AIChE

Oligonucleotide Optimization for DNA Synthesis

Tobias M. Louw, Scott E. Whitney, Joel R. TerMaat, Elsje Pienaar, and Hendrik J. Viljoen
Dept. of Chemical and Biomolecular Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588

DOI 10.1002/aic.12410
Published online September 29, 2010 in Wiley Online Library (wileyonlinelibrary.com).

Large DNA constructs can be synthesized from smaller oligonucleotides using the polymerase chain reaction. The set of oligonucleotides should be designed so that the melting temperature amongst oligonucleotide hybridization pairs do not vary greatly and the length of each oligonucleotide should not exceed 50 nucleotides. A near optimal oligonucleotide set is calculated using reliable gradient optimization methods. This was accomplished by defining a set of discrete arrays that is used to determine the melting temperature of a subset of the larger DNA sequence, depending on the subset start and end positions. These arrays were then incorporated into an objective function, which was optimized using the Broyden-Fletcher-Goldfarb-Shanno method. This method is adjusted slightly to incorporate explicit length and temperature constraints. Experimental results confirmed that the method performs better than similar software programs for the cases investigated and produces suitable oligonucleotide sets for DNA assembly. © 2010 American Institute of Chemical Engineers *AIChE J,* 57: 1912–1918, 2011
Keywords: DNA synthesis, bioengineering, optimization, melting temperature, oligonucleotide

Introduction

DNA molecules from 100 to 1000 bp in length can be de novo synthesized using the polymerase chain reaction (PCR). The DNA constructs are assembled from smaller oligonucleotides (single-stranded DNA molecules, typically with a length of up to 50 nucleotides), which overlap to form the complete DNA sequence. These are then extended by a polymerase enzyme. The cycle is repeated until the full-length product has been formed. This method has come to be known as polymerase chain assembly (PCA).[1]

The main difficulty in DNA synthesis is in designing the set of oligonucleotides so that certain criteria are met:

- the melting temperatures of each oligonucleotide hybridization pair must not vary greatly;
- the length of each oligonucleotide must not exceed 50 bp to reduce synthesis errors and associated purification costs; and
- oligonucleotides must bond to specific sites to prevent incorrect assembly—in other words, each oligonucleotide must be dissimilar from all others.

Many open-access software packages currently exist to optimize the oligonucleotide set to be used in the assembly of a larger DNA construct.[2–4] These software packages generally use combinatorial optimization techniques. Convergence is slow and there is no guarantee that the optimum will be reached. Most of these software packages do not incorporate the effects of monovalent salt and magnesium concentrations present in the process, further limiting the accuracy of predicted melting temperatures.

To make use of reliable gradient optimization methods, an objective function can be derived from the DNA sequence to be synthesized. The objective function presented here is calculated from a predetermined set of discrete arrays. Derivatives can be approximated using the central difference

Correspondence concerning this article should be addressed to H. J. Viljoen at hviljoen1@unlnotes.unl.edu.

operator. The objective function can be optimized using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method.[5]

Background

Sequence melting temperature

The melting temperature of DNA is usually understood to be the temperature at which 50% of the DNA molecules are single stranded and 50% are double stranded. The melting temperature can be estimated using thermodynamic data. This estimation involves using the nearest neighbor parameters[6] and several other correction factors for oligonucleotide, salt, and magnesium concentrations.[7–9] This method is used by the Integrated DNA Technologies website www.idtdna.com.[10,11] The melting temperature is calculated using Eqs. 1–5:

$$T_{\mathrm{m}} = \frac{\Delta H^0}{\mathrm{R}\ \ln(C/2) + \Delta S} \tag{1}$$

$$\Delta H^0 = \Delta H_{\mathrm{I}}^1 + \Sigma_{k=1}^{N-1} \Delta H_{\mathrm{NN}}^k + \Delta H_{\mathrm{I}}^N \tag{2}$$

$$\Delta G^0 = \Delta G_{\mathrm{I}}^1 + \Sigma_{k=1}^{N-1} \Delta G_{\mathrm{NN}}^k + \Delta G_{\mathrm{I}}^N + \Delta G_{\mathrm{sym}} \tag{3}$$

$$\Delta G = \Delta G^0 - 310.15K \times \Delta H^0 f_{\mathrm{corr}} \tag{4}$$

$$\Delta S = \frac{\Delta H^0 - \Delta G}{310.15\ K} \tag{5}$$

The symbols retain their standard thermodynamic interpretation. The subscript NN indicates "nearest neighbor" enthalpy and Gibbs free energy, whereas I indicates "initial." The superscript indicates the nucleotide position. The integer N is the sequence length (in nucleotides). f_{corr} is a correction factor to include the effects of magnesium, monovalent salts, and GC content on the Gibbs free energy.[7] C is the oligonucleotide concentration (of each oligonucleotide type, assuming they are in equal concentration) in moles per liter.

