We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX INDEXED — CLARIVATE ANALYTICS

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter

Accelerating DNA Computing via PLP-qPCR Answer Read out to Solve Traveling Salesman Problems

Fusheng Xiong, Michael Kuby and Wayne D. Frasch

Abstract

An asymmetric, fully-connected 8-city traveling salesman problem (TSP) was solved by DNA computing using the ordered node pair abundance (ONPA) approach through the use of pair ligation probe quantitative real time polymerase chain reaction (PLP-qPCR). The validity of using ONPA to derive the optimal answer was confirmed by *in silico* computing using a reverse-engineering method to reconstruct the complete tours in the feasible answer set from the measured ONPA. The high specificity of the sequence-tagged hybridization, and ligation that results from the use of PLPs significantly increased the accuracy of answer determination in DNA computing. When combined with the high throughput efficiency of qPCR, the time required to identify the optimal answer to the TSP was reduced from days to 25 min.

Keywords: DNA computing, traveling salesman problem, ordered node pair abundance, pair ligation probe-qPCR, PLP-qPCR

1. Introduction

The traveling salesman problem (TSP) computes the shortest route on the arcs of a network that visits a given set of nodes (cities) before returning to the starting point [1–6]. In many cases, *in silico* computers are incapable of quickly determining an exact solution of these nondeterministic polynomial (NP) problems because a linear increase in the number of variables leads to a factorial increase in the number of potential solutions. Although advanced heuristic methods have increased the ability of *in silico* computers to provide approximate answers to NP problems [7], they lack the massive parallelism and data storage required to find exact solutions. DNA computing, which uses the hybridization of DNA molecules as a means to make computations [8–15], is particularly well-suited to solve computationally intense NP problems such as the TSP because multiple sequences of DNA in a solution can hybridize simultaneously, thereby performing massively parallel computing.

Several technical limitations have prevented DNA computing from reaching its full potential. Although the computation can occur within seconds, current

applications of DNA computing are largely limited by time-consuming and labor-intensive answer sorting and determination processes. For example, a gradient PCR procedure that involved a series of PCR reactions with a variety of primer pair combinations was used to determine the solution to a 7-city directed Hamiltonian circuit problem [8]. Other answer determination methods include DNA sequencing [16], and denaturation temperature gradient-polymerase chain reaction (DTG-PCR) associated with denaturing gradient gel electrophoresis (DGGE) and/or temperature gradient gel electrophoresis (TGGE) [10, 17, 18]. Neither DNA sequencing nor DGDC/TGDC can be used to identify the optimal answer to a TSP because the feasible answers represent both optimal and suboptimal answers that differ only in the order in which the components were ligated when the answers were formed, and are flanked by the sequences that encode the start and end nodes. The dependence of the discriminatory powers of DTG-PCR, DGGE, and TGGE on the diversity of oligonucleotide components, especially the G and C content [19, 20] also limits their applicability for answer determination of NP-complete problems.

Xiong et al. [12] developed the ordered node pair abundance (ONPA) approach to identify the optimal answer of an asymmetric, fully-connected 15-city TSP, the largest problem solved to date, using DNA computing. In that study, 20-mer DNA sequences that specifically encoded the nodes and arcs in the network were added to the reaction mixture in the presence of ligase. The sequence of each arc was capable of linking two node sequences in a specific order via hybridization of the last 10 bases of the prior node to the first 10 bases of the subsequent node. The efficiency of each arc in the network was encoded as the concentration of DNA for that arc using saturating concentrations of DNA for each node. Ligation of these sequences then formed covalently linked ordered node pairs (ONPs) that assembled into answer sequences representing tours through the network, such that the optimal answer would be formed in greatest abundance.

The optimal answer to the 15-city TSP was identified by ONPA after infeasible answer sequences were removed by electrophoresis and magnetic bead separation [11, 12]. For each ONP examined by ONPA, probes were designed to hybridize to the 3′-end of the prior node and the 5′-end of the subsequent node in the ordered pair as well as a complementary pair for use in ligation chain reaction (LCR). Each pair of probes was then subjected to the same number of LCR cycles in the presence of an aliquot of answer sequences, and the relative abundance of LCR product was quantified by PAGE band intensity. This was repeated for every possible combination of ONPs to identify the ones present in greatest abundance. Although this provided a clear determination of the optimal answer, the accuracy and precision of the computation were limited by the error inherent in determining DNA abundance via bands in a PAGE gel and by the fact that LCR is an end-point assay. In addition, the procedure required days to complete.

Quantitative real-time PCR (qPCR), which uses an increase in fluorescent signal to monitor increases of PCR amplicons, has proven to be a fast, precise and reproducible method to rapidly quantify the relative abundance of many nucleic acid sequences. This results in part because qPCR is not an end-point assay. Instead, the cycle at which a positive signal is first consistently detectable, termed the cycle threshold (C_t), is proportional to the initial content of a given template in the sample. Higher initial target contents give rise to earlier detectable increases in signal, resulting in lower C_t values. Simultaneous amplifications of multiple DNA targets in a single reaction tube can be achieved by multiplex qPCR, which maximizes throughput and decreases the time required to collect information. Ibrahim et al. [21] reported a qPCR approach to readout answers to a Hamiltonian path problem. However, the protocol required multiple combinations of qPCR

amplifications and was ultimately dependent upon the use of *in silico* information processing to determine the answer.

Answer determination via a direct qPCR amplification of the entire length of TSP answer sequences is impossible. This is due to the intra-molecular heterogeneity of answers sharing the same starting and the ending nodes, which are the same limitations that eliminate the use of DTG-PCR. The application of qPCR for the quantification of the short 20-mer sequences using ONPA for answer determination of a TSP has also not been possible because standard PCR methodologies require two DNA primers that flank the region of the DNA sequence to be amplified. However, we developed a pair ligation probe (PLP)-dependent qPCR system (PLP-qPCR) that is capable of rapidly quantifying the abundance of short DNA sequences under the multiplex conditions [22]. Upon hybridization to its 20-mer target DNA, the PLP becomes circularized by ligation. This allows the abundance of adjacent short sequences in a DNA strand to be transformed into the copy number of circularized PLP that can be amplified by qPCR.

We have now solved an asymmetric, fully-connected 8-city TSP by the ONPA approach using PLP-qPCR without the need for *in silico* information processing. The validity of using ONPA to derive the optimal answer was confirmed using a reverse-engineering method to reconstruct the complete tours in the feasible answer set from the measured ONPA. The high specificity of the sequence-tagged hybridization and ligation that results from the use of PLPs significantly increased the accuracy of answer determination in DNA computing. When combined with the high throughput efficiency of qPCR, the time required to identify the optimal answer to the TSP was reduced from days to 25 min.

2. Materials and methods

2.1 Oligonucleotide design and construction

Oligonucleotide sequences were designed using Primer Express Version 2 for Windows (Applied Biosystems) and were synthesized by Invitrogen Inc. Nodes labeled B through H were represented by synthetic 20-mer sequences of DNA except for the starting and ending nodes (A_{start} and A_{end}) that were comprised of 30-mers (**Table 1**). Node sequences were chosen to minimize cross hybridization, and nodes B through H had a GC content from 30% to 35% such that melting points varied from 60.6 to 62.0 °C, while the A_{start} and A_{end} sequences contained GC contents of 66% and 72%, respectively. None of the arc sequences were complementary to the first 15-mer sequence for the A_{start} and the last 15-mer sequence for the A_{end}. Thus, incorporation of A_{start} or A_{end} into an answer sequence prevented further extension of that end of the answer sequence. The start and end sequences also served as primer sequences for downstream amplification by PCR, which provided the capability to increase the amount of answer sequences. The longer A_{start} and A_{end} sequences with higher GC content were important for downstream PCR amplification with enhanced fidelity and improved efficiency for answer purification. Any two node sequences could be linked together by a 20-mer arc sequence composed of two 10-mers that were complementary to the last and first 10 nucleotides in the sequences of the former and latter nodes, respectively (**Figure 1**). Arcs were made to complement every combination of nodes with the exception of the respective 5′ and 3′ ends of A_{start} and A_{end}. Oligo sequences (**Table 2**) designed for the PLP-mediated qPCR assay followed the protocol of Xiong and Frasch [23]. All oligonucleotides were purified by PAGE under denaturing conditions [12].

