DIMACS Series in Discrete Mathematics
and Theoretical Computer Science

Thermodynamic Simulation of Deoxyoligonucleotide Hybridization for DNA Computation

Alexander J. Hartemink and David K. Gifford

Abstract. Nearly every model of DNA computation proposed to date depends upon sequence-specific hybridization operations. In order to better predict the binding specificity of arbitrary deoxyoligonucleotides, a simulator named BIND is implemented. BIND operates on a single template DNA sequence and a number of shorter primer sequences. For each primer sequence, BIND calculates a theoretical melting temperature at every position of the primer along the template, yielding a measure of binding specificity between each primer and the template. The simulator differs from previous melting temperature programs in that it is intended to be used with oligonucleotides, is designed to handle mismatched base pairs, makes use of the latest thermodynamic parameters, and provides features with DNA computation expressly in mind. This paper describes how BIND is implemented, provides corroborating evidence as to its accuracy, and offers instances of its usefulness to a range of DNA computing applications.

1. Introduction

Recent activity in the field of DNA computing has generated interest in the dynamics of nucleic acid interaction. By creating another consumer of precise information regarding enzyme activity, thermal and ionic denaturation, binding specificity, and the polymerase chain reaction (PCR), DNA computing has enhanced the already compelling case for deeper fundamental understanding of these processes.

In particular, since many proposed designs for DNA-based computational systems rely on assumptions about the binding specificity between two or more strands of DNA, a thorough investigation of the hybridization process is of paramount importance. To our knowledge, *all* current DNA-based computation proposals make use of annealing steps, whether in constructing large libraries of DNA on which to perform computations, in using PCR to amplify or detect small quantities of DNA containing a specific sequence of nucleotides [**2, 3**], in performing sequence-specific separation operations using magnetic bead or affinity column techniques [**2, 3, 10**], in setting specific bits in sticker-based models [**10**], in marking specific strands in

1991 *Mathematics Subject Classification.* Primary 92-04, 92C40, 80-04; Secondary 68U20.

Hartemink is partially funded by a National Science Foundation Graduate Research Fellowship.

surface-based models [**6**], or in performing sequence-specific mutagenic operations in programmed mutagenesis models. Because these operations all rely on site-specific nucleic acid hybridization, it is critical that significant progress be made in more precisely characterizing the way the temperature, ionic environment, and degree of mismatch influence the hybridization process.

By evaluating the current state of DNA-based computation, and by looking in the direction in which the field will need to expand in the next few years, it becomes clear that many of the proposed computational systems will need to move "from the blackboard into the laboratory" in order to iron out the subtle details associated with their implementation. Before particular problems can be investigated, however, the algorithms (and the input to those algorithms) will need to be translated into concrete DNA sequences, which is a daunting task, especially in light of the necessary binding specificity demands made by the systems' specifications. With this in mind, a simulator capable of quickly evaluating the binding specificity of oligonucleotides serving as PCR primers, stickers, markers, probes, or mutagenic rewrite rules should prove helpful to the DNA computing community, as well as to the molecular biological community at large.

Programs which compute DNA melting temperatures do exist, but most are intended only for polynucleotide hybridization, do not handle mismatched nucleotide binding or multiple binding possibilities, or use outdated thermodynamic parameters. Moreover, to our knowledge, none of these programs is designed with DNA computation in mind. Therefore, there appears to be a niche for an accurate simulator that uses the latest thermodynamic parameters, is capable of handling nucleotide mismatches, and is designed to provide needed functionality to scientists investigating DNA computation.

Overview of this Paper. The preliminary implementation of such a simulator, evaluations of its accuracy, and demonstrations of its general applicability provide the basis of this paper. The simulator described here, named BIND, has been calibrated using melting temperature data reported in the literature and has already begun to provide useful information to our group's ongoing work in the area of programmed mutagenesis. In the following section, we briefly describe programmed mutagenesis in order to provide a contextual motivation for our interest in such a simulator. Then, in the next section, we proceed to discuss the BIND simulator itself and derive the thermodynamic equations used in its melting temperature calculations. Then we turn to an evaluation of BIND's accuracy by comparing it with melting temperature data determined in the laboratory and reported in the literature. This is followed by a discussion of the simulator's uses and some of the results obtained with the aid of the simulator. In the last section, we conclude by presenting a number of possible extensions to BIND which we are in the process of developing.

2. Programmed Mutagenesis

Programmed mutagenesis is a technique for programmatically rewriting DNA sequences by incorporating highly sequence-specific oligonucleotides into newly manufactured strands of DNA. There are several advantages to using programmed mutagenesis for DNA-based computation:

- The pool of oligonucleotides can be designed to cause sequence-specific programmed changes to occur, including the propagation of programmed

changes up and down a DNA molecule and the evolution of a programmed sequence of changes over the course of future replication events. Thus, sequential computations with programmatically evolving state can be carried out, resulting in *constructive computation*, as contrasted with *selective computation* which requires all possible solutions to a problem to be present *ab initio*.