DNA sequence assembly from oligonucleotides

DNA sequences are assembled using PCA in much the same way as they are copied using the conventional PCR for template amplification. The process consists of three steps, namely denaturation, annealing, and elongation.[1] The temperature at which the annealing step takes place is critical, especially during PCA (see Figure 1). It is very important that the melting temperature of each overlap (A:B, B:C, and C:D) is not lower than the PCA protocol's annealing temperature to ensure that annealing takes place throughout the sequence.

Furthermore, it is important that the possibility of incorrect annealing (i.e., A annealing to D instead of C annealing to D) is prevented by ensuring that the sequences of A and C are very dissimilar. If this is the case, the temperature at which A will anneal to D will be much lower than the annealing temperature prescribed by the PCA protocol.

One way of preventing similarity in sequences is to avoid using large oligonucleotides. This decreases the chances of similar sequences occurring on different oligonucleotides that may cause incorrect annealing. Very large oligonucleotides (>50 nucleotides) are also much more expensive due to purification costs.

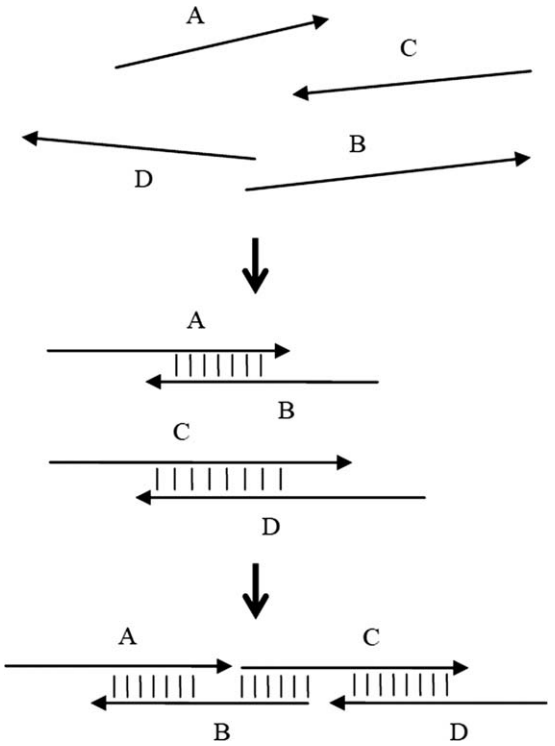

Figure 1. Oligonucleotides A, B, C, and D (with varying lengths) form a DNA sequence.

Each overlap A:B, B:C, and C:D has a unique melting temperature.

The BFGS method

The BFGS method is a well-known iterative optimization technique for continuous differentiable functions.[5] It is essentially a line search technique, defined by Eq. 6:

$$\mathrm{D}^{(k)}(\bar{\mathbf{x}}^{(k+1)} - \bar{\mathbf{x}}^{(k)}) = -t_k \nabla f(\bar{\mathbf{x}}^{(k)}) \tag{6}$$

The matrix $\mathrm{D}^{(k)}$ must be positive definite and updated with each iteration. This guarantees convergence if t_k is chosen according to a set of criteria known as Wolfe's conditions.

It is important to note that the BFGS method is an unconstrained optimization method. However, constraints can be implemented relatively simply. If the current vector $\bar{\mathbf{x}}^{(k)}$ satisfies any set of explicit constraints, then the next vector $\bar{\mathbf{x}}^{(k+1)}$ will satisfy the constraints as well if t_k is small enough. Each iteration will start by guessing a large value for t_k and decreasing this value until all conditions (Wolfe's conditions and explicit constraints) are met.

Calculations

Rewriting a sequence as a mathematical function

For a given DNA sequence, it is desirable to predict the melting temperature of a specific subset of the sequence. For instance, given the sequence shown in the first column of Table 1 and knowing parameters such as the salt and magnesium concentrations, one can determine the melting point of any smaller piece of the sequence, given the start and end

Table 1. Example Sequence and Sequence Subsets with Corresponding Melting Temperature in a Typical Solution

Sequence	x_1	x_2	Subset	T_m (°C)
5′-CCAGGGCAAGCTCTTCCGAGGCGCCAAGTCCTACACTTTT-3′	5	15	5′-GGCAAGCTCTT-3′	51
	20	38	5′-GGCGCCAAGTCCTACACTT-3′	68
	12	25	5′-TCTTCCGAGGCGCC-3′	65

points of each piece (x_1 and x_2). Therefore, once all "external" parameters are fixed, the melting temperature of a subset of a sequence is a function of only the start and end points of that sequence. The sequences in column four of Table 1 are subsets of the sequence in column one, starting at the position indicated by x_1 and ending at the position x_2.

