Int. J. Bio-Inspired Computation, Vol. 10, No. 1, 2017

Programmable DNA-mediated decision maker

Jian-Jun Shu*, Qi-Wen Wang and Kian-Yan Yong

School of Mechanical and Aerospace Engineering,
Nanyang Technological University,
50 Nanyang Avenue, 639798, Singapore
Email: mjjshu@ntu.edu.sg
Email: m090002@e.ntu.edu.sg
Email: kenny_yong@e.ntu.edu.sg
*Corresponding author

Abstract: DNA-mediated computing is a novel technology that seeks to capitalise on the enormous informational capacity of DNA and has tremendous computational ability to compete with the current silicon-mediated computing, due to massive parallelism and unique characteristics inherent in DNA interaction. In this paper, the methodology of DNA-mediated computing is utilised to enrich decision theory, by demonstrating how a novel programmable DNA-mediated normative decision-making apparatus is able to capture rational choice under uncertainty.

Keywords: deoxyribonucleic acid; DNA; processor; material; programmable biochemical operator.

Reference to this paper should be made as follows: Shu, J-J., Wang, Q-W. and Yong, K-Y. (2017) 'Programmable DNA-mediated decision maker', *Int. J. Bio-Inspired Computation*, Vol. 10, No. 1, pp.51–55.

Biographical notes: Jian-Jun Shu is a recipient of the British Institution of Mechanical Engineers 1992 BFPA Prize for Young Engineers. He is a member of the editorial board of the journals *(Nature) Scientific Reports* and *Mathematical Problems in Engineering*. Since 2000, he has been with Nanyang Technological University, Singapore. His research interests focus on DNA-mediated computing.

Qi-Wen Wang received his BEng in 2009 from Nanyang Technological University, Singapore, PhD in 2014 from the same university under the supervision of Dr. Jian-Jun Shu. His research interests focus on DNA-mediated computing.

Kian-Yan Yong received his BEng in 2004 from Nanyang Technological University, Singapore, MEng in 2007 and PhD in 2013, respectively from the same university under the supervision of Dr. Jian-Jun Shu. His research interests focus on DNA-mediated computing.

1 Introduction

In nature, most of our daily activities, such as interacting with information received from the outside world and responding with appropriate reactions, are processed in our brains, which is a magnificent decision-making apparatus. The actual mechanism on how our genetic material, deoxyribonucleic acid (DNA), is translated into functional amino acids and eventually dedicated in the construction of sophisticated decision-making system still remains mysterious. The intrinsic characteristics of DNA, on the other hand, reveal the feasibility that artificially-synthesised DNA can be utilised as information storage and processing substrate, and ultimately assembled as a splendid normative decision-making apparatus, which is capable of making rational decision based on received information.

DNA-mediated computational ability, through the in-vitro manipulation of artificially-encoded DNA sequences, has been successfully demonstrated to cope with some intriguing conundrums, including Hamiltonian path problem (Adleman, 1994), satisfaction problem (Lipton, 1995), maximal clique problem (Ouyang et al., 1997) and strategic assignment problem (Shu et al., 2011). These ideas have ignited the potential power of assembling artificially-synthesised DNA as an astonishing computational device (Shu et al., 2015; Wong et al., 2015). Most recently, further evidence reveals that DNA can be utilised in the construction of molecular logic circuits (Saghatelian et al., 2003; Seelig et al., 2006; Frezza et al., 2007; Qian and Winfree, 2011), simple artificial intelligent gaming device (Pei et al., 2010), and even the neural networks (Qian et al., 2011). In this paper, a novel programmable DNA-mediated normative decision-making apparatus is proposed.

2 Normative decision-making apparatus

Normative decision-making apparatus is a typical device, which is employed to simulate the rational behaves of individuals facing risky choices. Expected utility (EU) hypothesis (Keeney and Raiffa, 1993), which is vastly used in the analysis of decision making under risk, is adopted as a normative model of rational choice.

The classical EU model of decision making under risk can be abstractly described as follows: Given a situation with a set of alternative options O_i and by assuming that individual decision maker follows a series of predefined axioms, the rational choice is subjected to the option, whereas the EU U as described by (1) is maximised.

$$U = \max \sum P_j u\left(x_j\right) \tag{1}$$

where P_j denotes the probability at each possible outcome x_j, and $u(x_j)$ denotes the utility of receiving outcome x_j.

