

DNA Code Word Design for DNA Computing with Real-Time Polymerase Chain Reaction

¹Muhammad Faiz Mohamed Saaid, ¹Zuwairie Ibrahim, ¹Mohamad Shukri Zainal Abidin,
¹Marzuki Khalid, and ²Nor Haniza Sarmin
¹Centre for Artificial Intelligence and Robotics (CAIRO),
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor Darul Takzim, Malaysia
²Department of Mathematics, Faculty of Science, Universiti Teknologi Malaysia,
81310 UTM Skudai, Johor Darul Takzim, Malaysia

Abstract: Problem statement: A number of DNA computing models to solve mathematical graph problem such as the Hamiltonian Path Problem (HPP), Traveling Salesman Problem (TSP), and the Shortest Path Problem (SPP), have been proposed and demonstrated. Normally, the DNA sequences used for the computation should be critically designed in order to reduce error that could occur during a computation. We have proposed a DNA computing readout method tailored specifically to HPP in DNA computing using real-time Polymerase Chain Reaction (PCR). The DNA sequences were designed based on a procedure and DNASequenceGenerator was employed to generate the sequences required for the experiment. The drawback of the previous approach is that a pool of DNA sequences need to be generated by DNASequenceGenerator before the selection is done manually, based on several design constraints. Hence, an automatic and systematic approach is needed to generate the DNA sequences based on design constraints. **Approach:** In this study, a generate-and-test approach was proposed for the same problem subjected to several design constraints. The generate-and-test algorithm consists of two main levels. The first level considered the basic constraints of DNA sequence design, which were melting temperature, GC-percentage, similarity, continuity, hairpin, and H-measure. This was followed by the second level that includes specific constraints formulated based on five rules, which had been used in previous study. A generated sequence was chosen only if the sequence satisfies all the basic and specific constraints. **Results:** Sequences designed by generate-and-test approach have higher H-measure value than sequences generated by DNASequenceGenerator. However, the generated sequences show lower value for similarity as well as for additional constraints compared to sequences designed by DNASequenceGenerator. **Conclusion:** The generated DNA sequences were better compared to the sequences, obtained from DNASequenceGenerator.

Key words: DNA computing, generate-and-test, real-time polymerase chain reaction, TaqMan probes

INTRODUCTION

Since the discovery of PCR^[1], numerous applications have been explored, primarily in the life sciences and medicine, and importantly, in DNA computing as well. The subsequent innovation of real-time PCR has rapidly gained popularity and plays a crucial role in molecular medicine and clinical diagnostics^[2]. All real-time amplification instruments require a fluorescence reporter molecule for detection and quantitation, whose signal increase is proportional to the amount of amplified product. Although a number of reporter molecules currently exist, it has been found that the mechanism of the TaqMan hydrolysis probe is

very suitable for the design and development of readout method for DNA computing.

A TaqMan DNA probe is a modified, non-extendable dual-labeled oligonucleotides. The 5' and 3' ends of the oligonucleotide are terminated with an attached reporter, such as FAM, and quencher fluorophore dyes, such as TAMRA, respectively, as shown in (Fig. 1)^[3]. Upon laser excitation at 488 nm, the FAM fluorophore, in isolation emits fluorescence at 518 nm. Given proximity of the TAMRA quencher, however, based on the principle of Fluorescence Resonance Energy Transfer (FRET), the excitation energy is not emitted by the FAM fluorophore, but rather is transferred along the sugar-phosphate-

Corresponding Author: Zuwairie Ibrahim, Centre for Artificial Intelligence and Robotics (CAIRO), Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor Darul Takzim, Malaysia

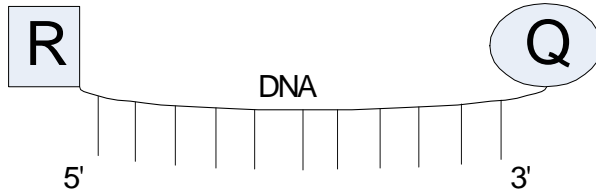


Fig. 1: Illustration of the structure of a TaqMan DNA probe. Here, R and Q denote the reporter and quencher fluorophores, respectively.

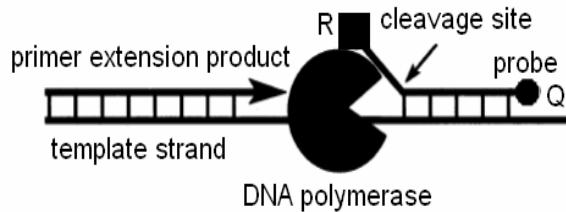


Fig 2: Degradation of a TaqMan probe, via cleavage by DNA polymerase.

backbone to TAMRA. As TAMRA emits this absorbed energy at a significantly longer wavelength (580 nm), the resulting fluorescence is not observable in Channel 1 of real-time PCR instruments^[4].

The combination of dual-labeled TaqMan DNA probes with forward and reverse primers is a must for a successful real-time PCR. As PCR is a repeated cycle of three steps (denaturation, annealing, and polymerization), a TaqMan DNA probe will anneal to a site within the DNA template in between the forward and reverse primers during the annealing step, if a subsequence of the DNA template is complementary to the sequence of the DNA probe. During polymerization, *Thermus Aquaticus* (*Taq*) DNA polymerase will extend the primers in a 5'-3' direction. At the same time, the *Taq* polymerase also acts as a "scissor" to degrade the probe via cleavage, thus separating the reporter from the quencher, as shown in (Fig. 2)^[5], where R and Q denote the reporter and quencher dyes, respectively. This separation subsequently allows the reporter to emit its fluorescence^[6]. This process occurs in every PCR cycle and does not interfere with the exponential accumulation of PCR product. As a result of PCR, the amount of DNA template increases exponentially, which is accompanied by a proportionate increase in the overall fluorescence intensity emitted by the reporter

group of the excised TaqMan probes. Hence, the intensity of the measured fluorescence at the end of each PCR polymerization is correlated to the total amount of PCR product, which can then be detected, using a real-time PCR instrument for visualization.

