International Journal of Information Technology and Knowledge Management
January-June 2011, Volume 4, No. 1, pp. 191-194

BIOLOGICAL COMPUTER MODEL TO SOLVE NP COMPLETE PROBLEM

Shalini Rajawat[1], Vijay Singh Rathore[2] & Naveen Hemrajani[1]

We have designed a special DNA computer model to NP complete problem. The main body of this DNA computer was polyacrylamide gel electrophoresis consisting of three parts, i.e. a melting region, an unsatisfied solution region and a solution region. The polyacrylamide gel was connected with a controllable temperature device, and the relevant temperature for the three parts was denoted by Tm1, Tm2 and Tm3 respectively. The library generating method had two parts: one, called storage library, was composed of all possible colorings represented by DNA strands, and the other was composed of many probes, were the relevant many complementary strands, representing the graph structure information.
Keywords: DNA, NP Complete Problem, DNA Computing

1. Fundamental Principle

This computer is composed of two parts: the hard ware and the soft ware. The hard ware was polyacrylamide gel electrophoresis containing three regions: The first region is called the melting region or melting box, in which dsDNA molecules can melt into ssDNA molecules. One of them is fixed on the polyacrylamide, and the other is kept on electrophoreses. The fixing technique was either AcryditeTM or using magnetic beads. For the theory and applications of AcryditeTM.We put all the dsDNA representing the possible solutions of the problem on a board, and name the board A. The second region is called the unsatisfied solution sticking region or then satisfied solution box, on which we put a board with lots of probes. As a matter of fact, the probes are some ssDNA sequences equal to the adjacency matrix containing all the information of a graph, we name this board B. It should be noted that the purpose of using these probes on the board B is to eliminate the solutions that do not satisfy the problem. The third region is called the solution region or solution box. Each ssDNA here represents a solution satisfying the problem. It should be emphasized that there are two boards, board A and board B. Board A, named storage library, is a set that contain all the possible solutions of the problem. The storage library is made on a glass full of polyacrylamide gel, and the "data" contained in the storage library are all dsDNA; The board B is composed of lots of ssDNA, which represent the information of a given graph, so different graphs have different information on board B. The main operation of this DNA computer was polyacrylamide gel electrophoresis.The polyacrylamide is connected with a controllable temperature apparatus to keep the temperature of the three regions at Tm1, Tm2 and Tm3, respectively. When the temperature of the melting region is Tm1, one dsDNA molecule could melt into two ssDNA molecules. And one was fixed on the polyacrylamide (board A), and the other unfixed would move to the unsatisfied solution region under the effect of electric field. Then we shift the temperature of the unsatisfied solution region to Tm2, and it is at right temperature that the probes could hybridize with ssDNA on board B (according to the experiment, the length of unsatisfied solution box can be adjustable). The temperature of the solution region is controlled at Tm3, and the ssDNA that reaches this box is the right solution to the problem.

1.1. The Principle of Library Design

We always assume that the treated graph in the graphical vertex-coloring DNA computer is n -order. We only consider the 3-coloring problem of a graph in this paper which is a np complete problem. Of course, the principle of the graphical vertex-coloring DNA computer is the same if the number of color $k \geq 4$. In this paper, we assume the vertex of the graph to be V(G) = {1,2,..,n}, and each vertex can have three different colors, ri, bi, yi, i =1,2,…,n. Here, ri means that the vertex i is red, while bi is blue, yi is yellow. Each color is represented by ssDNA with a length l Note that l is determined by such factors as the number of vertices and the number of colors. Actually,for the problem with fewer variables, the length l of ssDNA should satisfy $15 \leq l \leq 30$. Here, using DNA strands with 15 bases to represent each variable (coloring),we got a sequence set $T = \{x_1x_2..x_n; x_i \in \{ri,bi, yi\}; i = 1,2,\ldots, n\}$. (1) The number of the DNA sequences in this set, which represents all the possible