Given any sequence, the melting temperature can be calculated from Eqs. 1–5. Let us first create an array named $\mathbf{H}_I(x)$ (7):

$$\mathbf{H}_I(x) = \begin{cases} 0.42\,\frac{\text{kJ}}{\text{mol}} & x \rightarrow \text{G/C} \\ 9.63\,\frac{\text{kJ}}{\text{mol}} & x \rightarrow \text{A/T} \end{cases} \tag{7}$$

The value of $\mathbf{H}_I(x)$ depends solely on whether the sequence has a G/C or A/T base pair at position x where $x \in \{\mathbb{N}; x \leq N\}$ and N is the total number of base pairs in the sequence.[6] This is shown in Figure 2 for the total sequence in Table 1. The same can be done to determine ΔG_I.

To determine the enthalpies contributed by the sets of nearest neighbors, we set up a similar array, named $\mathbf{H}_{NN}(x)$ (8):

$$\mathbf{H}_{NN}(x) = \begin{cases} 0 & x = 1 \\ \Sigma_{k=2}^{x} \Delta H_{NN}^{k} & x \geq 2 \end{cases} \tag{8}$$

Notice that $\mathbf{H}_{NN}(x)$ is the sum of all the nearest neighbor pairs up to x. We can therefore determine the nearest neighbor enthalpies between x_1 and x_2 by subtracting $\mathbf{H}_{NN}(x_1)$ from $\mathbf{H}_{NN}(x_2)$. Again, the same can be done to determine $\Sigma\Delta G_{NN}$. This is shown in Figure 3.

Look at the sequence shown in Table 1 as an example. The total enthalpy of the subset from position 5 to 15 (GGCAAGCTCTT) must be determined (9):

$$\begin{aligned} &\Delta H^0(x_1, x_2) = \Delta H_I^{x_1} + \Sigma_{k=x_1}^{x_2} \Delta H_{NN}^{k} + \Delta H_I^{x_2} \\ &\therefore \Delta H^0(5, 15) = \mathbf{H}_I(5) + \mathbf{H}_{NN}(15) - \mathbf{H}_{NN}(5) + \mathbf{H}_I(15) \\ &\therefore \Delta H^0(5, 15) = 0.42\,\frac{\text{kJ}}{\text{mol}} - 451.78\,\frac{\text{kJ}}{\text{mol}} + 101.74\,\frac{\text{kJ}}{\text{mol}} \\ &\qquad + 9.63\,\frac{\text{kJ}}{\text{mol}} = -339.99\,\frac{\text{kJ}}{\text{mol}} \end{aligned} \tag{9}$$

The Gibbs free energy can be calculated in exactly the same way.

As mentioned above, the salt correction function, f_{corr}, is a function of the salt concentration, magnesium concentration, GC content, and sequence length. The salt and magnesium concentrations are independent of the subset positions x_1 and x_2. The sequence length is $N_{BP} = x_2 - x_1 + 1$. The fractional GC content can easily be calculated in the exact same way as the nearest neighbor enthalpies. Let $\mathbf{GC}(x)$ be defined by Eq. 10:

$$\begin{aligned} \mathbf{GC}(x) &= \Sigma_{k=1}^{x}\ \text{gc}(k) \\ \text{gc}(k) &= \begin{cases} 1 & k \rightarrow \text{G/C} \\ 0 & k \rightarrow \text{A/T} \end{cases} \end{aligned} \tag{10}$$

$\mathbf{GC}(x)$ gives the total amount of G/C base pairs from 0 to x. The fractional GC content of a sequence can be determined by Eq. 11:

$$f_{GC}(x_1, x_2) = \frac{\mathbf{GC}(x_2) - \mathbf{GC}(x_1 - 1)}{x_2 - x_1 + 1} \tag{11}$$

Five arrays have been defined in terms of x: $\mathbf{H}_I(x)$; $\mathbf{H}_{NN}(x)$; $\mathbf{G}_I(x)$; $\mathbf{G}_{NN}(x)$; and $\mathbf{GC}(x)$. These five arrays can be substituted into Eqs. 1–5, 11 to determine the melting

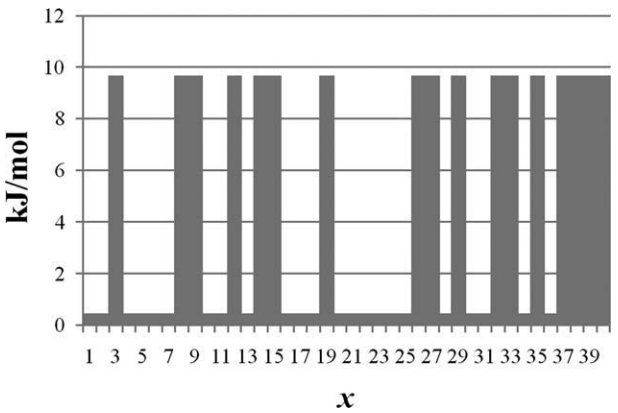

Figure 2. The initiation enthalpy is represented as $H_I(x)$.