The DNA sequence design problem arises in molecular self-assembly applications and DNA computing^[7,8], as well as in general purpose laboratory applications such as probe selection for DNA microarrays and primer design for PCR^[9,10]. A procedure for DNA sequence design for DNA computing with output visualization based on real-time PCR has been proposed in^[11]. The procedure utilized DNASequencesGenerator^[12], a graph-based approach for designing a set of good DNA sequences. In graph-based approach, nodes in the graph represent base strands and a node has four strands that can appear as successors in a longer sequences as its child nodes. Then, by travelling the graph from root to leaf, DNA sequences can be designed. This approach also can find a set of orthogonal DNA sequences within a predefined error rate quickly. Some modified rules were used for designing probes and primers, based on DNASequencesGenerator, for real-time PCR implementations^[11]. As those modified rules are not available in DNASequencesGenerator, a major drawback of this procedure is the manual filtering, which is required for obtaining the desired DNA sequences. Hence, all the calculations have to be carried out manually, which is a time consuming process.

In this study, we propose different procedure based on generate-and-test approach to obtain all the DNA sequences based on user requirements. By considering adding specific constraints for real-time PCR implementation, a good set of DNA sequences can be generated automatically. Generate-and-test algorithm approach has been used recently to design DNA sequences. For example, Pechovsky and Ackermann designed DNA sequences by a random search algorithm^[13]. They encoded binary information in DNA strands and demonstrated twelve-bit DNA library. Tanaka *et al.*^[14] have developed the method of encoding design based on DNA free energy, and used random generate-and-test to filter out the good DNA sequences.

Output Visualization of DNA Computation based on Real-Time PCR: A readout method tailored specifically to HPP in DNA computing based on real-time PCR has been proposed, which employs a hybrid *in vitro-in silico* approach^[11,15]. In the *in vitro* phase, $O(|V|^2)$ TaqMan-based real-time PCR reactions are

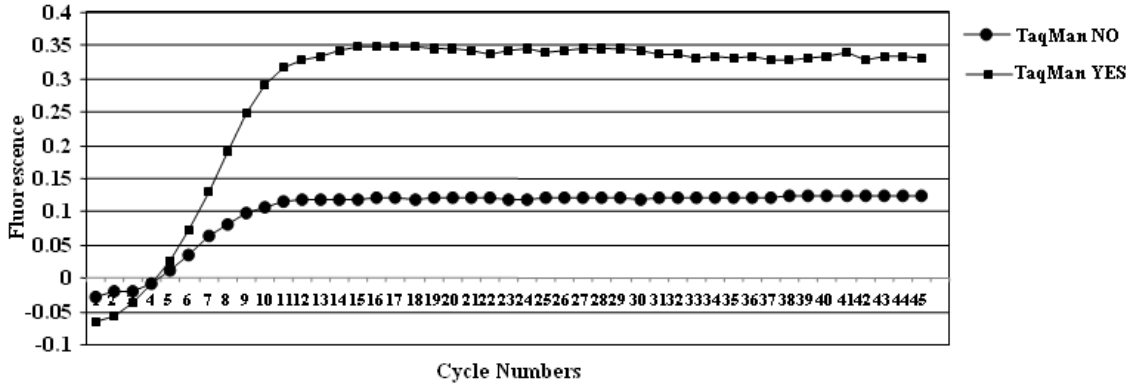


Fig. 3: An example of amplification plots corresponding to $TaqMan(v_0, v_k, v_l) = YES$ (first condition) and $TaqMan(v_0, v_k, v_l) = NO$ (second condition).

performed in parallel, to investigate the ordering of pairs of nodes in the Hamiltonian path of a $|V|$ -node instance graph, in terms of relative distance from the DNA sequence encoding the known start node. The resulting relative orderings are then processed *in silico*, which efficiently returns the complete Hamiltonian path. Previously, graduated PCR, which was originally demonstrated by Adleman^[16], was employed to perform such operations.

Notation and Basic Principle: First of all, $v_{1(a)}v_{2(b)}v_{3(c)}v_{4(d)}$ denotes a double-stranded DNA (dsDNA), which contains the base-pairs sub sequences, v_1, v_2, v_3 , and v_4 , respectively. Here, the subscripts in parenthesis (a,b,c,d) indicate the length of each respective base-pair subsequence. For instance, $v_{1(20)}$ indicates that the length of the double-stranded subsequence, v_1 is 20 base-pairs (bp). When convenient, a dsDNA may also be represented without indicating segment lengths (e.g., $v_1v_2v_3v_4$). A reaction denoted by $TaqMan(v_0, v_k, v_l)$ indicates that real-time PCR is performed using forward primer v_0 , reverse primer v_l , and TaqMan probe v_k . Based on the proposed approach, there are two possible reaction conditions regarding the relative locations of the TaqMan probe and reverse primer. In particular, the first condition occurs when the TaqMan probe specifically hybridizes to the template, between the forward and reverse primers, while the second occurs when the reverse primer hybridizes between the forward primer and the TaqMan probe. As shown in (Fig. 3), these two conditions would result in different amplification patterns during real-time PCR, given the same DNA template (i.e., assuming that they occurred separately, in two different PCR reactions). The higher fluorescent output of the first condition is a

typical amplification plot for real-time PCR. In contrast, the relatively lower fluorescent output of the second condition, which reflects the cleavage of a lower number of TaqMan probes via DNA polymerase due to the ‘unfavourable’ hybridization position of the reverse primer, is due to linear rather than exponential amplification of the template. Thus, $TaqMan(v_0, v_k, v_l) = YES$ if an amplification plot similar to the first condition is observed, while $TaqMan(v_0, v_k, v_l) = NO$ if an amplification plot similar to the second condition is observed.