Node	Length	Sequence[a]	GC (%)	T_M (°C)
A_S[e]	30	TCTGCGGGCGGACAGACATGGTTAGCGGCC[b,c]	66	70.0
B	20	TTTACGTCTACCATATCTAT[d]	30	61.2
C	20	ATAGCAACACTACATATGTC	35	61.6
D	20	TCGACTAATTCGTACTTATA	30	61.6
E	20	CCTGATACAAGTACTAAGTA	35	61.6
F	20	GCGTAAGGATATTTATACAA	30	61.6
G	20	GTTTGTTTAGTCCATCATTA	30	61.7
H	20	AGCATTATTTCTTCCAAATA	25	61.7
A_E[e]	30	CTACTGCCGCCGCCGGGTAGACGGCTCGGA	72	72.0

[a] All sequences read in the 5′ to 3′ direction.
[b] Blue sequences can hybridize to the 3′-ends of arc sequences.
[c] Green sequences are unable to hybridize to arc sequences but serve as primer sequences for PCR.
[d] Black sequences can hybridize to the 5′-ends of arc sequences.
[e] A_S is the sequence for node A_{start}; the A_E is the sequence for the node A_{end}.

Table 1.
Node sequences used in the calculation.
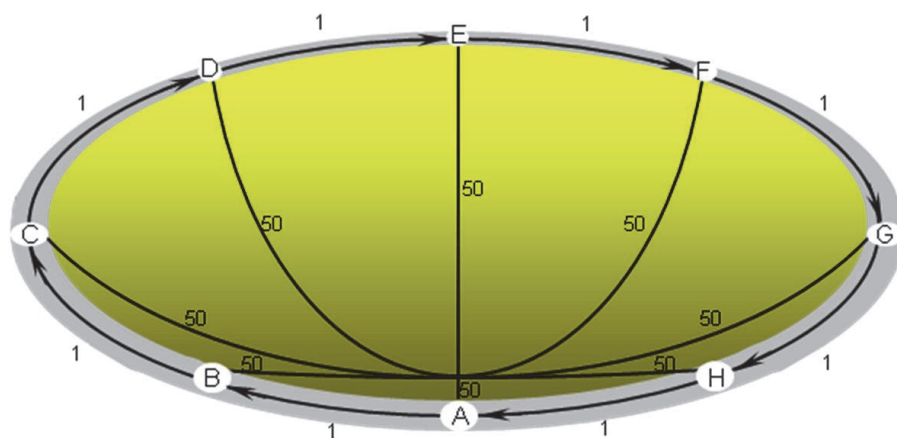
Figure 1.
Graphical representation of the 8-city TSP solved by DNA computing. The optimal tour through the network visits the nodes in alphabetical order. All other tours include arcs that are 100-fold less efficient, and cross at a common point.

2.2 Answer formation and purification

Hybridization was performed by heating the answer formation reaction medium (AFRM) to 92°C for 4 min, followed by a programmed annealing process at a cooling rate of 1°C per min to 8°C. The AFRM was composed of all relevant oligonucleotides in T4 Ligase Buffer (Fermentas). Ligation was performed by incubating at 8°C for ∼16 h after addition of 5 Weiss U of T4 ligase, 20 mM DTT, and 10 mM ATP to the AFRM.

Ligation products containing all answer sequences were separated by 6% denaturing PAGE containing 8 M urea and 15% formamide at 95 V for about 3 h. The gels were visualized with UV light after staining with ethidium bromide (1 mg/ml) for 10 min. Fragment sizes were determined by comparison with mobility of a 20-bp DNA ladder (Bayou BioLabs). The 190 bp band was excised from the gel that contained sequences consistent in length to those of feasible answers (i.e. containing one copy of each node sequence starting and ending with A_{start} and A_{end}). These sequences were amplified by PCR using PCR primers for A_{start} and

DNA sequence	
qPCR reporter-identifying sequences	
NED-TAqMan® spacer	CGCGGATGTCGGTCAGCCGAGTCTACCCAGCGCGCCACTATCGCCATCAGGCAGC
TaqMan-MGB® reporter	
NED	NED–TCGCCATCAGGCAGC–MGBNFQ
Pair ligation arms for node recognition	
5′-B′	ATAGATATGG
5′-D′	TATAAGTACG
5′-F′	TTGTATAAAT
3′-C″	GTGTTGCTAT
3′-E″	TTGTATCAGG
3′-G″	CTAAACAAAC
qPCR primers	
P_r	TAGACTCGGCTTGACCGACATCCGCG
P_f	TCTACCCAGCGCGCCAC

Table 2.
Oligonucleotide sequences for the 5′ and 3′ arms of the pair ligation probes, the forward (P_f) and reverse (P_r) primers, and the TaqMan reporters used for PLP-qPCR.

A_{end}, and subject to sequential magnetic affinity purification steps as reported by Spetzler et al. [11, 12] to purify feasible answers.

2.3 Preparation of target-specific pair ligation probes

Each PLP consisted of one 55-mer core and two 10-mer target-specific sequences that comprised the 5′ and 3′ arms located at ends of the core (**Table 2**). The core sequence contained the forward (P_f 19-mer) and reverse (P_r 21-mer) PCR primer-binding sequences, and a qPCR reporter-identifying sequence known as a TaqMan spacer for use with the TaqMan-MGB® (Applied Biosystems) NED reporter dye (λ_{max} = 580 nm). The PLPs specific for each ONP in the answer sequences were made from core and arm components by ligation as per Xiong and Frasch [23].

2.4 Circularization of the PLP and the qPCR assay

Aliquots of the purified answer sequences containing ∼2 pmol DNA were denatured and annealed with 20 pM of linear PLP. The hybridized PLPs were circularized by ligation at 10°C overnight. The ligation reaction mixture contained 2 μl of 10× ligation buffer (Fermentas), 50 mM DTT, and 5 Weiss U of T_4 DNA ligase, in a final volume of 20 μl. After ligation, 2 μl of ligation product was added to 18 μl of the exonuclease mixture that contained 10 mM Tris-HCl, pH 9.0, 5 mM $MgCl_2$, 0.1 mg per ml of BSA, 10 U Exonuclease I and 10 U Exonuclease III to remove any remaining linear PLPs. The samples were incubated at 37°C for 2 h followed by inactivation at 65°C for 20 min.

Quantitative real-time PCR (qPCR) assays were performed in a 96-well, closed plate using the AB 7500 Fast RT-PCR System (Applied Biosystems). In a typical qPCR assay, 20 μl of qPCR reaction mixture that contained 10 μl of 2× TaqMan® Fast universal PCR Master Mix including the ROX fluorescent reporter as a passive reference, 900 nM of each of the P_f and P_r primers, 0.25 μM of the NED

TaqMan-MGB® reporter, and ~2 ng of the circularized PLP were used to determine the abundance of each ordered node pair. Thermal cycling profiles for qPCR included heating at 95°C for 20s followed by 45 cycles with 95°C for 3s and 60°C for 30s.

