Heuristic Solution to a 10-City Asymmetric Traveling Salesman Problem Using Probabilistic DNA Computing

David Spetzler, Fusheng Xiong, and Wayne D. Frasch

Molecular and Cellular Biology Graduate Program, and
Faculty of Biomedicine and Biotechnology, School of Life Sciences,
Arizona State University, PO Box 874501 Tempe, AZ 85287-4501, USA
frasch@asu.edu

Abstract. DNA hybridization was used to make a probabilistic computation to identify the optimal path for a fully connected asymmetric 10 city traveling salesman problem. Answer set formation was achieved using a unique DNA 20mer for each edge capable of hybridizing to half of each neighboring vertex. This allowed the vertex 20mers to be linked in all possible combinations to form paths through the network. Hybridization occurred in the presence of an excess of vertex 20mers, while edge 20mers were added in limiting amounts inversely proportional to the weight of each edge, resulting in the paths with the least cumulative weight being the most abundant. Correct answers, 230bp in length, contained a single copy of each vertex and were purified by PAGE and by successive magnetic bead affinity separations with probes for each vertex. Answer detection was accomplished using LCR of probes complementary to each vertex in a manner that identified the sequential order of vertices in each path by identifying vertex pairs. Optimal answer identification was accomplished using a conventional computer by normalizing the abundance of vertex pairings, and was found to be the same as that calculated by *in silico*.

Keywords: DNA computing, Traveling Salesman Problem, Ligation, Hybridization, Denaturing PAGE, Magnetic affinity.

1 Introduction

The use of DNA for making computations was first demonstrated by the successful computation of the solution to a 7 node Hamiltonian path problem (HPP) [1]. Methods to solve numerical optimization problems have been developed to expand the types of problems able to be solved using DNA computing [2-11]. Yamamoto et al. [12] accounted for the weight of each edge using a DNA concentration-dependent regime to design a computational method for a 6 node shortest path problem. Temperature has also been used to solve other numerical optimization problems with limited success. Although a subset of optimal solutions was purified, it was not possible to determine the optimal solution using this method [13].

One of the major limiting factors of DNA computing is that the number of molecules required to form every possible solution to an NP-complete problem is too large to generate [14]. Current methods consume a significant amount of the DNA

M.H. Garzon and H. Yan (Eds.): DNA 13, LNCS 4848, pp. 152–160, 2008.

forming incorrect solutions which must be thrown away, thus requiring more DNA to generate a complete solution set. Recent evidence also indicates that formation of secondary structures can occlude the correct solutions to make answer determination extremely difficult. Both of these factors have proved to be barriers preventing larger problems from being solved.

We now report an approach to DNA computing that generates a subset of the solution space composed of the solutions with the highest optimality rating, and does not require that the answer set includes every possible solution. Using this method, we have successfully solved a random instance of the fully connected 10 city asymmetric traveling salesman problem (Table 1) which has 3.3 million possible solutions. The NP completeness result for the Traveling Salesman problem implies that not all problem instances are hard . There are many different approaches to solve this type of problem, including exact and heuristic solutions. Conventional computers have completed hard instances of the problem, with thousands of cities, though at a cost of years of computing time. The TSP remains the standard optimization problem used to test new approaches as the struggle to solve NP problems continues.

Table 1. The distance matrix for the problem solved

	A[a]	B	C	D	E	F	G	H	I	J
A	***	55.2	34.05	31.75	53.85	39.95	36	39.9	36.55	52.6
B	63.95	***	54.25	54.95	72.6	45.05	71.65	50.55	52.75	52.15
C	51.35	47.6	***	41.45	39.8	57.8	55.2	32.75	34.85	37.05
D	46.65	46.25	54.6	***	49.4	45.55	55.9	52	57.35	54.6
E	49.9	39.1	42.65	52.4	***	25.9	39.85	38.85	37.95	33.1
F	59.7	49.15	47.8	56.9	58.05	***	48.05	46.6	48.4	47.7
G	51.25	36.7	43.95	43	42.45	40.25	***	64.2	47.8	46.95
H	58.1	35.85	53.7	45.05	47.3	43	84.25	***	42.8	41.9
I	52.9	38.2	40.35	33.45	36.5	65.2	35	29.7	***	30.95
J	60.05	39.1	40.65	55.75	41	41.1	45.1	58.65	43.95	***

[a] Each letter is a different vertex in the graph.
*** indicate paths that are not contained in the subset generated.