The Readout Approach: Let the output of an *in vitro* computation of an HPP instance of the input graph be represented by a 120-bp dsDNA $v_{0(20)}v_{2(20)}v_{4(20)}v_{1(20)}v_{3(20)}v_{5(20)}$, where the Hamiltonian path $V_0 \rightarrow V_2 \rightarrow V_4 \rightarrow V_1 \rightarrow V_3 \rightarrow V_5$, begins at node V_0 , ends at node V_5 , and contains intermediate nodes V_2, V_4, V_1 , and V_3 , respectively. Note that in practice, only the identities of the starting and ending nodes, and the presence of all intermediate nodes will be known in advance to characterize a solving path. The specific order of the intermediate nodes within such a path is unknown.

The first part of the approach, which is performed *in vitro*, consists of $\frac{(|V|-2)^2 - (|V|-2)}{2}$ real-time PCR reactions, each denoted by $TaqMan(v_0, v_k, v_l)$ for all k and l , such that $0 < k < |V|-2$, $1 < l < |V|-1$, and $k < l$. For this example instance, so that the DNA template is dsDNA $v_0v_2v_4v_1v_3v_5$, these 6 reactions, along with the output in terms of “YES” or “NO” are as follows:

- $TaqMan(v_0, v_1, v_2) = NO$ (1)
- $TaqMan(v_0, v_1, v_3) = YES$ (2)
- $TaqMan(v_0, v_1, v_4) = NO$ (3)

TaqMan(v_0, v_2, v_3) = YES (4)

TaqMan(v_0, v_2, v_4) = YES (5)

TaqMan(v_0, v_3, v_4) = NO (6)

Note that the overall process consists of a set of parallel real-time PCR reactions, and thus requires $O(1)$ laboratory steps for *in vitro* amplification. The accompanying SPACE complexity, in terms of the required number of capillary tubes is $O(|V|^2)$. Clearly, only one forward primer is required for all real-time PCR reactions, while the number of reverse primers and TaqMan probes required with respect to the size of input graph are each $|V|-3$.

After all real-time PCR reactions are completed, the *in vitro* output is subjected to an algorithm for *in silico* information processing, producing the satisfying Hamiltonian path of the HPP instance in $O(n^2)$ TIME (here, n denotes vertex number) as follows:

```

Input: N[0...|V|-1]=2 // N[0, ?, ?, ?, 5]
A[1...|V|-2]=|V| // A[1, 1, 1, 1]
  for k=1 to |V|-3
    for l=k+1 to |V|-2
      if TaqMan( $v_0, v_k, v_l$ ) = YES
        A[l] = A[l]+1
      else A[k] = A[k]+1
      endif
    endfor
    N[ A[k] ] = k
  endfor
  N[ A[ |V|-2 ] ] = |V|-2

```

In this algorithm, an array (N[0...|V|-1]) that store all the nodes of the Hamiltonian path is defined. In addition, an array of aggregation values (A[1..|V|-2]) that is used to locate the Hamiltonian path in each array of nodes is also defined. Based on the modified algorithm, the input array N is first initialized to $N=\{0,?,?,?,5\}$ since the start and the end of the path are known, in advance. Next, the aggregation array A is initialized to $A=\{1,1,1,1\}$. During the loop operations of the algorithm, the value of in the array A is increased in each iteration steps. The aggregation array A[i] is used for indexing the nodes array for each value of k . After the loop operation $|V|-2$ is assigned to the $N[A[|V|-2]]$. The output of the *in silico* algorithm can be viewed by calling back all the nodes array N[0] to N[|V|-1].

Experiment: The experiment of output visualization of DNA computing based on real-time PCR includes two phases. The first phase is the preparation of input

molecules and the second phase is real-time PCR experiment, which 6 different TaqMan reactions were performed simultaneously on the real-time PCR machine.

A pool of 120-bp input molecules $v_{0(20)}v_{2(20)}v_{4(20)}v_{1(20)}v_{3(20)}v_{5(20)}$ is prepared, via standard protocol of parallel overlap assembly (POA) of single-stranded DNA strands (ssDNAs). For this purpose, 11 ssDNAs are required, including additional ssDNAs, which act as link sequences for self-assembly. These strands shown in Table 1, which were generated by DNASequencesGenerator^[12]. After completion, amplification via PCR was performed using the same protocol as POA. The PCR product was subjected to gel electrophoresis and the resultant gel image was captured, as shown in (Fig. 4). The 120-bp band in lane 2 shows that the input molecules have been successfully generated. Afterwards, the DNA of interest is extracted. The extracted DNA was used as a template in real-time PCR experiment. The real-time PCR reaction involves primers (Prologo, Japan), TaqMan probes (Prologo, Japan), and LightCycler TaqMan Master (Roche Applied Science, Germany). The sequences for forward primers, reverse primers, and TaqMan probes are shown in Table 2 and 3, basically derive from the generated sequences from Table 1. The resulting real-time PCR amplification plots are shown in (Fig. 5).