During the PLP-qPCR assay, DNA amplification was monitored by quantitatively analyzing fluorescence emissions using an ABI prism sequence detector. The intensity of the NED reporter dye was measured against the ROX internal reference dye signal to normalize for non-PCR-related fluorescence fluctuations that occurred from well to well. The threshold cycle (C_t) represented the refraction cycle number at which a positive amplification reaction was measured and was set at 10 times the standard deviation of the mean baseline emission calculated from PCR cycle 5–15.

2.5 Linear programming (LP) model of tour frequency from ONPA data

The model is based on LP approaches to data fitting [24–27], and solves for the tour frequencies X_t that best fit the observed ONPA F_{ij}.

$$\text{Minimize} \sum_{i=1}^{8} \sum_{j=1}^{8} \left(d_{ij}^{+} + d_{ij}^{-} \right) \tag{1}$$

$$\text{Subject to:} \sum_{t=1}^{5040} h_{tij} X_t + d_{ij}^{+} - d_{ij}^{-} = F_{ij} \quad \text{for all } i,j \tag{2}$$

$$X_t, d_{ij}^{+}, d_{ij}^{-} \geq 0 \quad \text{for all } t,i,j \tag{3}$$

where d_{ij}^{+} = positive deviation from observed frequency of arc ij; d_{ij}^{-} = negative deviation from observed frequency of arc ij; X_t = estimated frequency of tour t; h_{tij} = 1, if tour t contains ONP ij; 0 otherwise; F_{ij} = observed frequency of arc ij.

The number of possible tours t that begin and end with node A, and visit each node only once is P_7^7 = 7! = 5,040. Since the number of DNA molecules of tour t that were made by the DNA computer were not explicitly measured, one X_t variable is generated for each tour t, representing how many DNA molecules of tour t were likely to have been made, given the ONPA. The goal was to solve for the best-fitting values of X_t. Constraints **2** are written once for each ONP ij. The first expression on the left-hand side of the constraint, $\sum_{t=1}^{5040} h_{tij} X_t$, sums all occurrences of arc ij within all tours t. Constraints **2** then add a positive deviation or error (d_{ij}^{+}) and subtract a negative deviation (d_{ij}^{-}) to the left-hand side and equate it to the observed ONPA F_{ij}. Thus, if there is no set of X_t values that can perfectly reproduce the observed ONPA, the deviation variables are forced to take up any slack or surplus. The objective function **1** minimizes the sum of the deviations so as to find the best-fitting tour frequencies X_t for the observed ONPA F_{ij}. The deviations are separated into positive and negative components because linear programming cannot handle absolute values explicitly. The deviation variables, like the tour frequency variables, are constrained by **3** to be non-negative. If the deviation variables were permitted to be either positive or negative, minimizing their sum would not provide a good fit because the objective would reward large negative deviations pushed toward $-\infty$.

3. Results

The computation to find the optimal solution to an asymmetric, fully-connected 8-city TSP was defined by the distance matrix in **Table 3**. The problem was

Prior	Subsequent node								
Node	A_S	B	C	D	E	F	G	H	A_E
A_S	*a (0)b	1c (100)b	100 (1)	100 (1)	100 (1)	100 (1)	100 (1)	100 (1)	100 (1)
B	* (0)	* (0)	1 (100)	100 (1)	100 (1)	100 (1)	100 (1)	100 (1)	100 (1)
C	* (0)	100 (1)	* (0)	1 (100)	100 (1)	100 (1)	100 (1)	100 (1)	100 (1)
D	* (0)	100 (1)	100 (1)	* (0)	1 (100)	100 (1)	100 (1)	100 (1)	100 (1)
E	* (0)	100 (1)	100 (1)	100 (1)	* (0)	1 (100)	100 (1)	100 (1)	100 (1)
F	* (0)	100 (1)	100 (1)	100 (1)	100 (1)	* (0)	1 (100)	100 (1)	100 (1)
G	* (0)	100 (1)	100 (1)	100 (1)	100 (1)	100 (1)	· (0)	1 (100)	100 (1)
H	* (0)	100 (1)	100 (1)	100 (1)	100 (1)	100 (1)	100 (1)	* (0)	1 (100)
A_E	* (0)	* (0)	* (0)	* (0)	* (0)	* (0)	* (0)	* (0)	* (0)

a Arc does not exist.
b pmoles of arc molecules input into a final volume of 56 μl are shown in parentheses.
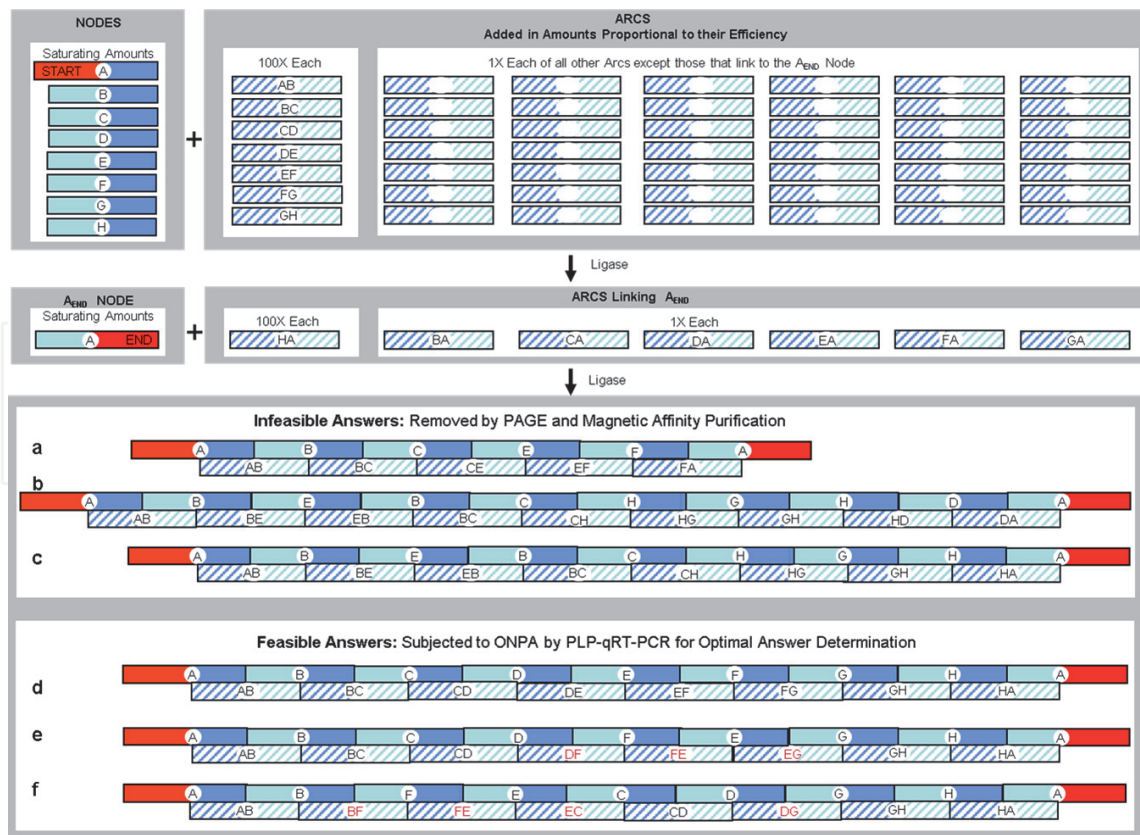*c Distances.*

Table 3.
Arc distance and corresponding concentration matrices used to make the computation.

designed such that the alphabetical order of the nodes was the optimal tour (**Figures 1** and **2d**). To accomplish this, the distances were translated into concentrations (1:1 pmol) of the arc sequences that connect the nodes in inverse proportion to the distances (**Table 3**). These sequences were then hybridized with an excess of all node sequences and ligated to form answer strands as summarized in **Figure 2**. Node and arc sequences were asymmetric in design so that two node sequences could become ligated in an order-specific manner to form an ordered node pair (ONP). For example, to compute the formation of the BC-ONP, the last (3′) and first (5′) halves of the sequences for nodes B and C hybridize to the first (3′) and last (5′) halves of the BC arc oligonucleotide, respectively, which then permits ligase to link them covalently.

