis was conducted in 8% polyacrylamide gels stained with ethidium bromide.



**Table 1.** Multiplicity of the occurrence of each literal in pool 0. The data of 223 literals were obtained from pool 0, and 730 additional data (parentheses) were collected during the computations using pool 0 (31). "c1" to "c10" indicate clauses 1 to 10, respectively. " $k_1$ ," " $k_2$ ," and " $k_3$ " represent the first, second, and third literals from each clause, respectively.  $k_1$  represents  $a$  for clauses 1, 2, 3, 5, 6, and 8 and  $\neg a$  for clauses 4, 7, 9, and 10.  $k_2$  represents  $b$  for clause 1;  $c$  for clauses 2, 7, 8, and 10;  $\neg c$  for clauses 3, 4, 5, and 9; and  $d$  for clause 6.  $k_3$  represents  $\neg c$  for clause 1;  $d$  for clauses 2, 4, and 7;  $\neg d$  for clauses 3, 8, 9, and 10;  $e$  for clause 5; and  $\neg f$  for clause 6. "x" represents any erroneous literal that is not included in the corresponding clauses of the formula  $F$ .

Literal	Multiplicity of occurrence									
	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10
$k_1$	7 (26)	7 (17)	4 (30)	7 (17)	12 (24)	3 (8)	15 (40)	4 (21)	6 (29)	15 (32)
$k_2$	5 (31)	3 (12)	9 (27)	9 (24)	2 (14)	6 (26)	2 (3)	6 (20)	5 (11)	5 (25)
$k_3$	7 (16)	10 (44)	6 (16)	15 (32)	16 (35)	9 (39)	4 (30)	7 (32)	8 (33)	9 (16)
x	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

this order ( $a$ -1- $b$ ), and molecule B, in the form of  $a$ -1- $\neg a$ . Figure 4A shows that molecule B actually forms a hairpin, because it migrates on the gel much faster than expected from its length, and molecule A stays in the nonhairpin form. It was found that only molecule B is susceptible to the restriction enzyme. The small population of molecule B remaining undigested was probably due to incomplete "deprotection" after chemical synthesis.

The second technique depends on the inability of DNA polymerase to duplicate a DNA template that forms stable hairpins (20). If the concentration of the template is kept low enough to allow hairpin formation, rather than intermolecular hybridization, then PCR is expected to amplify only nonhairpin molecules. To maintain such a low DNA concentration, we diluted the reaction mixture twice after each PCR cycle. The common PCR protocol was then utilized to recover the DNA. To test the feasibility of this technique (which we call "exclusive PCR" or "ePCR"), we processed dsDNA molecules C (pbs1- $a$ -1- $c$ -2- $\neg d$ -pbs2) and D (pbs1-7- $a$ -8- $\neg d$ -9- $\neg a$ -pbs2) by ePCR. As expected, molecule D (having a pair of  $a$  and  $\neg a$ ) was recovered in a much smaller amount than molecule C (Fig. 4B). This ePCR technique may be also effective for the hairpin DNA with a few point mutations in the restriction site; for such

DNA, the enzymatic digestion protocol does not work.

These techniques were applied to processing the pool of the literal strings, pool 0. The enzymatic digestion was performed first. The literal strings in the dsDNA form were biotinylated by amplification with the primers [either of which was biotinylated at its 5' end (Fig. 3B)] and then bound on the beads for conversion into the single-stranded form by alkali treatment. Another objective of this immobilization was to prevent the intermolecular hybridization between the literal strings during the renaturation process that facilitates hairpin formation. The ssDNA molecules on the beads were digested twice with the restriction enzyme Bst NI. The molecules that remained undigested were recovered by PCR from the beads. The band corresponding to the undigested DNA appeared on the gel, together with a ladder of the bands (Fig. 3B). It is likely that, during PCR, the digested molecules hybridized with different sites on the undigested ones and were extended with them as the template, thus producing literal strings of various lengths (21). Only the full-length strings obtained here (pool 1) were purified and subjected to the sequence determination. No satisfying literal strings were found among 11 clones picked up randomly. Then, this destructive process was performed once more on pool 1, followed by the ePCR