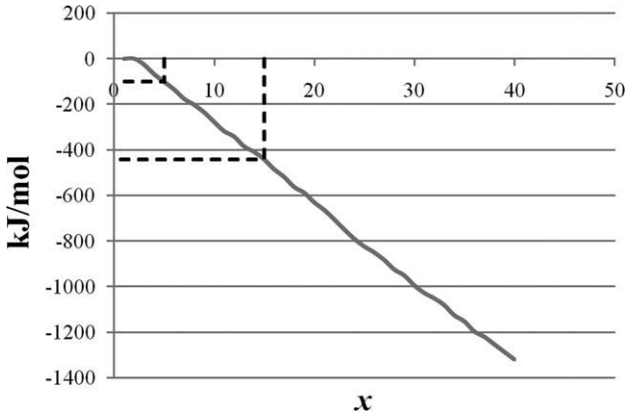

Figure 3. The nearest neighbor enthalpies are represented as $H_{NN}(x)$.

The dashed lines show that $H_{NN}(5) = -101.74$ kJ/mol and $H_{NN}(15) = -451.78$ kJ/mol.

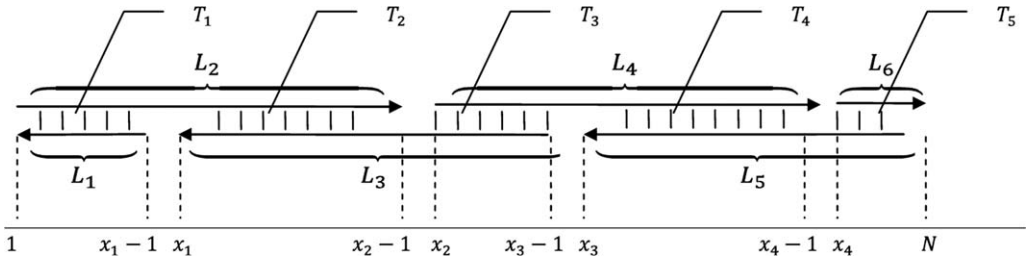

Figure 4. The numbering of overlap temperatures and oligonucleotide lengths are shown above for $n = 5$.

Notice that L_1 and L_6 are necessarily smaller than the other lengths, and the value of T_5 will be lower than the other temperatures due to the shorter length.

temperature of a subset of a predefined sequence. The melting temperature of this subset is therefore only a function of x_1 and x_2.

The five defined arrays are discrete, which means that they are not differentiable. We will, however, define derivatives using the central difference method as follows (12):

$$\frac{\partial \psi(x)}{\partial x} \approx \frac{\psi(x+1) - \psi(x-1)}{2} \tag{12}$$

where Ψ is any one of the arrays listed above. As each array has a weak derivative in terms of x, the derivative of the melting temperature of the subset defined by x_1 and x_2 exists in terms of x_1 and x_2 by the chain rule. However, x can only have integer values ranging from 1 to N, where N is the total number of base pairs in the sequence.

Optimization of the oligonucleotide set

A DNA sequence of total length N can be divided into n overlaps. These overlaps can be defined by a $[(n-1) \times 1]$ vector $\bar{\mathbf{x}}$ such that overlap 1 ranges from position 1 to position $x_1 - 1$, overlap 2 from x_1 to $x_2 - 1$, etc., until the nth overlap ranging from position x_{n-1} to N. Each overlap has a melting temperature T_i. Each oligonucleotide consists of two overlaps, thus, the length of each oligonucleotide can be calculated by Eq. 13. The size of each element in vector L_i is the difference between two components of the position vector x_i. This is illustrated in Figure 4.

$$L_i = \begin{cases} x_i - 1 & i = 1,2 \\ x_i - x_{i-2} & 3 \le i \le n-1 \\ N - x_{i-2} + 1 & i = n, n+1 \end{cases} \tag{13}$$

The optimal oligonucleotide set will be chosen such that:

- The overlap melting temperature variance is minimized.
- The oligonucleotide lengths are constrained and minimized.