2 Methods

Computer design included three sequential steps to solve a problem that will find the most efficient path to visit all vertices once and only once, and then return to the starting vertex. In step 1, answer set formation was achieved using a combination of unique DNA 20mers for each vertex and edge. Hybridization of the edges occurs between the first half of one vertex and the second half of another. This allowed the unique vertex 20mers to be linked sequentially in all possible combinations to form paths through the network upon addition of ligase. The start and end vertex sequences contained an additional GC end cap which raised the melting point which eliminated nonspecific annealing during PCR. The computation was performed by adding all

vertex 20mers in excess, while edge 20mers were added in limiting amounts that varied relative to the efficiency factor determined for that edge.

A set of 10 unique 20-mer sequences were designed using the software developed by Tanaka [15] to represent each vertex which were synthesized by Invitrogen. An additional 90 oligomers were synthesized containing all possible combinations of the complementary sequences to join any two city sequences together. The oligo sequences were designed to minimize cross hybridization, self-assembly and secondary structure formation and have similar thermal properties (melting temperature, between 61.3-61.8^{o}C and GC content (25-30%). The yield of the sequences synthesized was used to define the distance matrix for the problem solved.

The initial answer DNA pool was generated by combining saturating amounts of vertex sequences with limiting amounts of edge sequences. The concentration of each edge was inversely proportional to the cost of that edge. The result was a population of heterogeneous sequences that formed upon hybridization and ligation of the sequences. The approximate concentration ratio for the hybridization and ligation reaction was set at 10:1 (vertex:edge). This ensured that vertex oligo concentrations were saturated while edge concentrations were limiting and varied in concentration so that a potential linking between any pair of vertices was dependent upon concentration of the corresponding edge sequence. Table 1 shows concentration differences for all linkers that were used to solve the 10-city problem.

The initial answer pool was generated using a two step process. First, initial hybridization/ligation was conducted in the absence of the ending vertex sequence and all linkers to the ending vertex. This greatly reduced formation of shorter answer sequences, thus improving the hybridization/ligation efficiency. Secondly, the hybridization/ligation was allowed to continue with the addition of fresh ligase, the ending vertex sequence and corresponding linkers. Hybridization-ligation products were purified through sequential PCR amplifications using the 5'-starting and the 3'-ending vertex primers with the target DNA templates that were extracted from the profiled PAGE gel containing previous PCR products.

Unless specified otherwise, PCR was performed in a 0.5 ml microcentrifuge tube with a total volume of 50 µl reaction mixture containing 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$ and 0.001% gelatin, 200 µM each dNTP (ATP, GTP, CTP and TTP), 0.4 µM each of the starting and the ending primer, 2.5 U Taq DNA polymerase (New England BioLabs, MA, USA) and 50-100 ng DNA template. PCR reactions were carried out using a PJ2000 DNA thermal cycler that was programmed for a "Hot start" at 94^{o}C for 2.5 min followed by 35 cycles. Each cycle consisted of a denaturing step at 94^{o}C for 0.5 min, an annealing step at 68-70^{o}C for 40 sec, and an extension step at 72^{o}C for 30 sec. These cycles were concluded by a final extension for 3 min at 72^{o}C.

The PCR products were profiled on a 6% denaturing polyacrylamide gel (the ratio of acrylamide to bisacrylamide was 29:1) in 1 x TBE buffer (90 mM Tris-borate, pH 8.3, 83 mM boric Acid, and 2 mM EDTA) at room temperature under 10 volts/cm. After electrophoresis, the gel was stained with ethidium bromide (1 mg/ml) for 10 min. The image profile was visualized and photographed under a UV transilluminator (UVP BioDoc-It^{TM} system, UVP).