The programmable DNA-mediated normative decision-making apparatus is employed to achieve the identical function by means of the in-vitro manipulation of artificially-encoded DNA molecules. A simple case study, which is modified from that of the Ellsberg (1961) paradox, is selected for demonstration purpose. It is a ball-selecting game, which involves 90 balls in an urn. The number of ball is known to the decision maker: 40 red balls (R), 30 black balls (B), and 20 white balls (W).

The decision maker is offered with three distinct gambling scenarios – option 1: receive $u(20)$ if a red or black ball is drawn and $u(0)$ if a white ball is drawn; option 2: receive $u(20)$ if a red or white ball is drawn and $u(0)$ if a black ball is drawn; option 3: receive $u(20)$ if a black or white ball is drawn and $u(0)$ if a red ball is drawn. In the remaining part of the paper, $u(20)$ and $u(0)$ are interchangeably referred to as favourable outcome and unfavourable outcome, respectively. The question of the case study is which gambling scenario of three alternative options offers the best outcome to the decision maker. The entire case can be represented in terms of the decision matrix as shown in Table 1.

For the sake of convenience, the decision matrix can be transformed into an equivalent decision tree. The choice node stands for the point that decision maker picks up a gambling scenario from a set of predefined alternative options. The chance node is associated with various objectives probabilities leading to distinct outcomes. The chance node can be classified into two types, known as the favourable outcome $u(20)$, and unfavourable outcome $u(0)$. In addition to the decision tree, one additional node, which is named as termination node, is introduced in the graph. By doing so, the original decision tree is converted into a directed network, which begins at the choice node and ends at the termination node, as shown in Figure 1.

Figure 1 Decision tree (see online version for colours)