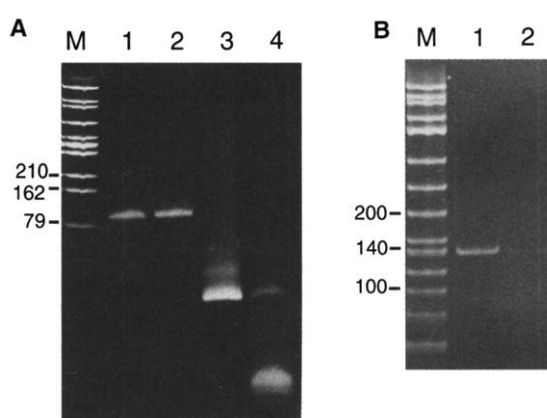
processing of 10 cycles, to generate pool 2; only one satisfying string was found among 16 clones. Instead of the second round of the destructive process, ePCR (20 cycles) (22) was directly performed on pool 1 to generate pool 3 (Fig. 3B), where we found six satisfying literal strings (five different ones) out of 37 sequenced clones (Table 2). All of these strings were identified with the correct unique solution.

The final result with pool 3 demonstrated the feasibility of the algorithm based on the intrinsic property of ssDNA to form hairpins, although the rate of hitting the real solution was much lower than that of the previous successful computations (6-8), which used common molecular biology techniques. Furthermore, the literal strings in pools 2 and 3 had accumulated point mutations around the restriction site, preventing further enrichment by digestion and weakening the effectiveness of ePCR. Increasing the fraction of satisfying strings in the final pool would require improvements in the fidelity of PCR or in the effectiveness of ePCR.

The instance solved here is comparable in size to the SAT problems solved in the previous computations (6-8); for example, Faulhammer and co-workers recently solved a nine-variable and five-clause instance (8). In these computations, each clause was examined by a few laboratory steps, whereas our computation processed all clauses simultaneously. In the previous computations, the long sequence of laboratory steps introduced many chances for errors to occur. The major drawback with our method is an inefficiency with respect to the required amount of DNA; the algorithms based on Lipton's method only require  $2^n$  or less molecules for encoding the value assignments with  $n$  variables, whereas we generated  $3^m$  literal strings for  $m$  clauses, where  $m$  is four times  $n$  for the hardest instances (23). This inefficiency comes from the necessity of a molecular representation of the given logical constraints.

For a scale-up of our method, we should also enumerate the value assignments as the literal strings comprising either of the two literals for each variable, which are then to be randomly concatenated with the literal strings that represent the given formula, being subjected to the hairpin-based computation. This modification

**Fig. 4.** (A) Electrophoresis of ssDNA molecules subjected to the Bst NI digestion. Molecules A (lanes 1 and 2) and B (lanes 3 and 4) were renatured (lanes 1 and 3) and then subjected to the enzymatic digestion (lanes 2 and 4). Lane M is a size marker. The electrophoresis was conducted in the denaturing 8% polyacrylamide gel stained with ethidium bromide. (B) Electrophoresis of the products of ePCR. Molecules C (lane 1) and D (lane 2) were each used as the template for a 10-cycle ePCR, followed by a 15-cycle PCR to recover the DNA. An aliquot (5  $\mu$ l) of the product was applied to the 8% polyacrylamide gel electrophoresis, stained with ethidium bromide. Lane M is a size marker.



**Table 2.** The satisfying literal strings obtained from pool 3. The numbers in the parentheses indicate the multiplicity of the occurrence of each string.

Literal string	Sequence of literals (clauses 1 to 10)
S1	$b-c-\neg d-\neg a-e-\neg f-\neg a-c-\neg a-\neg a$ (1)
S2	$b-c-\neg d-\neg a-e-\neg f-\neg a-c-\neg a-\neg d$ (1)
S3	$b-c-\neg d-\neg a-e-\neg f-\neg a-\neg d-\neg a-\neg a$ (1)
S4	$b-c-\neg d-\neg a-e-\neg f-\neg a-\neg d-\neg a-\neg d$ (2)
S5	$b-c-\neg d-\neg a-e-\neg f-\neg a-\neg d-\neg d-\neg a$ (1)

could reduce the required DNA amount, together with use of the "breadth-first" search proposed for large SAT problems (6, 24), where our algorithm must be repeated by the number of the clauses, although several clauses can be processed simultaneously. Upon tackling large instances, PCR may introduce serious errors into computation because all sequences may not be amplified with the same efficiency (7). However, a variation in sequence was still retained after our computation including as many as 70 PCR cycles (25). Further study is necessary for unraveling the nature of the possible bias during PCR.