In Step 2, Answer sorting was achieved in two stages. First, answer sequences were separated by PAGE, the 230mer band was excised from the gel, and the DNA

was amplified by PCR using the procedures described by Xiong et.al [16]. This PAGE separation, PCR amplification step was repeated five times to insure that only 230mer DNA was present. Second, the purified 230mer answer DNA was probed for the presence of each vertex sequence sequentially using magnetic affinity separation. Specifically, magnetic affinity separations were carried out by incubating 0.75 µl of vertex probe (400 µM), 149 µl 5x Binding/Washing (B/W) buffer with 150 µl of the M-280 beads (~1.5 mg, Dynal Biotech ASA, Osho, Norway), which were pre-washed 3 times with 500 µl B/W buffer for each time. After 45min at RT, the beads were separated using a magnetic separator. After washing 3 times using B/W buffer (500 µl each time), 90 µl of the ssDNA solutions was added along with 30 µl 20x SSC. After being gently vortexed and incubated at room temperature for 50-60 min, the DNA solutions hybridized to the immobilized probes were retained on the magnetic beads through biotin-streptavidin interaction. Those strands missing single or multiple vertex sequences were washed away (2 x SSC, 2 times with 500 µl each time, and then 0.5 x SSC, once, 500 µl). The captured duplexed answer sequences were dissociated from the biotinylated vertex probe in 80 µl of 0.1 N NaOH for 6-10 min. After separation using a magnetic separator, the collected supernatant containing the screened ssDNA solutions was neutralized by adding 8.2 µl of 1 N HCl, 10 µl of H_2O, 2 µl of 0.25 M Tris-Cl, pH 7.5.

In Step 3, ligation chain reaction (LCR) [17] combined with PAGE gel electrophoresis was implemented to characterize the answer pool. The answers were determined by performing a series of LCR reactions to determine the number of times one vertex preceded another. Complementary sequences to two vertex sequences were added at a time to determine the abundance of ordered pairs in the answer sequences. Since only one of the primers was phosphorylated, the probe that was phosphorylated on the 5' end dictated the order that the primers could link. Ligation chain reactions

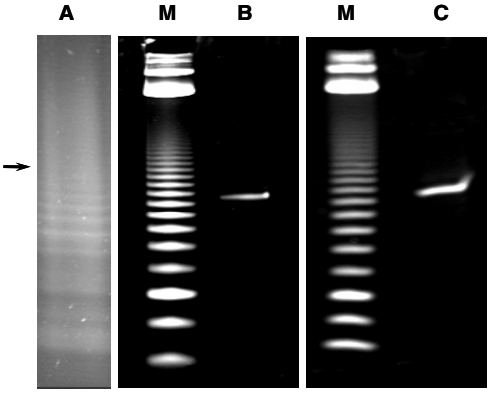

Fig. 1. A: The result of the two step hybridization and ligation reaction. The arrow shows where the 230 bp band was excised. B: Result of repeated rounds of PCR amplification on the 230bp band excised from the page gel for size separation. C: Result of PCR amplification of isolated 230-mer band after magnetic affinity purification containing the solutions to the problem. M: is the molecular marker, the brightest band is 100 bp.

were run for each possible vertex-vertex pairing, 90 reactions, and the products were profiled on PAGE gel. The relative abundance of each product measured as total density using UVP GDS-8000 BioImaging system.

Quantitative determination of the yield of LCR product was accomplished by the following procedure: (1) measure total density of the upper DNA band (the LCR product) and the lower DNA band (the PCR probes); (2) measure total density of the 100-mer band (the brightest one) from the DNA ladder; (3) normalize the total density of LCR product and probes against the 100-mer DNA ladder; (4) calculate the ratio of the normalized total density of PCR product over the normalized total density of PCR probe was calculated. This value represents a global measure of the abundance of each particular city pairing, and is used to determine the answer to the DNA calculation.

3 Results

Using a two-step hybridization/ligation protocol, an initial answer pool was generated with an amount of DNA distributed in the 230-mer region visible by PAGE, the required length for correct answers to the 10-city problem (Figure 1, Lane A). The hybridization/ligation products were actually distributed over a wide range of sizes and DNA sequences as large as ~500-mers were observed. This implies larger hybridization/ligation products were generated and that the techniques and protocols developed here for the 10-city problem are sufficient for larger problems.