Given that the concentration of arcs along the upper diagonal of the efficiency matrix was 100-fold that of the other arcs, strands containing node sequences in alphabetical order were anticipated to be produced in highest abundance. A comparable result was obtained previously using a similar approach to solve an asymmetric, fully-connected 15-city TSP [10]. Consequently, this computational method could be reliably used to examine the efficacy of PLP-qPCR in the answer determination step of a DNA computation.

A tour of the nodes was feasible, whether it was optimal or suboptimal, only if it contained one copy of each node sequence flanked by the A_{start} and A_{end} sequences (**Figure 2d–f**). **Figure 3** shows a stepwise summary of the procedure used to make the calculation. The PAGE profile of answer sequences at various stages of purification is shown in **Figure 4**. Selective amplification of answer sequences flanked by A_{start} and A_{end} (Lane 1) was achieved by PCR using primers specific for those sequences (**Figure 2a–f**). This PCR product (Lane 2) included the 190-mer band that was the size required for feasible answers (**Figure 2c–f**), which was then subjected to sequential magnetic bead affinity purification for every node sequence. This insured that all sequences used for answer analysis contained one and only one copy of each node sequence, and thus represented Hamiltonian circuits through the network. The purified sample composed only of answer DNA sequences for feasible tours (i.e. **Figure 2d–f**) appeared as a single 190-mer band (Lane 3) that was used to determine the optimal answer to the computation.
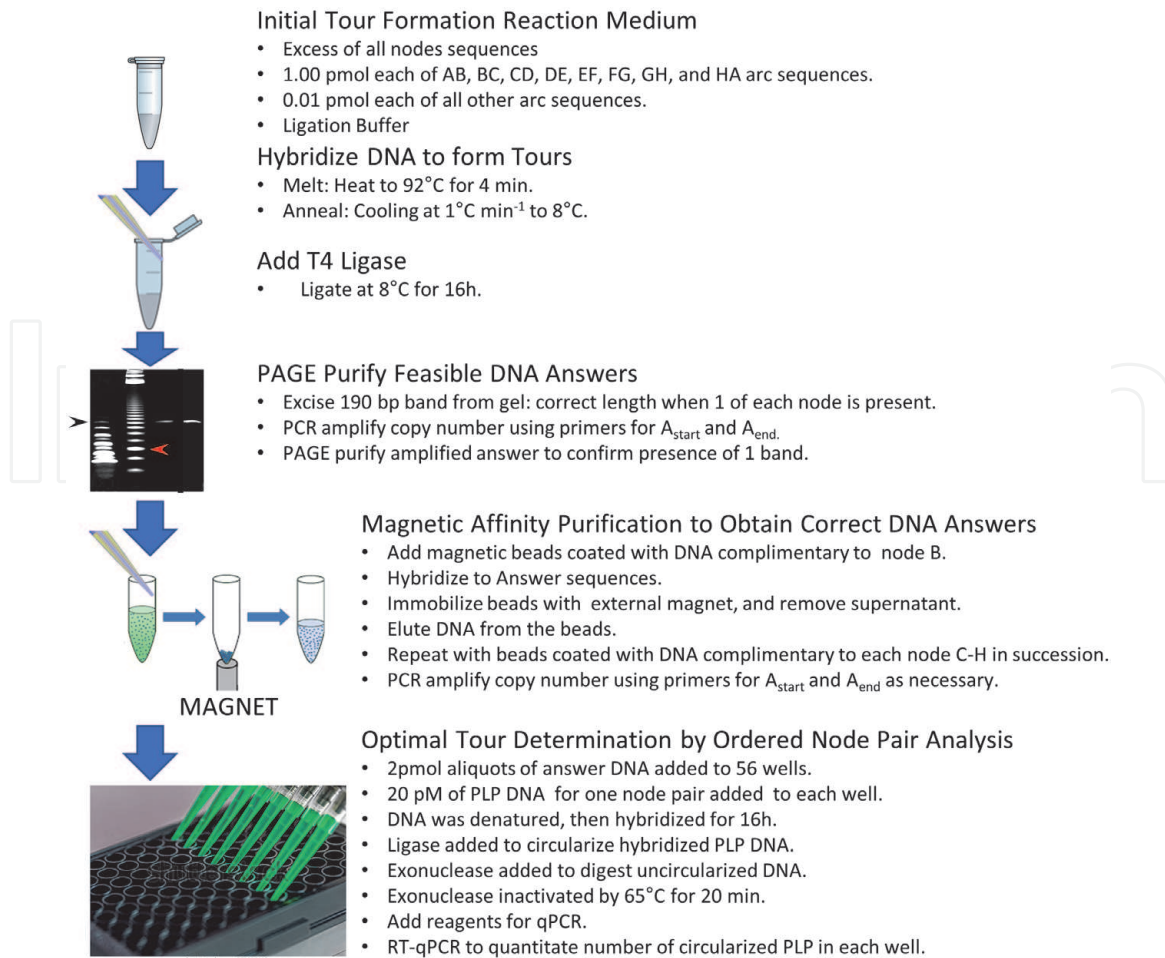
Figure 2.
*Protocol used to compute the 8-city TSP. Answers were formed by the addition of saturating amounts of all nodes and limiting amounts of arcs in inverse proportion to their respective distance. Upon hybridization, nodes and arcs were ligated to form answers. The ligation was accomplished in two stages to increase the probability of forming feasible answers that contained a single copy of each node flanked by the start and end sequences. Infeasible answers were removed by PAGE and by magnetic affinity purification as described in Section 2. Examples of infeasible answer sequences formed include those that are: (a) too short because they lack one or more nodes; (b) too long because they contain multiple copies of nodes; and (c) the correct length but lack at least one node. Feasible answers were subjected to PLP-qPCR for ONPA. Examples of feasible sequences include: (d) the optimal answer sequence with the nodes in alphabetical order; (e) a suboptimal sequence that contains a minimum substitution of 3 arcs from the optimal (see **Table 5**, normalized data row 2), and (f) an even less optimal sequence in which 4 arcs have been substituted from those found in the optimal answer (see **Table 5**, normalized data row 3).*

To identify the optimal answer, the feasible answers (**Figure 4**, Lane 3) were analyzed by the ONPA approach. This was accomplished using PLP-qPCR with the 56 PLPs that were specific to each of the possible ONPs. Each PLP was added to an aliquot of the solution containing the answer set to prepare it for PLP-qPCR. Preparation included the steps summarized in **Figure 5** for the BC-PLP that was used to detect the ONP in which the sequence encoding node B immediately precedes that of node C in a tour. For the BC-ONP, the 5′-arm of the PLP complementary to the last 10 bases of sequence B (designated B′), and the 3′-arm (designated C″) that were complementary to the first 10 bases of sequence C were created [23]. The hybridized PLP was circularized by ligase, after which the PLP was denatured from the target strand and exonuclease was added to eliminate any DNA that had not been circularized. Because the amount of circulated PLP was quantitatively related to the initial content of a target, the target content used in the determination was transformed into the copy number of the circularized PLPs. The presence of a qPCR reporter-identification sequence, and PCR primers in the PLP enabled the use of qPCR to measure the copy number.