[1]Department of Computer Science & Engineering, Suresh Gyan Vihar University, Jaipur, India

[2]Department of Computer Science & Engineering, Karni College, Jaipur, India

Email [1]rajawatshalini@yahoo.co.in, [2]vijaydiamaond@gmail.com, [1]naven_h@yahoo.com

colorings, is 3n. The length of each sequence is n × l, meaning that each sequence contains n × l bases. To construct the storage library of this DNA computer, we have to design each DNA sequence (i.e. encoding) in the storage library, and then with the codes, the DNA sequences are synthesized by DNA molecular synthesis apparatus, such as ABI 3900. These sequences can be fixed on board A.

1.2. The Encoding for the Storage Library

The codes can be obtained after the following three steps: (i) Finding all biological constraints; (ii)converting them into mathematic constraints; (iii) giving the encoding algorithms and corresponding codes. Just as mentioned above, board A contains many DNA sequences representing all possible colorings of the vertex 3-coloring problem. The length of each sequence is n × l, i.e. each sequence contains n × l bases. Actually, we encode every vertex with a 15-base sequence. we give the constraints as follows:

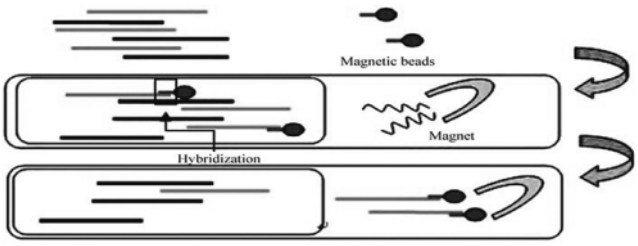
1. Library sequences contain A's, T's and C's.
2. All library and probe sequences have no occurrence of 5 or more consecutive identical bases.
3. Each probe sequence has at least 8 mismatches with all 15 bases alignment of any library sequence (except for its matching value sequence).
4. Each sequence of library sequence has at least 8 mismatches with all 15 bases alignment of its own or any other library sequence.
5. No probe sequence has a run of more than 7 matches with any 8 bases alignment of any library sequence (except for its matching value sequence).
6. No library sequence has a run of more than 7 matches with any 8 bases alignment of its own or any other library sequence.
7. Every probe sequence has 5 or 6 Gs in its sequence.
8. The hamming distance of every sequence is 8, m1-m10 ≥ 5, m6-m15 ≥ 5, m1-m5 ≥ 4, and m11-m1 5 ≥ 4. Here, m1-m10 represents the hamming distance from 1 to 10. As a paradigm, the codes of any graph with 20 vertices to solve 3-coloring problem are given as follows:

The Design of Coloring Library

A coloring scheme of an n-order graph is denoted by sequence $x1x2\ldots x^n$, which shows that the color of vertex i is xi, xi € {ri, bi, yi}, i = 1,2,…..,n. With this kind of enumeration method, a 20-order graph has 320 = 3486784401 possible colorings. Each DNA sequence with 300 bases represents a 3-coloring.The ssDNA in the single strand coloring library hybridizes with its W-C complement to form dsDNA. We take these dsDNA as the double strand coloring library. Actually, the sequences fixed on the board A are double strand coloring library. The DNA sequences will be synthesized after encoding.The DNA molecule synthesis is to form a DNA sequence by linking nucleotides with 3' -5' phosphodiester bond in order. Phosphoramidite synthesis method is commonly used to synthesize DNA sequence. The first base attached to the solid support by an ester linkage at the 3'-hydroxyl end is at first inactive by a DMT group at 5'-hydroxyl end because all the active sites are blocked or protected. To add the next base, the DMT group must be removed. A nucleotide will be added after a cycle, which includes four steps. Step 1 is protecting, step 2 is base condensation, step 3 is capping and step 4 is oxidation. Steps 1¯4 are repeated until all desired bases are added to the oligonucleotide. After all the bases are added, the oligonucleotide must be cleaved from the solid support and deprotected prior to use. This is done by incubating the chain in concentrated ammonia at a high temperature for an extended amount of time. At last desalting is done to purify the solution.