The DNA Sequence Design: In order to design a good sequence for each subsequence v_0, v_1, v_2, v_3, v_4 , and v_5 , five rules have been considered for designing the DNA sequences for primers and probes of real-time PCR according to^[17], which is based on the implementation of Primer Express® software (Applied Biosystems)^[18]. Those rules are as follows:

- Melting temperature for primers should be between 58-60°C and melting temperature for probes should be 10°C higher
- Primers should be 15-30 bases in length
- GC content of primers and probes should ideally be 30-80%
- For primers: The run of identical nucleotides should be avoided. This is especially true for G, where runs of 4 or more Gs are not allowed. Further, the total number of G and C in the last five nucleotides at the 3' end of the primer should not exceed 2
- For probes, there should be more C than G, and not a G at the 5' end

Table 1: The generated DNA sequences based on DNASequencesGenerator.

Name	DNA sequences (5'-3')
v ₀	CCTTAGTAGTCATCCAGACC
v ₂	CGCGCACCTTCTTAATCTAC
v ₄	CCACTGGTCTGCATGTAAC
v ₁	ATGCGCCAGCTTCTAACTAC
v ₃	TGGACAACCGCAGTTACTAC
v ₅	TCCACGCTGCACTGTAATAC
v ₀ v ₂	GTAGATTAAGGGTCTGGATGACTACTAAGG
v ₂ v ₄	GTAGTAACTGAAGGTGCGCG
v ₄ v ₁	GTATTACAGTCGGTTGTCCA
v ₁ v ₃	GTAGTTAGAAGCAGCGTGGA
v ₃ v ₅	GTTACATGCAGAACCAGTGGGCTGGCGCAT

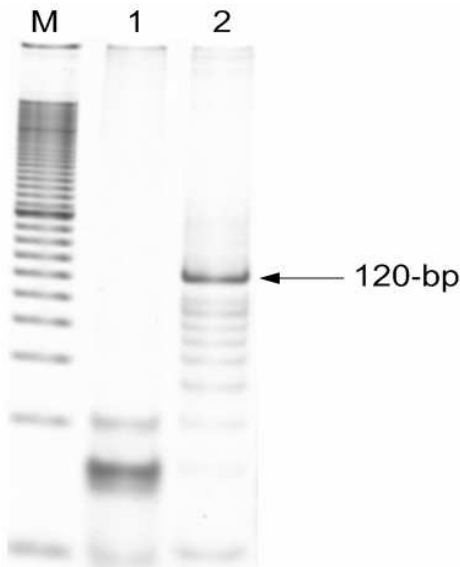


Fig. 4: Gel image for the preparation of input molecules. Lane M denotes a 20-bp molecular marker, lane 1 is the product of initial pool generation based on parallel overlap assembly, and lane 2 is the amplified PCR product.

Table 2: Sequences for forward and reverse primers.

Name	DNA Sequences (5'-3')
Forward primer, v ₀	CCTTAGTAGTCATCCAGACC
Reverse primer, \bar{v}_2	GTAGTTAGAAGCTGGCGCAT
Reverse primer, \bar{v}_3	GTAGTAACTGCGGTTGTCCA
Reverse primer, \bar{v}_4	GTATTACAGTCAGCGTGGA

Table 3: Sequences for TaqMan dual-labeled probes.

DNA probe	Sequences (5'-3')
Probe v ₁	R-ATGCGCCAGCTTCTAACTAC-Q
Probe v ₂	R-CGCGCACCTTCTTAATCTAC-Q
Probe v ₃	R-TGGACAACCGCAGTTACTAC-Q

In this research a DNA sequences design algorithm based on generate-and-test approach is applied to obtain a set of good DNA sequences. The generate-and-test algorithm consists of two main levels. The first level considers the basic constraints of DNA sequence design followed by the second level that includes specific constraints formulated based on 5 rules, which has been mentioned previously.

Basic Constraints: Normally, a set of sequences for DNA computing is designed subjected to six constraints^[19]. Those constraints are GC-percentage, melting temperature, continuity, hairpin, H-measure, and similarity.

GC-percentage is expressed as the percentage of nucleotides that are G and C bases in a strand or duplex. Since GC base pairs are held together by three hydrogen bonds while AT base pairs are held together by only two hydrogen bonds, double-stranded DNA with a high GC content is often more stable than DNA with a high AT content. The GC-percentage is simple to calculate; only the length and the number of GC bases are needed, where the length refers to the number of nucleotide base pairs. GC percentage formula is given by the Eq. 1:

$$GC\% = \frac{\text{(Number of G and C)}}{\text{(Length of DNA sequence)}} \times 100\% \quad (1)$$

Melting temperature, T_m , is the temperature at which 50% of the DNA strands are in denatured form and 50% are in double helical form. The simplest model to estimate the melting temperature of DNA duplexes simply counts the base pairs. This model, which is known as the Wallace 2-4 rule^[20], is formulated as follows:

$$T_m = 2(\text{number of AT pairs}) + 4(\text{number of GC pairs}) \quad (2)$$

An improved method to calculate melting temperature is given by the Eq.(3), which includes the consideration of the salt concentration^[21]:

$$T_m = 2(A+T) + 4(G+C) - 16.6 \log_{10}(0.050) + 16.6 \log_{10}([Na+]) \quad (3)$$

where A, T, C, and G represent the number of each bases in a DNA sequence and [Na+] is the concentration of sodium ions in a DNA solution.