Figure 6 shows the PLP-qPCR fluorescence amplification plots as a function of cycle number that determines the amount of each possible ONP in the answer sequences of the 8-city TSP. The ONPA was determined in groups according to the

Figure 3.
Graphical representation of the process to make, and purify DNA sequences that represent correct answers, and the use of PLP sequences for ordered node pair analysis to identify the optimal tour. Details of the methods are described in Section 2.

prior node in the ONP such that the relative abundance of any of the ONPs in that group could be compared directly. In each case, it is evident that the first sample in which fluorescence amplification consistently increased above the threshold corresponded to that ONP where the nodes appeared in alphabetical order (i.e. AB, BC, etc.). Thus, it was possible to determine the optimal answer to the 8-city TSP using PLP-qPCR by simple inspection of the raw data. Due to the quantitative nature of qPCR, the C_t values measured from the data in **Figure 3** were used to calculate the copy number of each ONP (**Table 4**). The results are organized so that each row and column define the prior and subsequent nodes of each ONP. Since these measurements were made in groups corresponding to each row, comparisons of the ONPA between rows were achieved by determining the fractional ONPA after normalizing the ONPA present in each row in highest abundance (data in parentheses). The average copy number for suboptimal ONPs was 0.35% of the optimal ONP in each row. However, the suboptimal ONPs were not present in equal abundance. Seven of the suboptimal ONPs were present in amounts that were below the limit of detection. However, the ability to detect these low abundance ONPs was not needed to obtain the optimal answer to the problem solved here. Some of the suboptimal ONPs were formed in much higher abundance than the average. **Figure 7A** shows the number of suboptimal ONPs from **Table 4** as a percentage of the preceding node in the pair. The fraction of suboptimal ONPs was significantly larger in rows C, D, E and F in **Table 4**. This was due to a relatively small number of specific alternate ONPs that included CE (0.2%), DA (0.2%), EC (0.47%), FD (0.5%), and FE (0.47%).
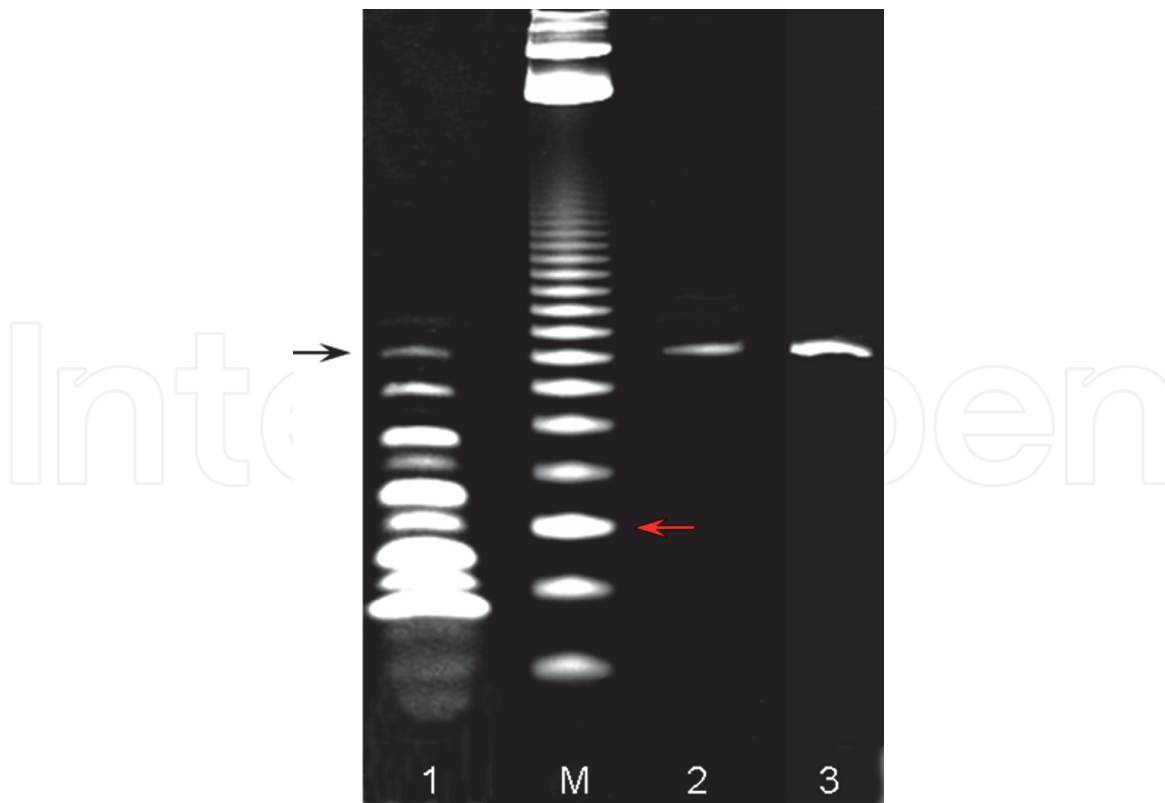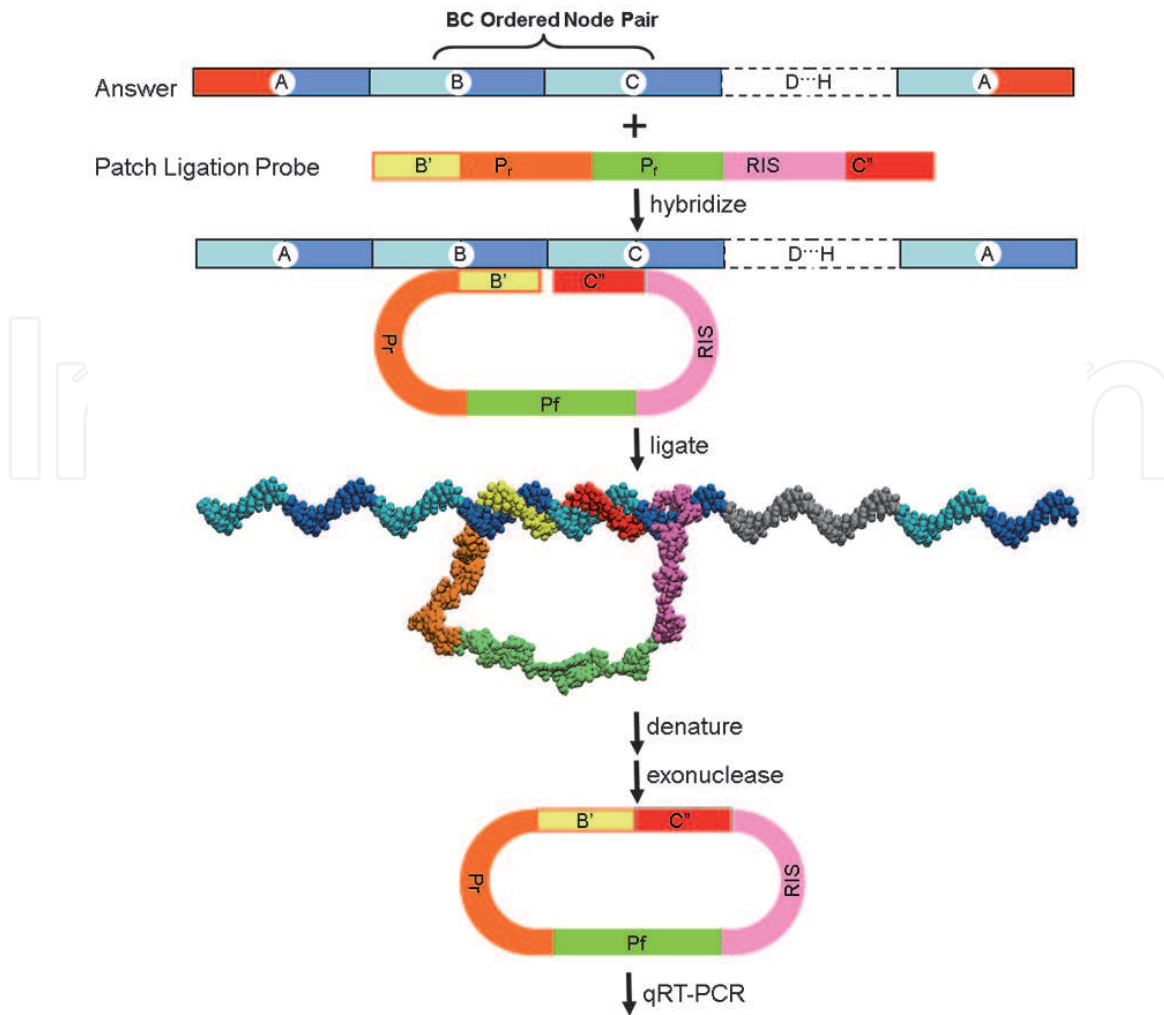
Figure 4.
Profiles of DNA computing products by super denaturing PAGE following amplification by PCR using primers specific to A_S and A_E. 1: Initial product of hybridization/ligation from the answer formation process. Black arrow indicates the 190-mer band that was excised for subsequent purification. 2: Purified 190-mer band excised from Lane 1. 3: Feasible answer sequences following sequential magnetic affinity purification to insure the presence of every node sequence. M: Molecular size reference using a 20-mer DNA ladder where the 100-mer was the brightest band (red arrow).