Sequences formed that did not correspond to correct answers to the traveling salesman problem being solved were removed in two stages. First, correct answers must contain a single copy of each city sequence and thus should be 230bp in length. The 230mer answer band was excised from a gel to eliminate incorrect answers with too few or too many cities and was amplified by PCR. Four successive excision-amplifications yielded pure correct answers only when the PAGE step was done at 65ºC (Figure 1, Lane B). This band corresponding to the correct length for solutions was then collected for subsequent purification by magnetic affinity separation. Second, avidin-coated magnetic beads were bound to biotinylated oligo probes complementary

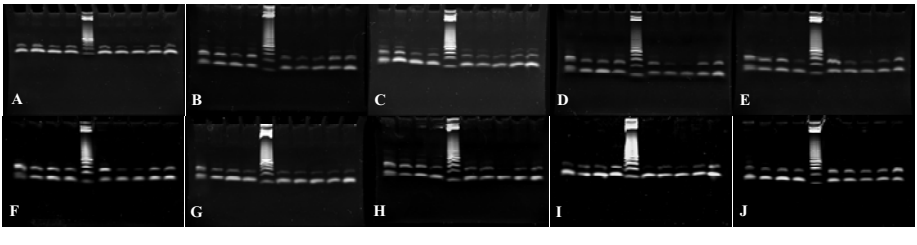

Fig. 2. PAGE profiles of ligation chain reaction product for all potential vertex pairings. For each LCR reaction, 230-mer DNA solutions plus two pairs of probes were included. The lower band is composed of probes that were not ligated. The upper bands are probes that were ligated and thus indicate the presence of that ordered pair in the answer pool.

the sequence for each vertex. These beads were used to probe the answer sequences sequentially to ensure every vertex was present. After all nine magnetic affinity separations, the remaining answer sequences that contained all vertices gave rise to a sharp 230mer band when separated by PAGE (Figure 1, Lane C).

The answer sequences were mixed with complementary sequences to two vertices that became covalently linked by ligase when the vertices were adjacent in a specific order. This was repeated for all combinations of vertex pairs, n^2-n required tests (Figure 2), and the relative abundance of all potential vertex pairings was determined. Thus, the concentration of each edge was determined from which the optimal pathway was deduced (Table 2). This was accomplished by normalizing the abundance of vertex pairings against the constant amount of city probes in each PAGE lane. Special attention was taken to perform the LCR with equal amounts of probe. The concentration of each probe was measured in triplicate using a NanoDrop ND-1000 spectrophotometer and a saturated concentration for each probe was used in the reaction. This ensured the yield of LCR product for a given link between two vertices was limited by the abundance of the corresponding answers.

Table 2. Matrix formed through the LCR gel read out

	A[a]	B	C	D	E	F	G	H	I	J
A	* * *	0.30	0.16	0.26	0.23	0.26	0.47	0.24	0.00	0.08
B	0.11	* * *	0.27	0.09	0.14	0.15	0.17	0.19	0.17	0.06
C	0.16	0.27	* * *	0.10	0.16	0.13	0.13	0.18	0.13	0.09
D	0.09	0.05	0.05	* * *	0.01	0.07	0.06	0.06	0.10	0.06
E	0.17	0.17	0.07	0.15	* * *	0.19	0.05	0.09	0.02	0.15
F	0.16	0.03	0.11	0.10	0.29	* * *	0.04	0.10	0.11	0.14
G	0.06	0.01	0.06	0.00	0.00	0.01	* * *	0.00	0.05	0.08
H	0.03	0.00	0.03	0.00	0.00	0.01	0.03	* * *	0.14	0.10
I	0.11	0.08	0.09	0.10	0.04	0.07	0.03	0.04	* * *	0.23
J	0.12	0.09	0.17	0.18	0.14	0.10	0.04	0.11	0.2	* * *

[a] Each letter is a different vertex in the graph.
*** indicate paths that are not contained in the subset generated.

The optimal answer obtained by the DNA computer was found to be the same as that calculated by a conventional computer. The DNA computer generated about 246,960 answer sets in total, which was 6.8% of the 3.3 million possible correct answers.