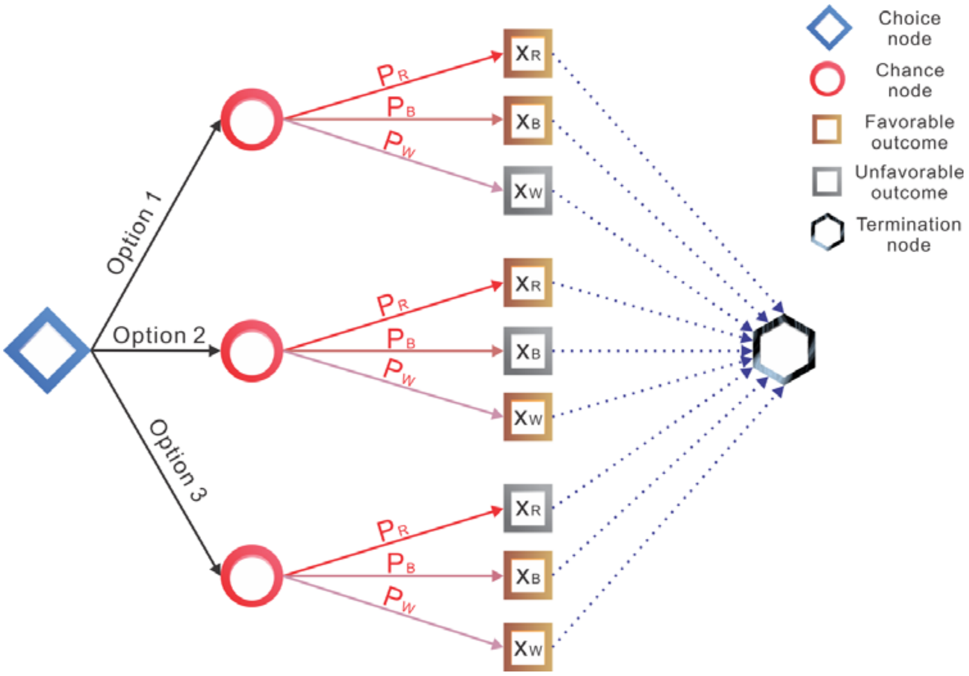
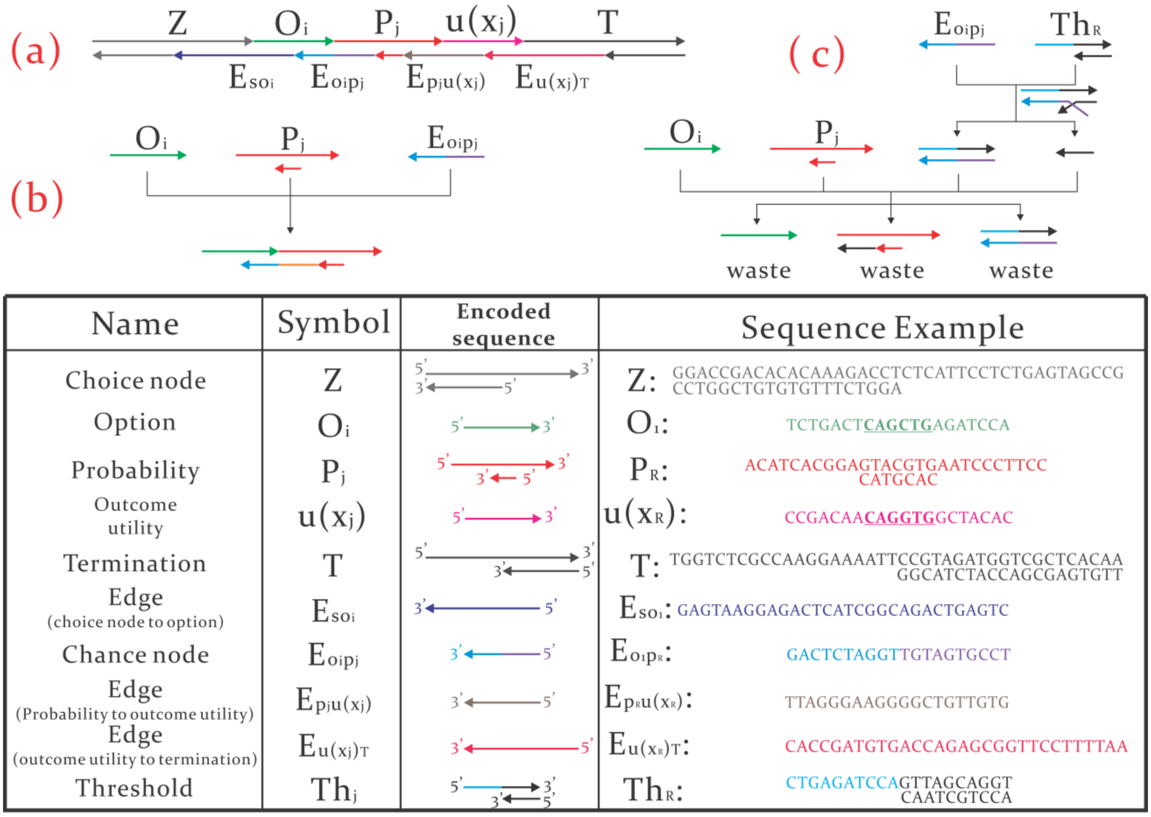


Figure 2 DNA sequence design motif (see online version for colours)



Name	Symbol	Encoded sequence	Sequence Example
Choice node	Z	5'→3' 3'←5'	Z: GGACCGACACACAAAGACCTCTCATTCCTCTGAGTAGCCG CCTGGCTGTGTGTTTCTGGA
Option	O_i	5'→3'	O_1: TCTGACT**<u>CAGCTG</u>**AGATCCA
Probability	P_j	5'→3' 3'←5'	P_R: ACATCACGGAGTACGTGAATCCCTTCC CATGCAC
Outcome utility	$u(x_j)$	5'→3'	$u(x_R)$: CCGACAA**<u>CAGGTG</u>**GCTACAC
Termination	T	5'→3' 3'←5'	T: TGGTCTCGCCAAGGAAAATTCCGTAGATGGTCGCTCACAA GGCATCTACCAGCGAGTGTT
Edge (choice node to option)	E_{so_i}	3'←5'	E_{so_1}: GAGTAAGGAGACTCATCGGCAGACTGAGTC
Chance node	$E_{o_ip_j}$	3'←5'	$E_{o_1p_R}$: GACTCTAGGTTGTAGTGCCT
Edge (Probability to outcome utility)	$E_{p_ju(x_j)}$	3'←5'	$E_{p_Ru(x_R)}$: TTAGGGAAGGGGCTGTTGTG
Edge (outcome utility to termination)	$E_{u(x_j)T}$	3'←5'	$E_{u(x_R)T}$: CACCGATGTGACCAGAGCGGTTCCTTTTAA
Threshold	Th_j	5'→3' 3'←5'	Th_R: CTGAGATCCAGTTAGCAGGT CAATCGTCCA