Another drawback with our method is the incompleteness of our understanding of the nature of hairpin molecules. The limit to the available length of hairpin remains to be determined; the available data suggest that an intact double helix of 30 base pairs is stable enough for closing a 2000-base hairpin loop (26–28). Sequence design or experimental conditions that ensure the hairpin formation and thus reduce the error rate also remain as the major issue for our algorithm. On the other hand, "negative" errors (loss of "solution" molecules) were successfully controlled, because the satisfying strings were finally obtained, despite their small population (0.04% of the total molecules) in the initial DNA pool. This approach for molecular computing, described on a theoretical basis (29) and then implied in some previous experiments (1, 3, 30), sheds a new light on the potential of DNA, not restricted to a carrier of information.

#### References and Notes

1. L. M. Adleman, *Science* **266**, 1021 (1994).
2. R. J. Lipton, *Science* **268**, 542 (1995).
3. Q. Ouyang, P. D. Kaplan, S. Liu, A. Libchaber, *Science* **278**, 446 (1997).
4. N. Morimoto, M. Arita, A. Suyama, in *DNA Based Computers III*, H. Rubin and D. H. Wood, Eds., vol. 48 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* (American Mathematical Society, Providence, RI, 1999), pp. 193–206.
5. A. G. Frutos et al., *Nucleic Acids Res.* **25**, 4748 (1997).
6. H. Yoshida and A. Suyama, in *Preliminary Proceedings of the Fifth International Meeting on DNA Based Computers*, E. Winfree and D. K. Gifford, Eds. (American Mathematical Society, Providence, RI, 1999), pp. 9–20.
7. Q. Liu et al., *Nature* **403**, 175 (2000).
8. D. Faulhammer, A. R. Cukras, R. J. Lipton, L. F. Landweber, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1385 (2000).
9. G. Paun, G. Rozenberg, A. Salomaa, *DNA Computing: New Computing Paradigms* (Springer-Verlag, Berlin, 1998).
10. M. H. Garzon and R. J. Deaton, *IEEE Trans. Evol. Comput.* **3**, 236 (1999).
11. E. Winfree, in *DNA Based Computers*, R. J. Lipton and E. B. Baumvol, Eds., vol. 27 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* (American Mathematical Society, Providence, RI, 1996), pp. 199–221.
12. M. Hagiya, *New Generation Comput.* **17**, 131 (1998).
13. J. Reif, in *DNA Based Computers III*, H. Rubin and D. H. Wood, Eds., vol. 48 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* (American Mathematical Society, Providence, RI, 1999), pp. 217–254.
14. T. Yokomori, in *Preliminary Proceedings of the Fifth International Meeting on DNA Based Computers*, E. Winfree and D. K. Gifford, Eds. (American Mathematical Society, Providence, RI, 1999), pp. 153–167.
15. E. Winfree, X. Yang, N. Seeman, in *DNA Based Computers II*, L. Landweber and E. Baum, Eds., vol. 44 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* (American Mathematical Society, Providence, RI, 1999), pp. 191–213.
16. E. Winfree, F. Liu, L. A. Wenzler, N. C. Seeman, *Nature* **394**, 539 (1998).
17. T. H. LaBean, E. Winfree, J. H. Reif, in *Preliminary Proceedings of the Fifth International Meeting on DNA Based Computers*, E. Winfree and D. K. Gifford, Eds. (American Mathematical Society, Providence, RI, 1999), pp. 121–138.
18. K. Sakamoto et al., *Biosystems* **52**, 81 (1999).
19. J. E. Hopcroft and J. D. Ullman, *Introduction to Automata Theory, Languages and Computation* (Addison-Wesley, Reading, MA, 1979).
20. The details of the DNA sequences and experimental manipulations are available at [www.sciencemag.org/feature/data/1046140.shl](http://www.sciencemag.org/feature/data/1046140.shl).
21. The priming of DNA synthesis by digested literal strings probably produced the "recombinant" strings of the full length, which cannot be distinguished from the undigested DNA. By the definition of literal strings, this unusual priming of DNA synthesis does not generate such literal strings that contain any erroneous literal that is absent in the corresponding clauses of the formula. The profile of the band ladder (the band intensity drastically decreased with the DNA length) indicated that the extent of this contamination was negligible for the size of the present instance. However, rigorous estimation may be necessary for larger instances, and molecular biology techniques could cope with this issue; for example, dideoxynucleotides can be added to the 3' end of the digested strings so that they cannot prime DNA synthesis. Recombination between the "solution-candidate" molecules during PCR amplification has been reported (8). This is another possible cause of the band ladder, although PCR amplification of pool 0 by itself did not produce any band ladder (Fig. 3B, lane 1).
22. To avoid the loss of DNA, we amplified the product once after 10 cycles of ePCR by the normal PCR protocol. This amplified DNA was then applied to another 10 cycles of ePCR and was finally recovered by PCR with the primers pbs1 and pbs2c that have additional bases of the Eco RI and Hind III sites, respectively, at the 5' end. The shift of the bands to the origin of the electrophoresis was due to the use of these longer PCR primers. The band ladder produced after ePCR was caused probably because the purification of the full-length products after the enzymatic digestion was incomplete, and the contamination of the shorter molecules was amplified during ePCR and the succeeding PCR. The recovered full-length DNA was purified and then digested with the restriction enzymes for cloning in an appropriate vector.
23. D. Mitchell, B. Selman, H. Levesque, in *Proceedings of the 10th National Conference on Artificial Intelligence (AAAI-92)*, P. Rosenbloom and P. Szolovits, Eds. (AAAI Press/MIT Press, Cambridge, MA, 1992), pp. 459–465.
24. M. Ogihara and A. Ray, in *DNA Based Computers III*, H. Rubin and D. H. Wood, Eds., vol. 48 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* (American Mathematical Society, Providence, RI, 1999), pp. 255–264.
25. K. Sakamoto et al., data not shown.
26. C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry Part III: The Behavior of Biological Macromolecules* (Freeman, San Francisco, 1980).
27. K. J. Breslauer, R. Frank, H. Blocker, L. A. Marky, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3746 (1986).
28. D. H. Jones and S. C. Winistorfer, *Nucleic Acids Res.* **20**, 595 (1992).
29. C. H. Bennett, *Int. J. Theor. Phys.* **21**, 905 (1982).
30. F. Guarnieri, M. Fliss, C. Bancroft, *Science* **273**, 220 (1996).
31. These additional data came from the 64 sequenced clones picked up from pools 1 through 3 and 9 clones from other prototype computations. These 73 clones include 71 different literal strings.
32. We thank A. Nishikawa for designing the DNA sequences used in the experiments, E. Winfree for the critical reading of the manuscript, and Y. Husimi for valuable discussions on the physicochemical nature of hairpins. This work was supported in part by the Japan Society for the Promotion of Science under the Research for the Future program (JSPS-RFTF 96100101).