1.3. The Design of the Probe Library and the Principle of Separation

The probe library is composed of different ssDNA which are complementary with the DNA strands representing the unsatisfied solutions. The length of the probe is based on different graphs. The basic choosing principle of the probe is as follows: Firstly, the length of a probe should be as short as possible; secondly, the probe should be W-C complements with one or more sequences in the set {ri, bi, yi; i = 1,2,..,20}. Thus, the length of a probe must be multiples of 15. The principle of synthesis is the same as the DNA molecular synthesis as mentioned before. The probe should be labeled. Here we can use two different methods to label the probe: If we separate the DNA by using magnetic beads, the 5' end of the probe should be labeled by biotin. The principle of the separation technique is depicted in Fig. below. If we use the Acrydite TM separating technique, the probe should be fixed on a polyacrylamide gel.



Experiment

We separate the unsatisfied solutions with the techniques based on magnetic beads as follows:

1. The storage library, probe library and primers are synthesized;
2. The ssDNA representing the unsatisfied solutions are hybridized with probes;
3. The unsatisfied solutions are separated using the magnetic beads;
4. PCR is done to obtain the solution;
5. The solutions that we want are gotten with electrophoresis.

Method

(i) Encoding. xi, the color of every vertex in the graph in Fig. below, can be ri,bi, yi, i =1,2,3,4,5, and each coloring is represented by oligonucleotide with 15 bases. xi denote the complementary strands of xi. Each coloring of the graph is one of the possible solutions, and each of them, (i.e., xi, i =1,2,3,4,5), can be represented by ssDNA with 75 bases. The codes

r1 = CATCACATTCTCAAT, b1 = CAAATCCCC ACTTAT, y1 = CACTAATTCCATCCA

r2 = CTTCTTAAACACTTC, b2 = CCCAATACT TTATCC, y2 = CCCTCAAACTATTAT

r3 = CCTAATCTTTCTACT, b3 = ACACCAATCA TAACA, y3 = ACTACCCTATCTAAC

r4 = ACCAAAATACCACTA, b4 = ATCAATAACCC ATCT, y4 = TTACATCCTTAC TCA

r5 = AATCCACAACTCACT, b5 = CACACATATATC ATC, y5 = TAACCACCTATA CCT

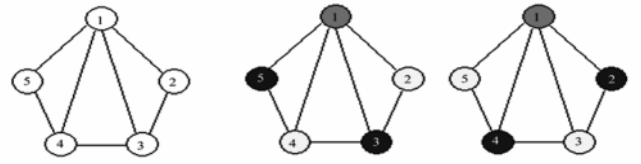


Fig. 6: Five Order Graph and its Two Solutions

According to these DNA sequences, 35 = 243 DNA strands would be contained in the storage library and each of them had 75 bases. These sequences were synthesized on the ABI 3900 synthesis apparatus.The probes should contain the information of the graph. And 24 probes were designed according to Fig. above. Some of them had 15 bases such as r1, and the others had 30 bases such as r1r2. All of the probes were labeled at 52 end with biotin.

(ii) Hybridization. Step 1. 0.5 μL of probes (r1,b1, y1; 0.2 p mol/μL) was put into 3 tubes respectively,and 5 ìL of ssDNA (8.35×10-4 pmol/μL) was added into each tube. These tubes were kept in hot wa ter (65 C) for10 minÿand then the tubes were removed and put into water (27 C) for 1 h. Then 5 μL Streptavidin MagneShere Promega Particles were added into each tube and the specific method was consistent with the protocol of Particles. After reaction the supernatants were discarded. Step 2. The Particles were washed with 5 μL 0.5 × SSC and the supernatants were discarded. 50 μL 0.5 × SSC was added into each tube, and each of them was kept in water (65C) for 10 min. Having been separated rapidly, the supernatants were removed to three new tubes respectively.Step 3. 0.15 μL of three probes (ri, bi and yi, i = 3,4,5; 40 pmol/μL) was added separately into three new tubes. These tubes were kept in water (65 C) for 5 min, and then they were kept in 27 C water for 2 h. Separated by using one fold particles under the same condition, the supernatants were put into new tubes.