Table 1 Decision matrix

Probability	*Outcomes*		
	X_R = 40	*X_B = 30*	*X_W = 20*
	4/9	*1/3*	*2/9*
Option 1	*u*(20)	*u*(20)	*u*(0)
Option 2	*u*(20)	*u*(0)	*u*(20)
Option 3	*u*(0)	*u*(20)	*u*(20)

3 Problem encoding

As being analogous to the very beginning step of contemporary silicon-mediated computer, whereas information is digitised in terms of binary expressions, the information as depicted in Figure 1 is converted into a combination of DNA sequences – adenine (A), thymine (T), guanine (G), and cytosine (C) – based upon the DNA sequence design motif specified in Figure 2. As the several section of proposed design motif is similar as described in (Shu et al., 2011, 2015), a detailed explanation of similar parts is therefore eliminated in this paper. By doing so, it is possible to concentrate on the newly-created problem encoding strategies.

It is to begin with the encoding scheme for option edges, as instructed in Figure 1. Each edge of option O_i, where $i \in \{1, 2, 3\}$, is encoded with 20-mer single-stranded DNA (ssDNA). To avoid undesired hybridisation in the subsequent procedure, all DNA sequences used to represent the option edges, and the remaining DNA sequences engaged in hybridisation, are accomplished based on the rules as specified in Tanaka et al. (2005). In addition to that, each DNA sequence of O_i contains a recognition site (as underlined in Figure 2) with respect to one specific restriction enzyme – option 1: *Proteus vulgaris* (PvuII); option 2: *Haemophilus parainfluenzae* (HpaI); option 3: *Streptomyces tubercidicus* (StuI). The selected restriction enzymes have two properties in common – all these chosen enzymes are only effective on double-stranded DNA (dsDNA) and cleave original 20 base pair (bp) dsDNA into two dsDNA of 10 bp length ended up with blunt ends.

The probabilities of picking red ball (P_R), black ball (P_B), and white ball (P_W) as mentioned in decision tree (Figure 1) are represented by introducing another class of DNA sequences, namely, threshold DNA sequence (Th_j), where $j \in \{R, B, W\}$. Without the presence of DNA sequence (Th_j), the DNA sequence of chance node $(E_{O_iP_j})$ is capable of joining the corresponding DNA sequence of option O_i and probability P_j, as illustrated in Figure 2(b), due to its designated nature – the former 10-mer ssDNA is complementary to the rear 10-mer ssDNA of O_i, and the rear 10-mer ssDNA is complementary to the former 10-mer stick end of P_j. The obtained sequence promotes the formation of DNA sequence as described in Figure 2(a). By manually introducing the threshold DNA sequence (Th_j) into DNA solution, DNA sequence representing chance node $(E_{O_iP_j})$ is likely to react with the corresponding threshold DNA sequence (Th_j) due to toehold mediated strand displacement mechanism. The resultant DNA sequences containing option O_i and probability P_j lead to three distinct parts, as shown in Figure 2(c). All these parts are regarded as waste because none of them contributes to the formation of DNA sequence as described in Figure 2(a).

By following the toehold mediated strand displacement mechanism, the probabilities of picking red ball (P_R), black ball (P_B), and white ball (P_W), as instructed in Figure 1, can be therefore represented by using various concentration of threshold DNA sequence (Th_j). The concentration ratio R_i between threshold DNA sequence (Th_j) and chance node DNA sequence ($E_{O_iP_j}$) is estimated by using formula (2):

$$R_i = 1 - P_i \qquad (2)$$

Therefore, in this study case, the ratios among threshold DNA sequences of red ball, black ball, and white ball are 5/9, 6/9, and 7/9, respectively.

As a result of gel electrophoresis, the DNA molecules of different lengths are used to indicate various concentrations. The DNA sequences representing probability (P_j), as instructed in Figure 2, are the double-stranded DNA (dsDNA) with two identical-length sticky ends. The length of the middle section (dsDNA) is a variable, depending on the following rule: The concentration of DNA is inversely proportional to the assigned length of dsDNA. Based on the same ratio as discussed above, the length of the middle section of probability DNA sequences P_j are set to be 7 bp for P_R, 16 bp for P_B, and 34 bp for P_W.