To test the validity of using ONPA to identify the optimal answer of a TSP, a linear programming (LP) model containing 5168 variables and 64 constraints was developed to "reverse engineer" the frequency of the feasible tours formed by DNA computing that would most closely match the individual ONP frequencies found in **Table 4** when broken down into their constituent ONPs. Since the data for each row in **Table 4** were obtained independently from the others, the model **1–3** was solved by inputting either the normalized and absolute ONP frequencies in **Table 4** as the F_{ij} parameters. In the normalized model, the F_{ij} were converted into percentages of each row total, whereas in the absolute model, the F_{ij} values were the total number of molecules detected (see **Table 5** footnotes a and b).

For the normalized model, the largest individual deviation is for the FG-ONP. The observed abundance F_{FG} was 0.98791. In the optimal solution, summed over all reverse-engineered tours ($\sum_{t=1}^{5040} h_{tFG}X_t$), arc FG was part of 0.99215 of the tours formed. Therefore, the FG-ONP would have a negative deviation $d_{FG}^- = 0.00425$ and a positive deviation $d_{FG}^+ = 0$. Many sets of positive and negative deviations exist that could equalize this constraint. However, the set that contributes the smallest amount to the objective results from the case in which one or both of the positive and negative deviations are zero, because the objective function adds both positive and negative deviations. In addition, there exist many possible solutions with X_t values such that $\sum_{t=1}^{5040} h_{tFG}X_t = F_{FG} = 0.98791$ exactly, but none of them were optimal because it would have increased the deviations for other ONPs. The same set of X_t values appeared in all 64 constraints and, while there is *no* set of X_t values that can satisfy all 64 instances of constraint **2** perfectly with no deviations, the model was designed to minimize the sum of the deviations for all ONPs simultaneously.
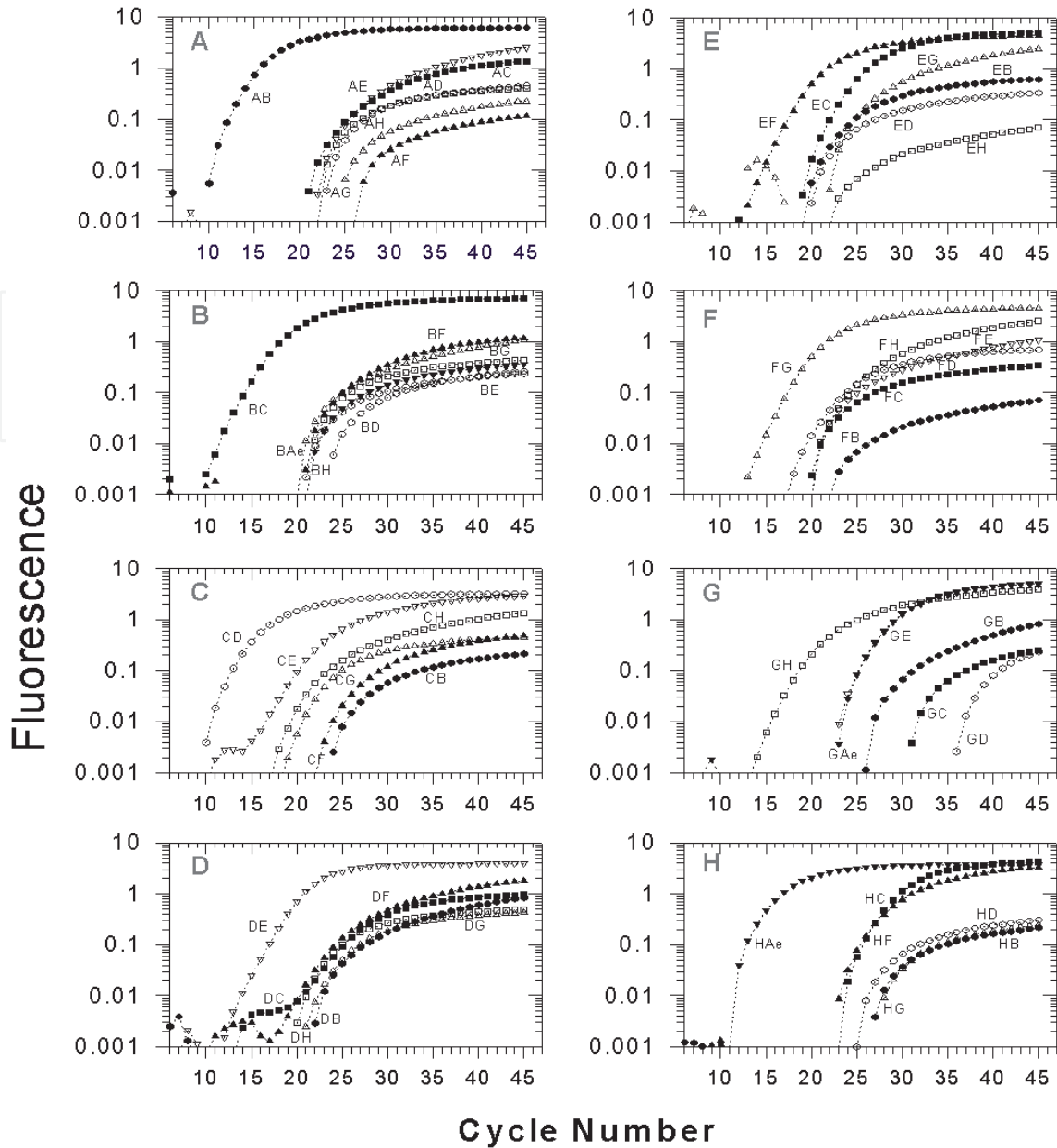
Figure 5.
Use of pair ligation probes (PLPs) to quantify the abundance of the nucleotide sequence encoding the BC ordered node pair (BC-ONP) where node B precedes node C. (1) The PLP hybridizes to 20 consecutive bases of the BC target sequence. (2) T4 ligase circularizes the hybridized PLP. (3) Circularized PLP is denatured from target, and Exonuclease I eliminated any nucleic acid not circularized. (4) qPCR quantifies PLP abundance upon addition of the TaqMan-MGB® reporter, PCR primers and nucleotides.