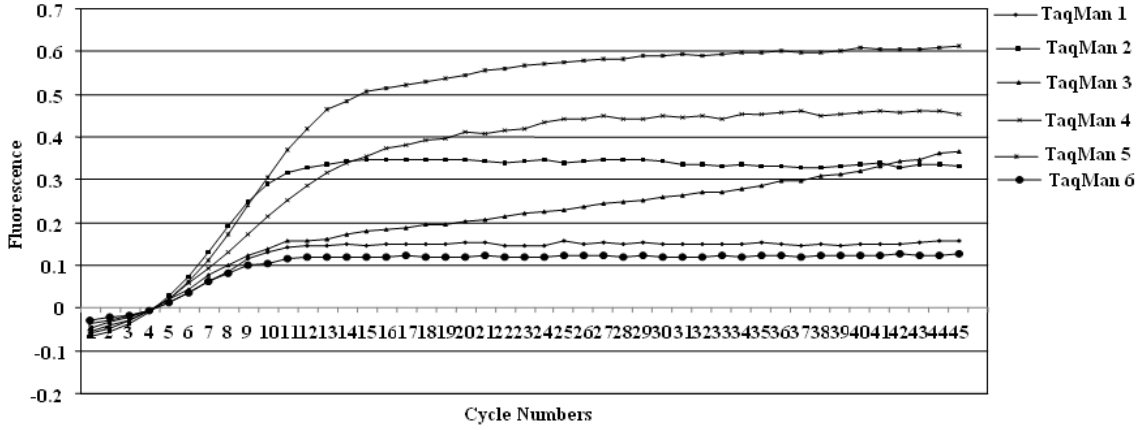


Fig. 5: Output of real-time PCR. Reactions 1 to 6 indicate the $\frac{(|V|-2)^2-(|V|-2)}{2}$ TaqMan(v_0, v_k, v_i) reactions of the input instance.

The most accurate thermodynamic model for predicting T_m is the nearest neighbor method that is based on experimentally derived values for the changes in enthalpy (ΔH) and entropy (ΔS) during the formation of a DNA duplex with a specific sequence. The model is formulated as:

$$T_m = \frac{\Delta H}{\Delta S + R \ln\left(\frac{C_t}{4}\right)} \quad (4)$$

where:

R = The gas constant [$1.987 \text{ cal Kmol}^{-1}$]
 C_t = The total strand concentration for non-symmetry DNA strand.

For self-complementary molecules, $C_t/4$ is replaced by C_t . ΔH and ΔS are calculated by summing the nearest-neighbor enthalpy and entropy changes for the entire hybrid. The values of the enthalpies and entropies are evaluated experimentally by^[22-25]. The Breslauer parameter is used, which the nearest-neighbor method is used for calculating T_m in this research.

Basic definitions: The similar basic definition in^[26] is used to formulate the hairpin, continuity, H-measure, similarity, and the other additional constraints. An alphabet consists of each single nucleotide and the gap can be defined as $\Lambda = \{A, C, G, T, -\}$, where $\{-\}$ denotes as a gap. Alphabet that consists of only single nucleotide can be defined as $\Lambda_{nb} = \{A, C, G, T, \}$. Then, the set of all DNA sequences is denoted as Λ_{nb}^* . Let $a, b \in \Lambda$ and $x, y \in \Lambda^*$. The length of sequence x is

denoted as $|x|$, and $x_i (1 \leq i \leq |x|)$ means i th nucleotide from 5'-end of sequence x . A set of sequences with the same length l is denoted by Σ , where i th member of Σ is denoted as Σ_i . \bar{a} is the complementary of a , and other basic definitions are as follows:

$$bp(a, b) = \begin{cases} 1, & a = \bar{b} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

$$eq(a, b) = \begin{cases} 1, & a = b \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

$$T(i, j) = \begin{cases} i, & i > j \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

For a given sequence $x \in \Lambda^*$, the number of nonblank nucleotides is defined as:

$$\text{length}_{nb}(x) = \sum_{i=1}^{|x|} nb(x_i) \quad (8)$$

where:

$$nb(a) = \begin{cases} 1, & a \in \Lambda_{nb} \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

and a shift of a sequence x by i bases is denoted as follows:

$$\text{shift}(x, i) = \begin{cases} (-)^i x_1 \cdots x_{1-i}, & i \geq 0 \\ x_{i+1} \cdots x_i (-)^i, & i < 0 \end{cases} \quad (10)$$

Furthermore, a single DNA sequence can be reversed by using this formulation:

$$\text{rev}(x) = \sum_{i=1}^{\lfloor l/2 \rfloor} \text{swap}(x_i, x_{l+1-i}) \quad (11)$$

H-measure: The H-measure computes how many nucleotides are complementary to prevent cross-hybridization of two sequences including position shift using the formulation as follows:

$$H(x, y) = \text{Max}_{-l < i < l} \left(\sum_{j=1}^l \text{bp}(x_j, \text{shift}(\text{rev}(y), i)_j) \right) \quad (12)$$

H-measure for the given set of sequences Σ is defined as follows:

$$f_{\text{H-measure}}(\Sigma_i) = \sum_{j, j \neq i}^n H(\Sigma_i, \Sigma_j) \quad (13)$$