This can be realized by optimizing the following objective function (14):

$$O(\bar{\mathbf{x}}) = W_1[\Sigma_{i=1}^{n-1}(T_{\text{ave}} - T_i)^{P_1}] + W_2[\Sigma_{i=2}^{n}(\exp(L_i))^{P_2}] \tag{14}$$

The parameters W_i and P_i are weights used to fine tune the objective function. The first term in the objective function represents the melting temperature variance if $P_1 = 2$. By increasing the value of P_1, the objective function will penalize temperature outliers more severely. The second term in the objective function minimizes the oligonucleotide length, as an increase in the length directly increases the cost. It was found that a value of $P_2 = \frac{2}{L_{\max}}$ provides the algorithm with enough flexibility for the oligonucleotide length to vary within the length constraints (15), while the value of the second term increases drastically if $L_i > L_{\max}$.

$$0 < L_{\min} < L_{\text{i}} < L_{\max} \tag{15}$$

where $L_{\min}$ is the minimum oligonucleotide length and $L_{\max}$ the maximum length. The summation for each term ensures that only the relevant melting temperatures and oligonucleotide lengths are taken into account (as explained below).

To determine the initial guess, $\bar{\mathbf{x}}^{(0)}$, an initial annealing temperature T_0 is selected. Starting with the first overlap, x_1 is increased until $T_1 = T_{\text{m}}(1, x_1) > T_0$ or $L_1 = L_{\max}$. The second overlap temperature is then selected by increasing x_2 in the same way, until $T_{\text{m}}(x_1, x_2) > T_0$ or $L_2 = L_{\max}$. This procedure is repeated until the entire vector $\bar{\mathbf{x}}^{(0)}$ has been specified.

Since the BFGS method is a line search method and the initial vector $\bar{\mathbf{x}}^{(0)}$ meets the constraints, the optimization step size can be decreased until the next step $\bar{\mathbf{x}}^{(1)}$ (and subsequent steps) are within the constraints.

The gradient of the variances exists for integer values of $\bar{\mathbf{x}}$ and can be found using the chain rule. As the gradient exists, the BFGS method can be used to optimize the objective function.

In Figure 4, the vector $\bar{\mathbf{x}}$ has four elements, indicating the positions between x_1 through x_4 (not including the beginning, 1, and the end, N).

It is evident that L_1 and L_{n+1} will always be smaller than the other lengths. Furthermore, T_n will usually be lower than the other temperatures. To compensate for this, L_1, L_{n+1}, and T_n are not included in the calculation of the objective function (14). After optimization, T_n can be adjusted by adding additional base pairs to the final two oligonucleotides. This will extend the total DNA sequence slightly but should have no effect on the actual functionality.

It is also important that the vector $\bar{\mathbf{x}}$ has only integer elements. This is easily accomplished by merely rounding off to the nearest integer during BFGS optimization.

To begin the optimization sequence, an initial matrix $D^{(0)}$ must be chosen. The Hessian of the objective function is very difficult to compute, therefore Eq. 16 is used:

Table 2. Reaction Mixtures and Thermocycler Protocols Used in the Assembly of the *rpoB* and Chimeric Gene Oligonucleotide Sets (PCA) and Amplification of the Resulting Product (PCR)

	rpoB		Chimera	
	PCA	PCR	PCA	PCR
Reactions mixture				
Each dNTP (μmol/L)	200	200	200	200
$MgSO_4$ (mmol/L)	4	4	5	5
Nonacetylated BSA (μg/mL)	400	400	400	400
KOD hot-start polymerase (units)	0.5	0.5	0.5	0.5
Manufacturers polymerase buffer	1×	1×	1×	1×
Each oligonucleotide (nmol/L)	100	–	25	–
Primers (μmol/L)	–	0.2	–	0.7
PCA product (μL)	–	0.5	–	2
Thermocycler protocol				
Hot start	30 s at 94°C	30 s at 94°C	30 s at 94°C	30 s at 94°C
Cycles	30×	30×	30×	35×
Denaturing	2 s at 94°C	2 s at 94°C	2 s at 94°C	2 s at 94°C
Annealing	10 s at 58/60/62°C	3 s at 62°C	10 s at 56°C	3 s at 58°C
Elongation	10 s at 72°C	6 s at 72°C	10 s at 72°C	10 s at 72°C
Final elongation	25 s at 72°C	25 s at 72°C	25 s at 72°C	25 s at 72°C

$$D^{(0)} = \nabla \mathbf{O}(\bar{\mathbf{x}}) \otimes \nabla \mathbf{O}(\bar{\mathbf{x}}) + \max(|\lambda|) \tag{16}$$

where λ is a vector containing the eigenvalues of $\nabla \mathbf{O}(\bar{\mathbf{x}}) \otimes \nabla \mathbf{O}(\bar{\mathbf{x}})$ and $\otimes$ signifies the outer product of the two vectors. Equation 16 ensures that $D^{(0)}$ has the right dimensions and is positive definite. All calculations were done using GNU Octave v. 3.2.3.[12]