When the normalized ONPA data from **Table 4** were input, 99.01% of feasible tours formed by the DNA computer are estimated to have been the optimal tour ABCDEFGHA. Given the relative proportions of the ONPs formed, no other tour is likely to have comprised more than 0.38% of the tours formed (**Table 5**), which was the percentage for the second-most frequently formed tour in the reverse-engineering model (A_SBFECDGHA$_E$). Of the 5040 possible tours in the LP model, the optimal reverse-engineered set of tours had only 30 tours with a non-zero frequency. The objective function value in the model of the normalized data (the sum of the positive and negative deviations over all 64 ordered node pairs) was 0.02997.

Using the *absolute* molecule frequencies as the F_{ij} values, the LP model estimated that 339,742 molecules, or 99.13% of all feasible tour molecules, were the optimal answer. The next most frequently produced tour by the DNA computer was esti-mated to have formed only 1259 molecules (**Table 5**). As expected, the error levels increased dramatically when the model was used to reverse-engineer the tours formed from the absolute molecule frequency. This occurred because the row sums of molecules in **Table 4** were determined in separate qPCR measurements without controlling for the absolute amount of answer sequences used in the samples between rows. Had this been held constant during the measurement, the total

Figure 6.
Abundance of ONPs in the feasible answer sequences determined by PLP-qPCR. A separate PLP was designed to detect the amount of each of the 56 possible ONPs (A_SB, A_SC, etc.). Each PLP contained a qPCR reporter identification sequence for use in fluorescence detection with the TaqMan-MGB® reporter for NED. Plots of fluorescence intensity as a function of PCR cycle number are grouped by the preceding node in the ONP. Each of these groups was run as single-plex assays on a 96-well plate using aliquots from a common solution of answer sequences so that the assays within each group are comparable. The latter node in the ONP is indicated as B (●), C (■), D (○), E (▽), F (▲), G (△), H (□), and A_E (▼).

number of molecules in each row of **Table 4** should be the same because each feasible answer contains one copy of each node sequence. The total deviations summed to 9,057,230, which represents 81.57% of the total $\Sigma_i\Sigma_j F_{ij}$, compared with only 0.37% for the normalized model. It is noteworthy that, in the model based on the absolute data, the five ONPs with the greatest deviations from the observed frequencies (HA_E, A_SB, GH, BC, CD) were all part of the optimal tour, whereas only two such ONPs were among the top five deviations (FG, BF, DG, BC, and FD) when matching to the normalized ONPA. Despite the fact that the deviations derived from the absolute ONPs were much higher, the DNA computation still returned the optimal tour over 99% of the time.

A fully connected 8-city TSP was solved by the ordered node pair abundance (ONPA) approach through the use of a pair ligation probe quantitative real time

Prior node	Subsequent node							
	B	**C**	**D**	**E**	**F**	**G**	**H**	**A_E**
A_S	1.4×10^{6a}	10	3.5	2.9	0.019	0.19	1.4	*
B	*	4.8×10^5	0.49	6	10	9	4.9	21
C	0.04	*	2.2×10^5	420	0.18	6.4	22	0
D	28.5	190	*	3.5×10^5	395	320	48.5	750
E	640	1600	380	*	3.4×10^5	87	1.5	0
F	1.5	0	1800	1600	*	3.4×10^5	380	380
G	0.87	0.057	0	47	0	*	1.2×10^5	31
H	0.4	22	1.6	0	47	0.34	*	7.9×10^6

[a]All values are in 1000s of molecules.

Table 4.
ONPA in the set of feasible answers from the 8-city TSP determined by PLP-qPCR.
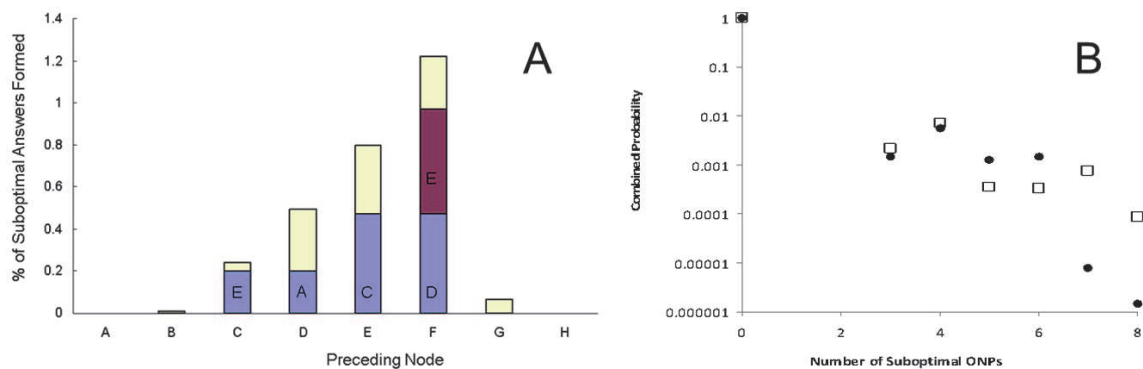
Figure 7.
*(A) Dominant suboptimal ONPs present in the answer sequences. The subsequent node in the ONP is indicated in each bar, and the sum of all other suboptimal ONPs in a given row from **Table 5** is indicated in yellow. (B) Combined probability of tours reverse engineered by the LP Model as a function of the Number of Suboptimal ONPs. The combined probability is the sum of the probabilities of all tours containing a given number of suboptimal ONPs generated from the normalized (●) and absolute data (□ of **Table 5**). The optimal tour did not contain suboptimal ONPs. From 1 to 9 tours were reverse engineered for each number of suboptimal ONPs shown.*

polymerase chain reaction (PLP-qPCR) system. The high specificity of the sequence-tagged hybridization and ligation that results from the use of PLPs significantly increased the accuracy of answer determination in DNA computing. When combined with the high throughput efficiency of qPCR, the time required to identify the optimal answer to the TSP was reduced from days to 25 min.

The reverse-engineering LP model provides additional evidence that the DNA computer can distinguish the quality of solutions to the TSP. Although the problem was set up artificially to have a single solution that was far better than any other possible solution, the reverse-engineering LP model confirms that the optimal solution was produced in far greater abundance than any other answer, and among the suboptimal solutions, those with greater optimality were generally produced in greater abundance than those that were less optimal.