Similarity: The similarity measure computes the similarity in the same direction of two given sequences to keep each sequence as unique as possible including position shift. The formulation for similarity uses the same approach as H-measure:

$$\text{Sim}(x, y) = \text{Max}_{-l < i < l} \left(\sum_{j=1}^l \text{eq}(x_j, \text{shift}(y, i)_j) \right) \quad (14)$$

Similarity for the given set of sequences Σ is defined as follows:

$$f_{\text{Similarity}}(\Sigma_i) = \sum_{j, j \neq i}^n \text{Sim}(\Sigma_i, \Sigma_j) \quad (15)$$

Hairpin: In^[23], the hairpin measure, as formulated in Eq.(16), calculates the probability of a single-stranded DNA, to form a secondary structure:

$$\text{hp}(x) = \sum_{r=R_{\min}}^{l-2*P_{\min}} \sum_{p=P_{\min}}^{l-P_{\min}-r} T \left(\sum_{i=1}^{\text{pinlen}(p,r)} \text{bp}(x_{p+l-i}, x_{p+r+i}) \right) \frac{\text{pinlen}(p,r)}{2} \quad (16)$$

where:

$$\text{pinlen}(p, r) = \min(p, l-p-r) \quad (17)$$

Continuity: If the same bases occur continuously in a sequence, the sequence could form an unexpected structure. This can be measured by continuity, which is defined as follows:

$$\text{Con}(x) = \text{Max}_{1 \leq i \leq l} \left(\sum_{a \in \Lambda_{ab}} c(a, i) \right) \quad (18)$$

where:

$$c(a, i) = \begin{cases} n, & \text{if } \exists n, \text{ s.t. } \text{eq}(a_i, a_{i+j}) = 1, \\ & \text{for } 1 \leq j < n, \text{ eq}(a_i, a_{i+n}) = 0 \\ 0, & \text{otherwise} \end{cases} \quad (19)$$

Specific or Additional Constraints: Based on the design rules for DNA sequences, probes, and primers, of the real-time PCR implementation^[11], three constraints for DNA sequence design are defined as follows:

- LAST5GC3 is the total number of G and C in the last five nucleotides at the 3'end of the DNA sequence.
- CMINUSG is the total number of C minus total number of G.
- G5END is an indicator whether there is a G at 5'end of the DNA sequence. G5END = 0 for no G and G5END = 1 if G exists at the 5'end.

Generate-and-Test Algorithms: A generate-and-test approach is implemented to generate a set of DNA sequences subjected to basic and specific/additional constraints. The flowchart of the generate-and-test approach is shown in Fig. 6.

Basically, all the constraints can be divided into two types. The first type of constraints only consider single DNA sequence for the computation of the constraint, which are hairpin, continuity, GC-percentage, melting temperature, and all the additional constraints. The second type of the constraints considers two DNA sequences for the calculation of the constraint, which are H-measure and similarity.

The algorithm starts by initializing all the constraint values based on user requirements. The first sequence is generated randomly. This sequence has to satisfy melting temperature, GC-percentage, continuity, hairpin, and all the additional constraints before it is stored in a DNA set for the next iterations. After the first sequence is obtained, the second sequence is randomly generated and the sequence must satisfy all the first type of constraints. This sequence however

must satisfy H-measure and similarity constraints, calculated with the first sequence before the sequence is stored in the archive set. The next sequences also must satisfy all the first type constraints and has to be compared with all sequences in the set by calculating H-measure and similarity. The iteration continues until the amount of DNA sequences required is obtained.

RESULT

The sequence generated by generate-and-test algorithm is performed on the Visual C++ 6.0, using a computer with 3.0 GHz Intel P4 processor and 2 GB RAM. The results obtained are compared with the sequence generated in^[11], where the DNASequenceGenerator was employed^[12]. All the variables for the generation of six DNA sequences based on HPP are listed in Table 4. The comparison is shown in Table 5 and Fig. 7. For the H-measure, similarity, LAST5GC3, and CMINUSG, the maximum value is allowed for calculating the constraints. For example, if the H-measure value is set to 8, this means that H-measure for DNA sequences cannot exceed 8. This approach is applied for similarity, LAST5GC3, and CMINUSG.

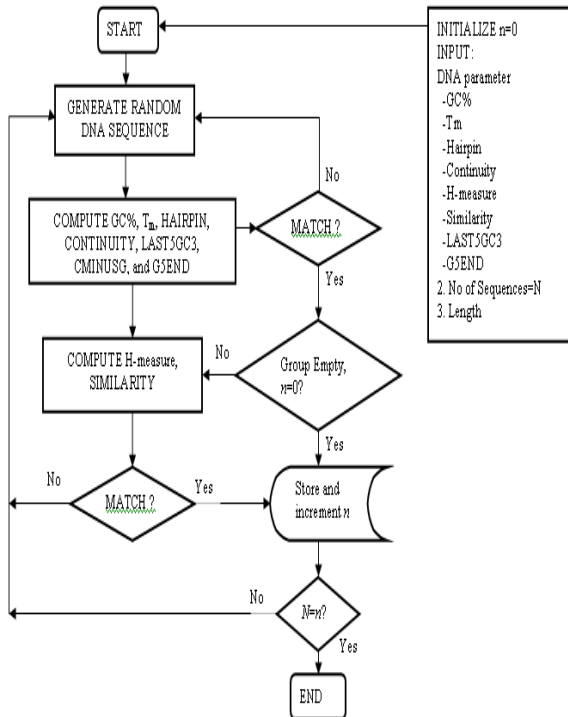


Fig. 6: The flowchart of the proposed generate-and-test approach.