A conventional computer ranked the 1000 most optimal answers from best to worst and compared the answers generated by the DNA computer (Figure 3). The DNA computer successfully generated the 24 most optimal answers. The first answer not included was the 25th most optimal. The number of answers excluded by the DNA computer increased proportionately to the decrease in optimality. Of those sequences that did represent a correct answer, the majority of them corresponded to answers that had a high optimality rating. This occurred because the reaction mixture was

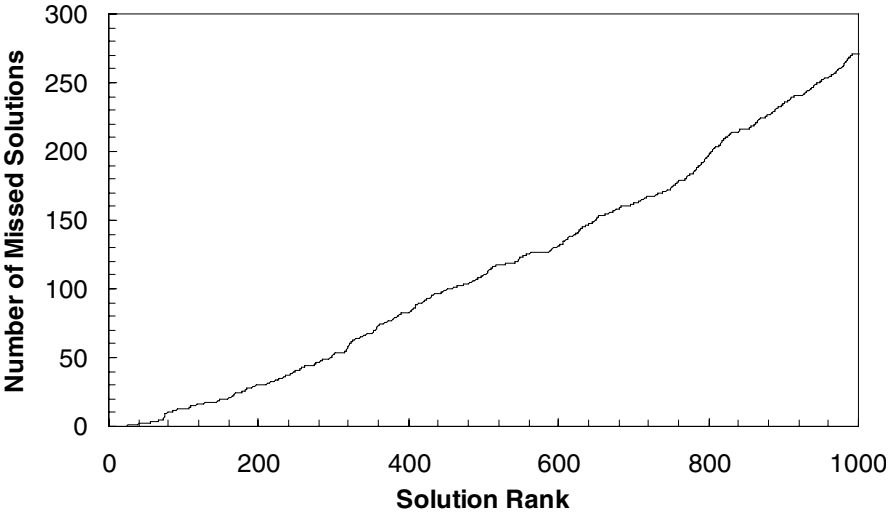

Fig. 3. The number of solutions that were missed as ranked by their optimality

composed of a variable amount of pathway sequences, such that the lower the cost of travel between two vertices, the more abundant the pathway. Thus, those sequences that represent good answers form a larger portion of the population.

4 Conclusion

These results demonstrate that our DNA computer presented here has successfully computed the 10 city problem. The first pathway through the network that was not contained by the subset of solutions generated by the DNA computer was the 25th most optimal solution. Thus, our method reduced the number of possible solutions while still retaining the most efficient pathway. Since this is a stochastic calculation, there is a chance that the optimal solution will not be created. However, since we do not readout particular solutions, but instead study the profile of the ordered pairs of vertices contained in the answer pool, it is likely that the paths that are involved in the optimal pathway will still be present. Thus, our readout method will allow the optimal path to be determined even though it might not be present in the DNA solution.

The distance matrix we chose to solve was not constrained, and solving it suffices to show that the technique can be used to solve any problem of lesser complexity. In the problem solved here, the number of possible solutions paths through the network of cities is limited by the rows and columns with the fewest possible transitions. Each row or column with fewer than 9 transitions limits the number of degrees of freedom that any path may travel. For example, vertex H may only be traveled to from vertices I or J, thus it has a degree of freedom of 6. To find the maximum number of potential solutions, we begin with the path that has the smallest degree of freedom and continue to the next smallest from there. In this case the number of possible solutions

can be calculated by taking the minimum of the lowest degree of freedom and the number of remaining vertices to move to. In this way we can determine that the matrix generated by the DNA computer has at most (7*7*7!) = 246,960s or 6.8% of the possible solutions of the original problem. Thus the DNA computer has served to reduce a problem with 3.3 million possible solutions to one with 246,960. This is a statistical sampling of the total population that is weighted towards better solutions. Although there can be no guarantee that the optimal solution will be found, but it is highly likely that a near optimal solution will be. Since the number of solutions is small enough to be searched, we used a standard laptop computer to perform a brute search of all possible solutions to find the optimal solution, AFEJBCDIHGA. We compared this solution to the optimal solution of the original problem defined by the initial concentrations of all the pathways. The power of such a method to solve large optimization problems lies in the combination of biological and silicon computing and represents the most practical implementation of biological computing to date. However there are still significant obstacles to useful DNA computing. With all current methods as the number of vertices increase, the reaction volume increases as well. This problem is limiting due to the large percentage of sequences that may not form correct solutions. Until a method can be developed where each molecule forms a solution, acquiring a large enough sample of the solution population may be prohibitive for problems with more variables. Work is underway to establish this upper limit.