Materials and Methods

The methods described above were used to optimize the oligonucleotide set for two DNA sequences of length 550 and 1416 bp. The sequences used were a subset of an *rpoB* gene (550 bp) and a custom chimeric gene (1416 bp) consisting of protein-C receptor and thrombomodulin domains. Both DNA constructs are based on genes found in nature and pertinent to other research conducted in our laboratory.

The PCA reaction was carried out as described in literature.[13] The oligonucleotide sets for the *rpoB* and chimeric genes (as calculated using the BFGS method) were assembled experimentally. A different oligonucleotide design tool (GeneOligo[2]) was used to find an oligonucleotide set for the synthesis of the *rpoB* gene. This set was also assembled experimentally for comparison. All three of these oligonucleotide sets were purchased from Integrated DNA Technologies (Coralville, IA). For the *rpoB* gene, the BFGS method prescribed a set of 33 oligonucleotides and the Gene2Oligo tool 44 oligonucleotides, whereas the BFGS method prescribed a set of 78 oligonucleotides for the chimeric gene. The *rpoB* gene was assembled using three different annealing temperatures, 58, 60, and 62°C to determine how much this affects the assembly.

A PCRJet® thermocycler (Megabase Research Products, Lincoln, NE) was used to conduct the PCA reaction as well as the PCR amplification of the assembled product, using a reaction volume of 25 μL. The reaction mixtures and thermocycler protocols used are summarized in Table 2. Internal primers were used for the *rpoB* amplification, resulting in a product of length 450 bp.

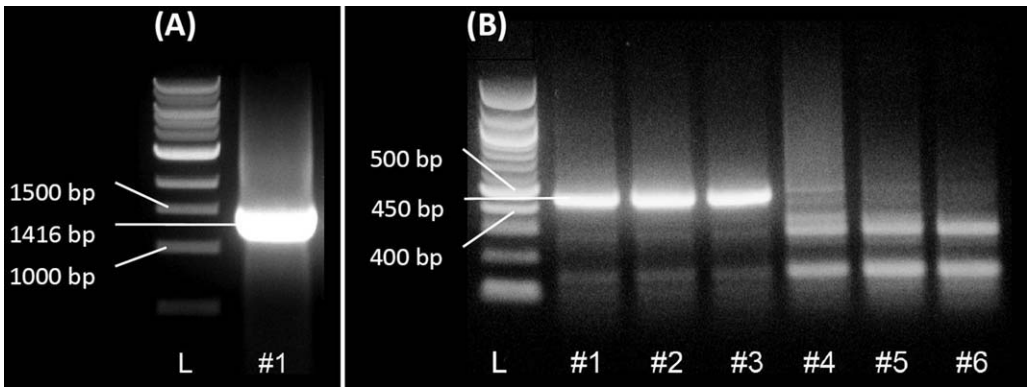

Figure 5. Gel electrophoreses results showing: A: the assembled 1416 bp chimera gene (1) alongside a 1 kbp DNA ladder (*L*): it is clear that the PCA process produced a construct of correct length, as is evident from the extremely bright product lane; B: 100 bp DNA ladder (*L*), the assembled and amplified 450 bp *rpoB* sequence using oligonucleotides prescribed by the BFGS method (1–3) and the same sequence using Gene2Oligo[2] (4–6).

The oligonucleotide set prescribed by the BFGS method assembled correctly—it is clear that the other set assembled two smaller fragments.

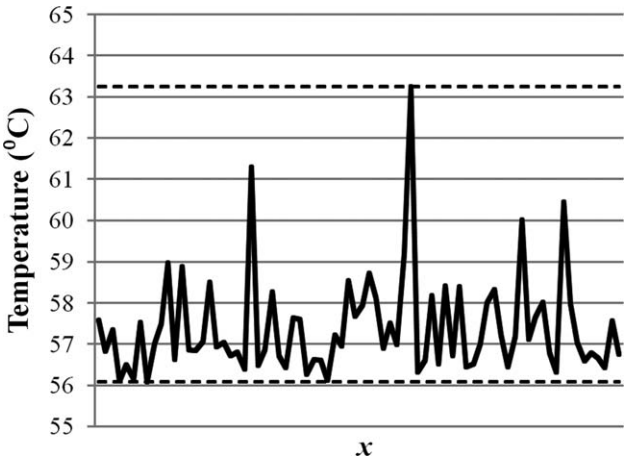

Figure 6. Overlap melting temperature for chimera sequence (1416 bp), showing that melting temperature varies between 56 and 63°C.