Step 4: Three probe sets (r1r2, bjbj+1, y j y j+1; b1b2, rjrj + 1, y j y j + 1; y1y2, rjrj + 1, bjbj + 1; 35p mol/μL)were added separately into supernatants absorbed by probes ri, bi and yi in step 3. These tubes were kept in water (65C) for 5 minÿand then they were kept in 50 □ water overnight. The reactants were separated four times by using three fold particles. After washing the particles, the supernatants were removed to new tubes. In this step, the supernatants were mixed for PCR reaction.(iii) Achievement of the solution. The supernatant was used as DNA template. The PCR condition was as follows: the primers were x1, xi, x€{r,b, y}, i = 2,3,4,5, predenaturing at 94 C for 1 min; 94 C, 30s, 34 C, 30 s, 62 C, 30 s, 30 cycles; prolonging at 65 C for 5 min.

Result

(i) Detection of storage library. The set x1, ri was used as primer for PCR. Here x € {r, b, y}, i = 2,3,4,5 Result shown in the Fig below

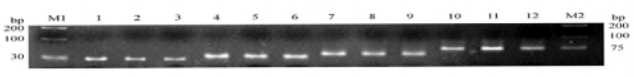


Fig. 1: Analysis of the Storage Library. PCR Products were Analyzed on 4% Agarose Gel. Lanes 1, 2 and 3 Correspond to Primer Pair x1, r2; Lanes 4, 5 and 6 Correspond to Primer Pair x1, r3; Lanes 7, 8 and 9 Correspond to Primer Pair x1, r4; Lanes 10, 11 and 12 Correspond to Primer Pair x1, r5. M1 and M2 are Molecular Weight Markers

(ii) Sensitivity Test: The concentration of the storage library DNA molecular was 8.35×10^{-4} pmol/μL and

this solution was done using a ladder like dilution. So, we got serial DNA solutions whose folds of dilution were 100, 200, 400, 600, 800, 1000, 5000, 10000, respectively.The PCR sensitive test condition was the same as the PCR's. By analyzing the result, a clue could be found that the concentration of templates in PCR system should be more than 16.7×10^{-6} pmol under the condition in this paper.

(iii) Hybridization Reaction: In sensitivity test, the probes and the ssDNA were mixed and hybridized. After reaction, the unsatisfied solutions were eliminated under the same condition. The hybridization products were used for PCR.

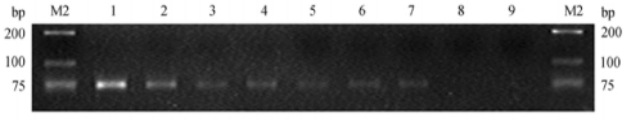


Fig. 2: The Result of the Sensitivity Test. The Products were Analyzed on 4% Agarose Gel. All the Lanes Correspond to Primer Pair r1, r5. Lane 1 Correspondsto the Liquor with Original Concentration; Lanes 2–9 Correspond to 100- to 10000-Fold Dilutions

(iv) PCR: The supernatant was used as DNA template. The PCR condition was: the primers were x1, xi, x€{r,b, y}, i = 2,3,4,5, predenaturing at 94 C for 1 min; 94 C, 30 s, 34 C, 30 s, 62 C, 30 s, 30 cycles; prolonging at 65 C for 5 min.