12 October 1999; accepted 3 April 2000

## Coalescence of Single-Walled Carbon Nanotubes

M. Terrones,<sup>1\*</sup> H. Terrones,<sup>1,2†</sup> F. Banhart,<sup>3‡</sup> J.-C. Charlier,<sup>4</sup> P. M. Ajayan<sup>5</sup>

The coalescence of single-walled nanotubes is studied in situ under electron irradiation at high temperature in a transmission electron microscope. The merging process is investigated at the atomic level, using tight-binding molecular dynamics and Monte Carlo simulations. Vacancies induce coalescence via a zipper-like mechanism, imposing a continuous reorganization of atoms on individual tube lattices along adjacent tubes. Other topological defects induce the polymerization of tubes. Coalescence seems to be restricted to tubes with the same chirality, explaining the low frequency of occurrence of this event.

The driving force for coalescence of particles and supramolecular structures is the reduction in surface and strain energy, however high energy barriers may have to be overcome, particularly in the case of supramolecular systems where factors such as bond rigidity and rotation, structural geometry, and atomic mobility play an important role. In this context, mass spectrometric measurements on hot fullerene vapors (1) and microscopy studies (2, 3) have provided some evidence for fullerene coalescence. However, single-walled carbon nanotubes (SWNTs)

are large molecular assemblies consisting of several thousand atoms. Their basic structure is made of seamless cylinders of sp<sup>2</sup>-like C. Recent reports suggest the possibility of coalescence between these massive molecular structures when they are annealed at high temperatures in the presence of H<sub>2</sub> (4, 5).

Here, we show the coalescence of SWNTs by in situ irradiation and heating in a high-resolution transmission electron microscope (HRTEM) and consider in detail the possible