DISCUSSION

Sequences designed by generate-and-test algorithm have higher H-measure average value than sequences generated by DNASequenceGenerator. However, our sequences show lower similarity average value than sequences designed by DNASequenceGenerator. The previous sequences designed by DNASequenceGenerator have average value of hairpin equals to zero and average value of continuity equals to two. Based on the results, our sequences have lower average value than sequences produced by DNASequenceGenerator for additional constraints.

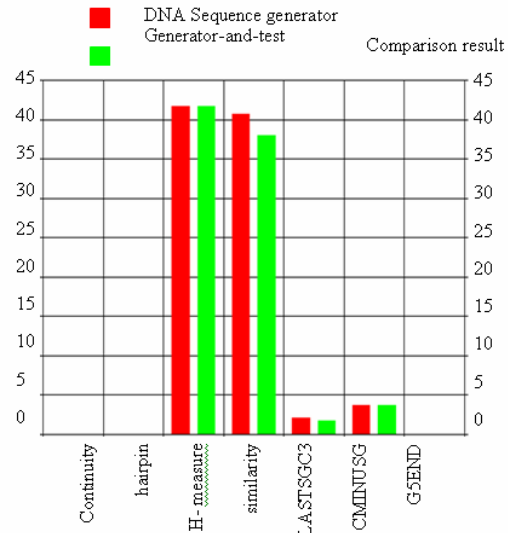


Fig. 7: Comparison results of^[11] and the proposed generate-and-test algorithm.

Table 4: Parameters setting for generation of DNA sequences for 6 nodes HPP.

Constraints (allowable range)	Value
Number of sequences (1-10)	6
Length of sequences (8-20)	20
Continuity (1-3)	1
Hairpin	0
H-measure	8
Similarity	8
LAST5GC3	2
CMINUSG	2
GSEND	0
GC percentage	50%
Melting Temperature	
Method	Breslauer
DNA concentration	200nM
Salt concentrations	0.05 M
Forward primer	50-55°C
Reverse primer	58-60°C

algorithm.

Table 5: Comparison results of the sequences in^[11] and the sequences based on the generate-and-test algorithm.

Sequence (5'-3')	Continuity	Hairpin	H-measure	Similarity	Tm	GC%	LAST5GC3	CMINUSG	G5END
DNASequenceGenerator:									
CCTTAGTAGTCATCCAGACC	2	0	42	36	53.5	50.0	3	4	0
CGCGCACCTTCTTAATCTAC	2	0	39	46	59.7	50.0	2	6	0
CCACTGGTTCTGCATGTAAC	2	0	43	42	59.4	50.0	2	2	0
ATGCGCCAGCTTCTAACTAC	2	0	40	41	59.1	50.0	2	4	0
TGGACAACCGCAGTTACTAC	2	0	43	40	58.9	50.0	2	2	0
TCCACGCTGCACTGTAATAC	2	0	43	39	59.7	50.0	1	4	0
Generate-and-Test:									
AATCCTCTACCGAGCTAACG	2	0	42	39	53.7	50.0	2	4	0
AACCTCCGCCTCTGTTCAT	2	0	41	39	58.4	50.0	1	6	0
TCAACTACCGCCAACGTCAT	2	0	42	36	58.8	50.0	2	6	0
AGGCCGAATCCAGATCTCAA	2	0	41	39	59.4	50.0	2	2	0
CCGTGTTAACCGGTCCAATT	2	0	41	37	59.3	50.0	1	2	0
CGGCCTCAACATATGGA	2	0	43	38	58.4	50.0	2	2	0

CONCLUSION

These studies showed that the generate-and-test algorithm can be used to design a set DNA sequences for DNA computing readout method based on real-time PCR. Sequences generated based on this algorithm are comparable to sequences designed by DNASequenceGenerator. By using 6 basic constraints for DNA sequence design with additional constraints proposed in this research, a set of DNA sequences generated by generate-and-test can be used in DNA computing with output visualization based on real-time PCR. Generated-and-test algorithm is easy to apply and suitable for designing DNA sequences that have many constraints, where no optimization process is needed, to generate a set of good DNA sequences. However, the DNA sequences generated by generate-and-test algorithm are not the optimized solution, because the resultant sequences generated by generate-and-test algorithm are based on threshold values set by user. Hence, for the future study, advanced optimization algorithms, such as Ant Colony Optimization and Particle Swarm Optimization would provide a lot of advantages in designing good DNA sequences.

ACKNOWLEDGEMENT

This research is supported financially by the Ministry of Science, Technology, and Innovation (MOSTI), Malaysia under eScienceFund Research Funding (Vot 79034) and Ministry of Education under Fundamental Research Grant Scheme (FRGS) (Vot 78225). The researchers are indebted to MOSTI, MOHE, and University Technology Malaysia for granting them a financial support and opportunity to do this research.