The PCR products were mixed with 5 μL 6× loading dye (New England Biolabs, Ipswich, MA). For the *rpoB* construct, 5 μL of this mixture was used to verify the product using gel electrophoreses. A 1% agarose gel, stained using GelRed (Biotium, Hayward, CA), was used. The same was done for the chimeric construct, however, 12 μL of the PCR product/loading dye mixture was used and the gel was stained using ethidium bromide. DNA ladders purchased from New England Biolabs (Ipswich, MA) were used as markers.

Results

The gel electrophoreses results showing the assembled chimera (1416 bp) and *rpoB* gene (450 bp after PCR) are shown in Figure 5.

The oligonucleotide set produced for the chimera sequence proved that the BFGS method produces a set suitable for assembly (Figure 6). The gel electrophoreses results proves that the gene assembled correctly (Figure 5A). The *rpoB* sequence shows that better results can be obtained using this method than using older software (Figure 7)—this is clearly seen by comparing the gel electrophoreses results (Figure 5B) obtained when using the oligonucleotide set prescribed by the BFGS method and Gene2Oligo.[2] The *rpoB* sequence, while short, proves difficult to assemble due to distinct stretches of GC-rich DNA. The BFGS method can accommodate these subsets as it uses the latest salt-correction factors.

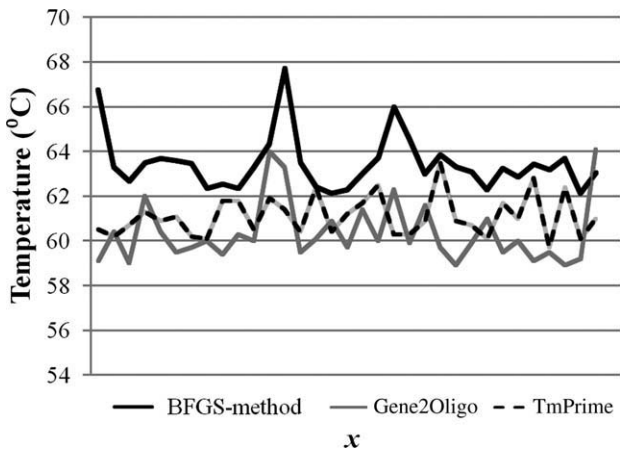

Figure 7. Overlap melting temperature for the *rpoB* sequence (550 bp) as predicted by the BFGS method, Gene2Oligo,[2] and TmPrime.[4]

It is clear that the BFGS method prescribes higher temperatures than the other methods—this may be due to added salt correction and minimum overlap length constraints.

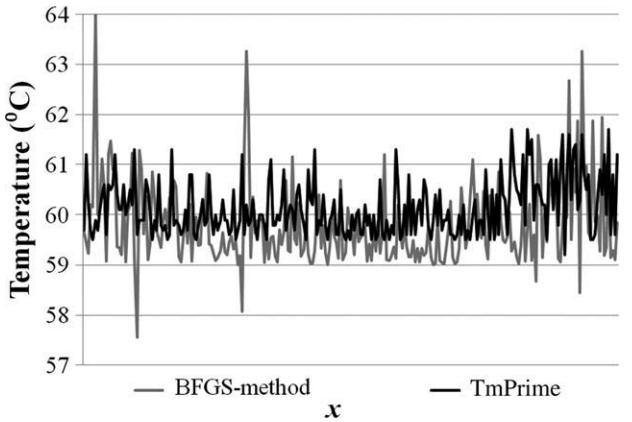

Figure 8. Overlap melting temperature for plasmid sequence (6890 bp) using the oligonucleotide sets prescribed by the unconstrained BFGS method and TmPrime.[4]

Clearly, TmPrime gives better results. However, the fact that no constraints may be applied poses additional problems.

The oligonucleotide set required to assemble a plasmid gene subset (6890 bp) was also calculated, although not experimentally assembled. The BFGS method was compared

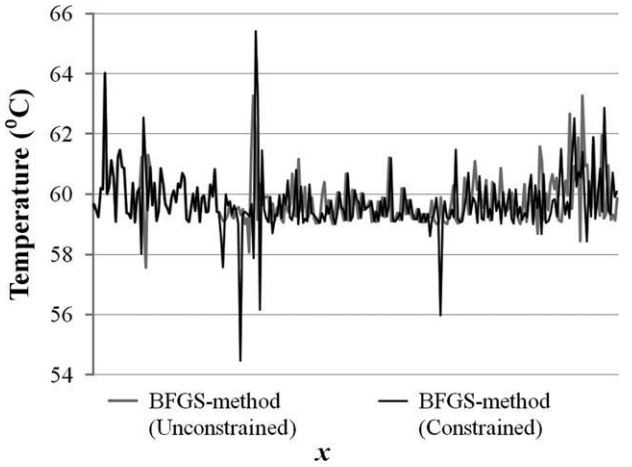

Figure 9. Overlap melting temperature for plasmid sequence (6890 bp) using the oligonucleotide sets prescribed by the unconstrained and constrained BFGS method.