Acknowledgements

We would like to thank Justin York for helpful discussions. This work was supported by funding from DARPA/DSO and AFOSR to W.D.F.

References

1. Adleman, L.M.: Molecular computation of solutions to combinatorial problems. Science 266(5187), 1021–1024 (1994)
2. Reif, J.H., LaBean, T.H., Sahu, S., Yan, H., Yin, P.: Design, simulation, and experimental demonstration of self-assembled DNA nanostructures and motors. Unconventional Programming Paradigms 3566, 173–187 (2005)
3. Seeman, N.C., Wang, H., Yang, X.P., Liu, F.R., Mao, C.D., Sun, W.Q., Wenzler, L., Shen, Z.Y., Sha, R.J., Yan, H., Wong, M.H., Sa-Ardyen, P., Liu, B., Qiu, H.X., Li, X.J., Qi, J., Du, S.M., Zhang, Y.W., Mueller, J.E., Fu, T.J., Wang, Y.L., Chen, J.H.: New motifs in DNA nanotechnology. Nanotechnology 9(3), 257–273 (1998)
4. Schmidt, K.A., Henkel, C.V., Rozenberg, G., Spaink, H.P.: DNA computing using single-molecule hybridization detection. Nucleic Acids Research 32(17), 4962–4968 (2004)
5. Shin, S.Y., Lee, I.H., Kim, D., Zhang, B.T.: Multiobjective evolutionary optimization of DNA sequences for reliable DNA computing. Ieee Transactions on Evolutionary Computation 9(2), 143–158 (2005)
6. Qu, H.Q., Zhu, H., Peng, C.: New algorithms for some NP-optimization problems by DNA computing. Progress in Natural Science 12(6), 459–462 (2002)

7. Shao, X.G., Jiang, H.Y., Cai, W.S.: Advances in biomolecular computing. Progress in Chemistry 14(1), 37–46 (2002)
8. Tostesen, E., Liu, F., Jenssen, T.K., Hovig, E.: Speed-up of DNA melting algorithm with complete nearest neighbor properties. Biopolymers 70(3), 364–376 (2003)
9. Kim, D., Shin, S.Y., Lee, I.H., Zhang, B.T.: NACST/Seq: A sequence design system with multiobjective optimization. DNA Computing 2568, 242–251 (2003)
10. Lipton, R.J.: DNA Solution of Hard Computational Problems. Science 268(5210), 542–545 (1995)
11. Faulhammer, D., Cukras, A.R., Lipton, R.J., Landweber, L.F.: Molecular computation: RNA solutions to chess problems. Proceedings of the National Academy of Sciences of the United States of America 97(4), 1385–1389 (2000)
12. Yamamoto, M., Kameda, A., Matsuura, N., Shiba, T., Kawazoe, Y., Ohuchi, A.: A separation method for DNA computing based on concentration control. New Generation Computing 20(3), 251–261 (2002)
13. Lee, J.Y., Shin, S.Y., Park, T.H., Zhang, B.T.: Solving traveling salesman problems with DNA molecules encoding numerical values. Biosystems 78(1-3), 39–47 (2004)
14. Hartmanis, J.: Response to the Essays on Computational-Complexity and the Nature of Computer-Science. Acm Computing Surveys 27(1), 59–61 (1995)
15. Tanaka, F., Kameda, A., Yamamoto, M., Ohuchi, A.: Design of nucleic acid sequences for DNA computing based on a thermodynamic approach. Nucleic Acids Res. 33(3), 903–911 (2005)
16. Xiong, F., Spetzler, D., Frasch, W.: Elimination of Secondary Structure for DNA Computing. In: Proceedings of DNA13 (2007)
17. Yamanishi, K., Yasuno, H.: Ligase chain reaction (LCR). Hum. Cell. 6(2), 143–147 (1993)