REFERENCES

1. Mullis, K., 1986. Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction. Cold spring Harb. Symp. Quant. Biol., 51: 263-273. <http://www.ncbi.nlm.nih.gov/pubmed/3472723>
2. Overbergh, L., 2003. The use of real-time reverse transcriptase PCR for the quantification of cytokine gene expression. J. Biomolecular Tech., 14: 33-43. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2279895>
3. Walker, N.J., 2002. A technique whose time has come. Science, 296: 557-559. DOI: 10.1126/science.296.5567.557
4. Bubner, B. and I.T. Baldwin, 2004. Use of real-time PCR for determining copy number and zygosity in transgenic plants. Plant Cell Reports, 23: 263-271. <http://www.ncbi.nlm.nih.gov/pubmed/15368076>
5. Heid, C.A., 1996. real-time quantitative PCR. Genome Res., 6: 986-994. DOI: 10.1101/gr.6.10.986
6. Holland, P.M., 1991. Detection of specific polymerase chain reaction product by utilizing the 5'→3' exonuclease activity of *termus aquaticus* dna polymerase. Proceedings of the National Academy of Sciences of the United States of America, Washington, DC., pp: 7276-7280. <http://cat.inist.fr/?aModele=afficheN&cpsidt=4975395>.
7. Chen, J. and N.C. Seeman, 1991. Synthesis from DNA of a molecule with the connectivity of a cube. Nature, 350: 631-633. DOI: 10.1038/350631a0
8. Deaton, R., R.C. Murphy, M. Garzon, D.T. Franceschetti and Jr. S.E. Stevens, 1996. Good encodings for DNA-based solutions to combinatorial problems. Proceedings of the 2nd Annual Meeting on DNA Based Computers, June 10-12, American Mathematical Society, Princeton University, pp: 159-171. <http://www.csce.uark.edu/~rdeaton/dna/papers/mcggood.pdf>

9. Ben-Dor, A., R. Karp, B. Schwikowski and Z. Yakhini, 2000. Universal DNA tag systems: A combinatorial design scheme. *J. Comput. Biol.*, 7: 503-519. DOI: 10.1089/106652700750050916
10. Li, F. and G.D. Stormo, 2001, Selection of optimal DNA oligos for gene expression arrays. *Bioinformatics*, 17: 1067-1076. <http://bioinformatics.oxfordjournals.org/cgi/content/abstract/17/11/1067>
11. Ibrahim, Z., T.B. Kurniawan and M. Khalid, 2007. A DNA sequence design for molecular computation of HPP with output visualization based on Real-Time PCR. *IEEE Congress on Evolutionary Computation*, Sept. 25-28, IEEE Xplore, Singapore, pp: 1823-1828. DOI: 10.1109/CEC.2007.4424694
12. Feldkamp, U., S. Saghafi, W. Banzhaf and H. Rauhe, 2001. DNASequencesGenerator: A program for the construction of DNA sequences. *Proceedings of the 7th International Workshop DNA Based Computers*, June 10-13, Tampa, USA., pp: 179-188.
13. Pechovsky, R. and J. Ackerman, 2003. DNA library design for molecular computation. *J. Comput. Bio.*, 10: 215-229. DOI: 10.1089/106652703321825973
14. Tanaka, F., A. Kameda, M. Yamamoto and A. Ohuchi, 2005. Design of nucleic acid sequences for DNA computing based on a thermodynamic approach. *Nucleic Acids Res.*, 33: 903-911. DOI: 10.1093/nar/gki235
15. Ibrahim, Z., Y. Tsuboi, J.A. Rose, O. Ono and M. Khalid, 2006. A New Readout Approach in DNA computing based on real-time PCR with TaqMan probes. *Lecture Notes Comput. Sci.*, 4287: 350-359. DOI: 10.1007/11925903_27
16. Adleman, L., 1994. Molecular computation of solutions to combinatorial problems. *Science*, 266: 1021-1024 <http://www.ncbi.nlm.nih.gov/pubmed/7973651>
17. Tevfik, M., 2002. Real Time PCR. <http://dorakmt.tripod.com/genetics/realttime.html>.
18. Tools, Tutorials, Maintenance, and Troubleshooting 2009. *Applied Biosystems*, <http://www.appliedbiosystems.com/support/apptech/>
19. Shin, S.Y., I.H. Lee, D. Kim and B.T. Zhang, 2005. Multiobjective evolutionary optimization of DNA sequences for reliable DNA computing. *IEEE Trans. Evolut. Comput.*, 9: 143-158. DOI: 10.1109/TEVC.2005.844166
20. Wallace R.B., J. Shaffer, R.F. Murphy, J. Bonner, T. Hirose and K. Itakura, 1979. Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: The effect of single base pair mismatch. *Nucleic Acids Res.*, 6: 3543-3558. <http://nar.oxfordjournals.org/cgi/content/abstract/6/11/3543>
21. Wetmur, J.G., 1991, DNA probes: Applications of the principles of nucleic acid hybridization. *Critical Rev. Biochem. Molecular Bio.*, 26: 227-259. DOI: 10.3109/10409239109114069
22. Breslauer, K.J., R. Frank, H. Blocker and L.A. Marky, 1986. Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. USA.*, 83: 3746-3750. <http://www.pnas.org/content/83/11/3746.full.pdf>
23. Lucia Jr, J.S. H.T. Allawi and P.A. Seneviratne, 1996. Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, 35: 3555-3562. DOI: 10.1021/bi951907q
24. Lucia, Jr. JS. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceeding of National Academy of Science*, Feb. 17-17, USA., pp: 1460-1465. <http://www.pnas.org/content/95/4/1460.abstract>
25. Sugimoto, N., S. Nakano, M. Yoneyama, K. Honda, 1996. Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nucleic Acid Res.*, 24: 4501-4505. <http://nar.oxfordjournals.org/cgi/content/abstract/24/22/4501>
26. Garzon M., P. Neathery, R. Deaton, R.C. Murphy, D.R. Franceschett, S.E. Stevens Jr, 1997. A new metric for DNA computing. *Proc. Genetic Program*, 472-478. <http://www.csce.uark.edu/~rdeaton/dna/papers/gp97code.pdf>