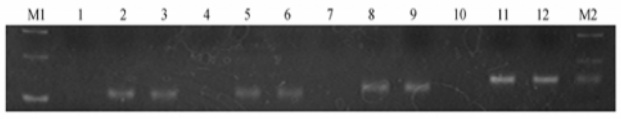


Fig. 3: The Result of PCR. The Products were Analyzed on 4% Agarose Gel. All the Lanes Correspond to Primer Pair r1, xi, here x (r, b, y), i = 2,3,4,5. There into i=2, x=r (Lane 1), i=2, x=b (Lane 2), i=2, x=y (Lane 3), i=3, x=r (Lane 4), i=3, x=b (Lane 5), i=3, x=y (Lane 6), i=4, x=r (Lane 7), i = 4x = b (Lane 8), i=4, x=y (Lane 9), i=5, x=r (Lane 10), i=5, x=b (Lane 11), and i=5, x=y (Lane 12)

(v) Agarose Electrophoresis: According to r1, b1, y1, the electrophoresis results were divided into three parts (Figs. 3–5). When the vertex 1 is red, namely r1, the truth value of the 5-order graph is shown as earlier. By analyzing Figs. 6 and 4, a group of truth value could be obtained, i.e. r1b2 y3b4 y5 and r1y2b3 y4b5. When the vertex 1 was blue, namely b1, the truth value of the 5-order graph was obtained as shown in Fig. 5.By analyzing Figs. 6 and 5, a group of truth value could be obtained, i.e. b1r2 y3r4 y5 and b1y2r3 y4r5. When the vertex 1 was yellow, namely y1, the truth value of the 5-order graph was obtained as shown in Fig. 5. By analyzing Figs. 6 and 5, a group of truth value could be obtained, i.e., y1r2b3r4b5 and y1b2r3b4r5.

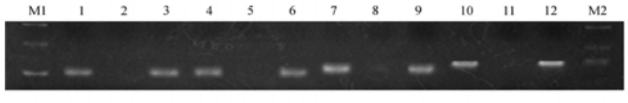


Fig. 4: The Result of PCR. The Products were Analyzed on 4% Agarose Gel. All the Lanes Correspond to Primer Pair b1, xi, x ∈ **{r, b, y}, i = 2,3,4,5, and the Rest are the Same as in Fig. 3**

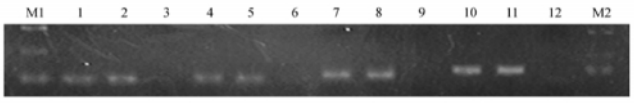


Fig. 5: The result of PCR. The products were Analyzed on 4% Agarose Gel. All the Lanes Correspond to Primer Pair y1, xi, x ∈**{ r,b, y }, i = 2,3,4,5, and the rest are as the same as in Fig. 3**

REFERENCES

[1] Adleman L., "Molecular Computation of Solution to Combinatorial Problems". Science, 1994.

[2] Lipton R J., "DNA Solution of Hard Computational Problems", Science, 1995.

[3] Roweis S, Winfree E, Burgoyne R, et al. "A Sticker based Architecture for DNA Computation", In: Baum E B, et al., eds. DNA Based Computers, Proc. 2nd Annual Meeting, Princeton, 1999.

[4] Braich R S, Chelyapov N, Johnson C, et al. "Solution of a 20-variable 3-SAT Problem on a DNA Computer", Science, 2002, 296(19).

[5] Sakamoto K, Gouzu H, Komiya K, et al., "Molecular Computation by DNA Hairpin Formation".

[6] Ouyang Q, Kaplan P D, Liu S M, et al. "DNA Solution of the Maximal Clique Problem".

[7] Liu Q H, Wang L M, Frutos A G, et al. "DNA Computing on Surfaces", Nature, 2000.

[8] Benenson Y, Gil B, Ben-Dor U, et al. "An Autonomous Molecular Computer for Logical Control of Gene Expression", Nature, 2004.

[9] Benenson Y, Paz-Elizur T, Adar R, et al. "Programmable and Autonomous Omputingmachine Made of Biomolecules", Nature, 2001, 414(22): 430-434.