Examination of the less frequently produced tours provides a measure of the sensitivity of DNA computer to the abundance of the arc sequence molecules initially input to carry out the computation. To make any change to the optimal tour A_SBCDEFGHA$_E$, at least three ONPs in the tour must be different. For example, the tour with penultimate optimality (**Figure 2e**) must have 3 ONPs that are different

Normalized ONPA F_{ij}[a]			Absolute ONPA F_{ij}[b]			
Top five tours[c]	# suboptimal ONPs	Probability	Top five tours[d]	# suboptimal ONPs	Molecules (in 1000s)	Probability
$A_SBCDEFGHA_E$	0	0.990131	$A_sBCDEFGHA_E$	0	339,742	0.991252
$A_SBFECDGHA_E$	4	0.003831	$A_sBFDECGHA_E$	4	1,259	0.003672
$A_SFDEBCGHA_E$	4	0.001860	$A_sFDEBCGHA_E$	4	512	0.001494
$A_SBCEDHGFA_E$	6	0.001099	$A_sBCEDFGHA_E$	3	258	0.000754
$A_SBCDFEGHA_E$	3	0.000733	$A_sBFECDGHA_E$	4	217	0.000634

[a]*Based on ONPA as a percent of total molecules in each row in **Table 4** (i.e. ΣF_{ij} on Row A_s = 1.00, ΣF_{ij} on Row B = 1.00, etc.).*
[b]*Based on ONPA in **Table 5** (i.e. ΣF_{ij} on Row A_s = 1.5 e9, ΣF_{ij} on Row B = 4.5 e8, etc.).*
[c]*A total of 30 non-zero tours were reversed engineered from Normalized data with 0.0299702 total absolute deviations (objective function value).*
[d]*A total of 32 non-zero tours were reverse engineered from Absolute data with 9,057,230,000 total absolute deviations (objective function value).*

Table 5.
Tour probabilities reverse engineered from observed ONPA by the LP model.

from the optimal (**Figure 2d**), while tours containing 4 different predominant ONPs (**Figure 2f**) would be even less optimal, and so forth. As shown in **Table 5**, among the 2nd-5th most frequently produced tours for models predicted from both the absolute and normalized data, all but one exhibit either three or four ONP changes from the optimal tour. When the probabilities of the tours were summed over all tours with a given number of suboptimal ONPs, the frequency of tours formed with more suboptimal ONPs exponentially declined as a function of an increase in the number of suboptimal ONPs incorporated (**Figure 7B**). Thus, not only did the DNA computer generate the optimal answer far more frequently, but other answers were generated in amounts that were inversely proportional to their objective function value.

Even though the data of **Table 4** were obtained in separate sets grouped according to each row without controlling for the abundance of molecules between rows (absolute data), the computation returned the correct answer, demonstrating the robustness of this approach. Several ONPs were not detected as products of the computation (e.g. FB). However, this does not imply that no molecules of these ONPs were formed. Although the amounts of these ONPs were below the level of detection, this did not alter the determination of the optimal answer to this problem because the efficiency of the tour that gave rise to the optimal answer was so much greater than any other tour. Identification of the optimal answers to some less obvious problems may require a more sensitive discrimination of ONPA. Under these circumstances, to quantify smaller amounts of ONPs or to discriminate between smaller differences in abundance of two specific ONPs, the PLP-qPCR can be repeated using larger aliquots of feasible answer solutions.

Acknowledgements

Author details

Fusheng Xiong[1], Michael Kuby[2] and Wayne D. Frasch[1*]

1 School of Life Sciences, Arizona State University, Tempe, AZ, USA

2 School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ, USA

*Address all correspondence to: frasch@asu.edu

IntechOpen

References

[1] Lin S. Computer solutions of traveling salesman problem. Bell System Technical Journal. 1965;**44**:2245

[2] Rosenkrantz DJ, Stearns RE, Lewis PM. An analysis of several heuristics for the traveling salesman problem. SIAM Journal of Computing. 1977;**6**:563-581

[3] Crowder H, Padberg MW. Solving large-scale symmetric traveling salesman problems to optimality. Management Science. 1980;**26**:495-509

[4] Jünger M, Reinelt G, Rinaldi G. The traveling salesman problem. In: Ball MO, Monma CL, Nemhauser GL, editors. Handbooks in Operations Research and Management Science. Amsterdam: Elsevier Science B.V; 1995. pp. 225-330

[5] Ryu H. A revisiting method using a covariance traveling salesman problem algorithm for landmark-based simultaneous localization and mapping. Sensors. 2019;**19**:E4910

[6] Miao K, Duan H, Qian F, Dong Y. A one-commodity pickup-and-delivery traveling salesman problem solved by a two-stage method: A sensor relocation application. PLoS One. 2019;**14**: e0215107

[7] Kahng AB, Reda S. Match twice and stitch: A new TSP tour construction heuristic. Operations Research Letters. 2004;**32**:499-509

[8] Adleman LM. Molecular computation of solutions to combinational problems. Science. 1994; **266**:1021-1024

[9] Lipton RJ. DNA solution of hard computational problems. Science. 1995; **268**:542-545

[10] Lee JY, Shin SY, Park TH, Zhang BT. Solving traveling salesman problems with DNA molecules encoding numerical values. Biosystems. 2004;**78**: 39-47

[11] Spetzler D, Ziong F, Frasch WD. Heuristic solution to a 10-city Asymmetric traveling salesman problem using probabilistic DNA Computing. Lecture Notes in Computer Science. 2008;**4848**:152-160

[12] Xiong FS, Spetzler D, Frasch WD. Solving the fully-connected 15-city TSP using probabilistic DNA computing. Integrative Biology. 2009;**1**:275-280

[13] Sharma D, Ramteke M. A note on short-term scheduling of multi-grade polymer plant using DNA computing. Chemical Engineering Research and Design. 2018;**135**:78-93

[14] Xu F, Wu T, Shi X, Pan L. A study on a special DNA nanotube assembled from two single-stranded tiles. Nanotechnology. 2019;**30**(115602):1-6

[15] Woods D, Doty D, Myhrvold C, Hui J, Zhou F, Yin P, et al. Diverse and robust molecular algorithms using reprogrammable DNA self-assembly. Nature. 2019;**567**:366-372

[16] Yamamoto M, Kameda A, Matsuura M, Shiba T, Kawazoe Y, Ohuchi A. A separation method for DNA computing based on concentration control. New Generation Computing. 2002;**20**:251-261

[17] Lee J-Y, Shin S-Y, Augh SJ, Park TH, Zhang B-T. Temperature gradient-based DNA computing for graph problems with weighted edges. Lecture notes in Computer Science. 2003;**2568**:73-84

[18] Yamamura M, Hiroto Y, Matoba T. Solutions of shortest path problems by

concentration control. Lecture Notes in Computer Science. 2002;**2340**:231-240

[19] Henco K, Harders J, Wiese U, Riesner D. Temperature gradient gel electrophoresis (TGGE) for the detection of polymorphic DNA and RNA. Methods in Molecular Biology. 1994;**31**:211-228

[20] Riesner D, Steger G, Wiese U, Wulfert M, Heibey M, Henco K. Temperature-gradient gel electrophoresis for the detection of polymorphic DNA and for quantitative polymerase chain reaction. Electrophoresis. 1992;**13**:632-636

[21] Ibrahim Z, Rose JA, Suyama A, Khalid M. Experimental Implementation and analysis of a DNA computing readout method based on real-time PCR with TaqMan probes. Natural Computing. 2008;**7**:277-286

[22] Szemes M, Bonants P, de Weerdt M, Baner J, Landegren U, Schoen CD. Diagnostic application of padlock probes-multiple detection of plant pathogens using universal microarrays. Nucleic Acids Research. 2005;**33**:e70

[23] Xiong F, Frasch WD. Padlock probe-mediated qRT-PCR for DNA computing answer determination. Natural Computing. 2010;**10**:947-959

[24] Brüggemann W. A minimal cost function method for optimizing the age-depth relation of deep-sea sediment cores. Paleoceanography. 1992;7: 467-487

[25] Kuby MJ, Cerveny RS, Dorn RI. A new approach to paleoclimatic research using linear programming. Palaeogeography Palaeoclimatocology Palaeoecology. 1997;**129**:251-267

[26] Leinen M, Pisias N. An objective technique for determining end-member compositions and for partitioning sediments according to their sources. Geochimica et Cosmochimica Acta. 1984;**48**:47-62

[27] Narula SC, Wellington JF. Selection of variables in linear-regression using the minimum sum of weighted absolute errors Criterion. Technometrics. 1979; **21**:299-306