Finally, the outcome utility DNA sequence $u(x_j)$ is encoded with 20-mer single-stranded DNA (ssDNA). As being similar to the option DNA sequence O_i, each DNA sequence of outcome utility contains a recognition site (as underlined in Figure 2) with respect to one specific restriction enzyme – $u(x_R)$: *Pseudomonas maltophilia* (PmlI); $u(x_B)$: *Escherichia coli* (EcoRV); $u(x_W)$: *Streptomyces caespitosus* (ScaI). The underlying principle is that, if the resultant utility of outcome is favourable, $u(x_j) = u(20)$, the presence of related restriction enzyme should be inhibited in the final mixture of DNA solutions. Otherwise, the assigned restriction enzyme is required to be added into the solution.

4 Implementation

After all DNA strands are synthesised in individual test tubes based on the DNA sequence design motif (Figure 2), a small amount (0.1 ml of solution at concentration of 0.1 µg/µl) of DNA amount from all test tubes is extracted and mixed in one new test tube. At the end of hybridisation, T4 DNA ligase is added into the solution, and sufficient time is allowed for the reaction.

The mixed solution is equally divided into three new test tubes, which are labelled as test tube 1 (*TT1*), test tube 2 (*TT2*), and test tube 3 (*TT3*). After that, three distinct restriction enzymes, HpaI, StuI, and ScaI, are added into *TT1*. The addition of the former two restriction enzymes, HpaI and StuI, are to cleave the DNA molecules, which contain the sequence representing the option O_2 and O_3. The concept is to retain only DNA molecules containing the specific sequence of option O_1. The addition of the last restriction enzyme ScaI is to eliminate the solution containing sequences representing the unfavourable outcome $u(0)$. By following the guideline, restriction enzymes, PvuII, StuI, EcoRV and PvuII, HpaI, PmlI are added into *TT2* and *TT3*, respectively. All three test tubes, containing a mixture of DNA solutions, proper restriction enzyme, and appropriate master mix, are immersed into a water tank with water temperature maintaining constant at 37°C for effective incubation. As a consequence of restriction enzyme digestion, the DNA sequences that do not satisfy the design paradigm are not amplified in the subsequent procedure, and ultimately, removed as the result of solution purification.

Polymerase chain reaction (PCR) is utilised to amplify, or to make the duplicate copies of target DNA sequence by varying the 'input signal' known as the DNA primer. Two DNA oligonucleotides, the sequences '3' CCTGGCTGTG 5'' and '3' AGCGAGTGTT 5', are exploited as the DNA primers for all three test tubes. After n complete thermal cycles of PCR, the amount of target DNA sequences is multiplied by 2^n. PCR has two distinct features in this model: One is to selectively amplify and retain the DNA sequences beginning with choice node Z and ending with termination node T; Another is to quantitatively enlarge concentration difference among amplified DNA sequences (prior to PCR, the initial concentration of amplified DNA solutions are distinct due to the involvement of threshold DNA sequences). For instance, the initial quantitative difference between any two DNA solutions is one unit. After, let us say, five complete thermal cycles of PCR, the difference is extended to 32 units, which is 32 times as compared with the initial difference. Such a difference can be easily captured by using spectrophotometer.

Prior to the gel electrophoresis, the solution obtained from PCR is subjected to the solution purifier, which significantly improves the separation outcome. The solution purifier is utilised to remove the 'noise signal', including DNA waste due to toehold mediated strand displacement method, cleaved DNA sequences due to restriction enzymes, unamplified DNA sequences, excessive DNA oligonucleotide (primer), and other degradation factors like salts.

Purified DNA solution in all test tubes (*TT1*, *TT2*, and *TT3*) are inserted into a pre-made 4 – well agarose gel slab immersed in the TAE (a mixture of Tris base, acetic acid, and EDTA) buffer. For optimal DNA resolution, the recommended agarose gel concentration is 2.5 – 3%. The well labelled '1', '2', and '3' are used to hold the DNA solution with the same labelling number as that of test tubes. Well '4' contains the 10 bp DNA ladder, which holds 33 fragments ranging in size from 10 bp to 200 bp. After that, the gel slab is subjected to a constant electric field. Due to the working principle of gel electrophoresis, the final DNA sequence containing the shorter portion of probability sequence (higher probability) migrates faster towards anode electrode through the pores of gel matrix as compared with that of longer DNA probability sequence. For favourable resolution, the process terminates once the visible tracking dye (XCFF), whose migration speed is close to 100 bp double-stranded DNA, migrates 2/3 distance of the entire length of gel slab. Once the stained agarose gel is projected under the UV light of 302 nm wavelength, each well (from well ‘1’ to well ‘3’) contains two DNA bands – well ‘1’: 147 bp of relative concentration 4 and 156 bp of relative concentration 3; well ‘2’: 147 bp of relative concentration 4 and 174 bp of relative concentration 2; well ‘3’: 156 bp of relative concentration 3 and 174 bp of relative concentration 2. Consequently, based on the observations obtained from the gel, it is possible to conclude that the gambling scenario 1 among all three alternatives offers the best possible outcome for the decision maker.