Constraining the problem increases the melting temperature variance, as would be expected.

with the latest freely available oligonucleotide optimization software, TmPrime.[4] Figure 8 shows that the results obtained by the BFGS method and the other tool[4] are comparable when oligonucleotide lengths are unconstrained. However, this tool does not allow for any oligonucleotide length constraints. The BFGS method provides a second, workable oligonucleotide set under the constraint $L_{max} < 60$ bp, as shown in Figure 9.

Clearly, the melting temperature can be confined to a range of ± 4°C if the outliers are excluded. This becomes more difficult as the length of the sequence increases.

It is noteworthy to point out the speed of calculation. The oligonucleotide set for the plasmid (>5000 bp) was computed in less than 10 s (using a Dell laptop with a 1.83 GHz Intel® Centrino® Duo core). Other methods report times of <4 min for similarly sized sequences.[4]

Conclusions

It was shown that the BFGS optimization routine could be used to determine a near optimal set of oligonucleotides to be used in the assembly of longer DNA constructs. The method predicts similar if not superior sets than other software programs but with markedly decreased computation times.

No conclusions were reached about the optimality of the solution produced. Given different initial conditions, slightly different results were obtained. The main problem is that different initial conditions separate the sequence into a different number of overlaps. This effectively changes the dimensionality of the problem. However, as an engineering application the focus does not lie on achieving a true optimum in this case—rather in reaching a workable solution with minimal computation time. Furthermore, due to the rapid computation time, users can select different initial conditions and compare results in quick succession.

Acknowledgments

The authors gratefully acknowledge the financial support of the National Institutes of Health through grant R33 RR022860.

Literature Cited

1. Czar MJ, Anderson JC, Bader JS, Peccoud J. Gene synthesis demystified. *Trends Biotechnol*. 2009;27:63–72.
2. Rouillard JM, Lee W, Truan G, Gao X, Zhou X, Gulari E. Gene2-oligo: oligonucleotide design for in vitro gene synthesis. *Nucleic Acids Res*. 2004;32:176–180.
3. Hoover DM, Lubkowski J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res*. 2002;30:e43.
4. Bode M, Khor S, Ye H, Li MH, Ying JY. TmPrime: fast, flexible oligonucleotide design software for gene synthesis. *Nucleic Acids Res*. 2009;37 (Web Server issue):214–221.
5. Peressini A, Sullivan F, Uhl J. *The Mathematics of Nonlinear Programming*. New York: Springer, 1988.
6. SantaLucia J. A unified view of polymer, dumbbell, and oligonucleotide DNA. *Proc Natl Aacd Sci USA*. 1998;95:460–1465.
7. Owczarzy R, Moreira B, You Y, Behlke M, Walder J. Predicting stability of DNA duplexes in solutions. *Biochemistry*. 2008;47:5336–5353.
8. Allawi H, SantaLucia J. Thermodynamics and NMR of internal G.T mismatches in DNA. *Biochemistry*. 1997;36:10581–10594.
9. Owczarzy R, Dunietz I, Behlke M, Klotz I, Walder J. Thermodynamic treatment of oligonucleotide duplex-simplex equilibria. *Proc Natl Acad Sci USA*. 2003;100:14840–14845.
10. Owczarzy R, Tataurov A, Wu Y, Manthey J, McQuisten K, Almabrazi H. IDT SciTools: a suite for analysis and design. *Nucleic Acids Res*. 2008;36:163–169.
11. IDT SciTools OligoAnalyzer 3.1. 2009. Integrated DNA Technologies: Available at: http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/. Accessed on June 3, 2009.
12. Eaton JW. GNU Octave v. 3.2.3, Octave homepage, 2010. Available at: http://www.gnu.org/software/octave/. Accessed on July 30, 2010.
13. Mamedov TG, Padhye NV, Viljoen H, Subramanian A. Rational de novo gene synthesis by rapid polymerase chain assembly (PCA) and expression of endothelial protein-C and thrombin receptor genes. *Biotechnology*. 2007;131:379–387.

Manuscript received Feb. 19, 2010, and revision received Aug. 6, 2010.