5 Conclusions

DNA-mediated computing has inestimable potentials to be extended into other seemingly-unrelated disciplines, and ultimately changes the existing configurations in these fields. In this paper, DNA-mediated computing has been correlated with the discipline of normative decision theory. A case study is made to illustrate how to handle DNA-mediated decision making. It is believed that the DNA-mediated technique proposed in this paper may facilitate effective complex decision making in a rapidly changing environment.

References

Adleman, L.M. (1994) ‘Molecular computation of solutions to combinatorial problems’, *Science*, Vol. 266, No. 5187, pp.1021–1024.

Ellsberg, D. (1961) ‘Risk, ambiguity, and the savage axioms’, *Quarterly Journal of Economics*, Vol. 75, No. 4, pp.643–669.

Frezza, B.M., Cockroft, S.L. and Ghadiri, M.R. (2007) ‘Modular multi-level circuits from immobilized DNA-based logic gates’, *Journal of the American Chemical Society*, Vol. 129, No. 48, pp.14875–14879.

Keeney, R.L. and Raiffa, H. (1993) *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, Cambridge University Press, Cambridge, UK.

Lipton, R.J. (1995) ‘DNA solution of hard computational problems’, *Science*, Vol. 268, No. 5210, pp.542–545.

Ouyang, Q., Kaplan, P.D., Liu, S.M. and Libchaber, A. (1997) ‘DNA solution of the maximal clique problem’, *Science*, Vol. 278, No. 5337, pp.446–449.

Pei, R.J., Matamoros, E., Liu, M.H. Stefanovic, D. and Stojanovic, M.N. (2010) ‘Training a molecular automaton to play a game’, *Nature Nanotechnology*, Vol. 5, No. 11, pp.773–777.

Qian, L.L. and Winfree, E. (2011) ‘Scaling up digital circuit computation with DNA strand displacement cascades’, *Science*, Vol. 332, No. 6034, pp.1196–1201.

Qian, L.L., Winfree, E. and Bruck, J. (2011) ‘Neural network computation with DNA strand displacement cascades’, *Nature*, Vol. 475, No. 7356, pp.368–372.

Saghatelian, A., Völcker, N.H., Guckian, K.M., Lin, V.S-Y. and Ghadiri, M.R. (2003) ‘DNA-based photonic logic gates: AND, NAND, and INHIBIT’, *Journal of the American Chemical Society*, Vol. 125, No. 2, pp.346–347.

Seelig, G., Soloveichik, D., Zhang, D.Y. and Winfree, E. (2006) ‘Enzyme-free nucleic acid logic circuits’, *Science*, Vol. 314, No. 5805, pp.1585–1588.

Shu, J-J., Wang, Q-W. and Yong, K-Y. (2011) ‘DNA-based computing of strategic assignment problems’, *Physical Review Letters*, Vol. 106, No. 18, p.188702.

Shu, J-J., Wang, Q-W., Yong, K-Y., Shao, F. and Lee, K.J. (2015) ‘Programmable DNA-mediated multitasking processor’, *Journal of Physical Chemistry B*, Vol. 119, No. 17, pp.5639–5644.

Tanaka, F., Kameda, A., Yamamoto, M. and Ohuchi, A. (2005) ‘Design of nucleic acid sequences for DNA computing based on a thermodynamic approach’, *Nucleic Acids Research*, Vol. 33, No. 3, pp.903–911.

Wong, J.R., Lee, K.J., Shu, J-J. and Shao, F. (2015) ‘Magnetic fields facilitate DNA-mediated charge transport’, *Biochemistry*, Vol. 54, No. 21, pp.3392–3399.