- The sequence specificity of the oligonucleotides permits a set of oligonucleotides to be present at each step of the reaction, with only a fraction of them being active during each cycle. This reduces human effort since it allows the computation to be carried forward by thermocycling the reactants in the presence of thermostable polymerase and ligase. Ideally, there would be no need for human (or robotic) intervention between computation steps.
- Although input and output of data from programmed mutagenesis reactions can be accomplished by the direct synthesis of input molecules and sequencing of output molecules, other more indirect methods exist which lend this technique greater flexibility and modularity. For example, input molecules could be created by ligating computational DNA subunits together, with the subunits chosen to encode specific problems.
- Programmed mutagenesis can be used in conjunction with other proposed computational systems by creating output strands to be used as input to other systems. In this manner, input strands for other systems could be generated in a systematic fashion. Because of this "intelligent design" property, the need to have all possible solutions present *ab initio* may be obviated.
- All the components necessary to implement programmed mutagenesis are present *in vivo*. Therefore it may eventually be possible to harness the internal workings of the cell for computation, thereby capitalizing on the cell's homeostatic capabilities to ensure that the computation takes place in a stable chemical environment.

We have begun to build programmed mutagenic machines in the laboratory and believe the techniques to be generally extensible. In the subsection below, we describe in detail how we have applied the technique of programmed mutagenesis to build a simple unary counter.

2.1. The Unary Counter Machine. To demonstrate the function of programmed mutagenesis, we implemented a counter that will in theory be incremented every time DNA is replicated. We encode the value of the counter in a DNA molecule as the number of **X** and **Y** symbols it contains, where **X**, **Y**, and **Z** symbols are shorthand representations of specific 12-nucleotide sequences, designed so that both **X** and **Y** differ from **Z** at two base positions, but from each other at four. The initial template molecule includes the sequence **ZZZZZX**, which encodes the number one. The mutagenic oligonucleotides which implement the counting mechanism are $\mathbf{X'Y'}$ and **XY**, denoted ρ_1 and ρ_2, respectively.

This counter construction presumes that 24-mer oligonucleotides which anneal to the template with two mismatches can be extended and successfully ligated to other polynucleotides, while oligonucleotides with a greater number of mismatches cannot be effectively incorporated into the new strand. Note that we do not need as many oligonucleotides as counter values; in this case, two oligonucleotides are sufficient to allow the counter to advance to an arbitrary value.

Each cycle, the counter uses as its input template the product of the immediately preceding cycle, and the product of cycle N encodes the number $N+1$. The counter is implemented by thermocycling a reaction that begins with the initial template strand, ρ_1, ρ_2, thermostable polymerase and ligase, and outside primers LP and RP (needed to produce the full-length product on each cycle). Each thermal cycle consists of a high temperature step, which denatures the double-stranded DNA and prepares it for the polymerization-ligation step; and a low temperature step, which is permissive for primer annealing, polymerization, and ligation. Ideal values for these temperatures were chosen on the basis of data from laboratory experiments. In a later section of this paper, the BIND simulator is used to demonstrate how such choices might be made more easily, circumventing the need for detailed laboratory experiments.

2.2. Summary. The salient point regarding programmed mutagenesis is that it relies on the binding specificity of its rewrite rules to ensure that the strand of DNA is being rewritten in a systematic way. For example, if rewrite rule ρ_i is meant to be applied to a strand of DNA representing state σ_i, producing a strand representing state σ_{i+1}, and rewrite rule ρ_{i+1} is subsequently meant to be applied to the strand of DNA representing state σ_{i+1} to produce a strand representing σ_{i+2}, it should be the case that ρ_{i+1} cannot be applied to σ_i and ρ_i cannot be applied to σ_{i+1}. If this condition is satisfiable, then both of the rewrite rules can be present in the reaction and yet the system can only evolve from the state representing σ_i to the state representing σ_{i+2} by first passing through σ_{i+1}, with each rewrite rule being applied in sequence, thereby capturing the notion of programmatic computation.

3. Simulator Design

In order to provide a means for rapidly testing binding specificity hypotheses, a simulator called BIND was developed. After being provided a single template sequence and any number of primer sequences, the BIND simulator considers individually how each primer can bind to the template strand, at the specified ionic and reactant concentrations. It does this by calculating, for each possible position of the given primer along the template strand, a theoretical melting temperature for the primer-template duplex at that position. A straightforward loop generates melting temperatures for all such positions along the template. This procedure is repeated for each primer. The resulting data can be used to generate plots revealing the most stable binding sites along the template for each of the primers.

Note that if the template is the same length as the primer, BIND just calculates the melting temperature between the two strands at a single position, the position where the two are perfectly aligned. Therefore, calculation of the melting temperature of a simple double-stranded segment of DNA is easily handled. Consequently, we can test BIND's melting temperature calculation by comparing its output with data reported in the literature for a number of oligonucleotide duplexes.

3.1. The Calculation of Theoretical Melting Temperature. Theoretical melting temperature is typically calculated assuming that the coil-helix transition is two-state, which is a justifiable assumption for small oligonucleotides.[1] SantaLucia, *et al.* [**11**] suggest that the two-state model is capable of providing a reasonable approximation of melting temperature for duplexes with non-two-state transitions,

[1] Usually defined in this context as fewer than 15-25 bases.

but the applicability of the assumption obviously decreases as the size of the duplex under consideration increases. For a two-state model of the

$$\text{single-strand} + \text{single-strand} \rightleftharpoons \text{double-strand}$$

transition between two distinct oligonucleotides in equimolar concentration, the equilibrium constant is given by

$$K = \frac{2f}{(1-f)^2[C_T]} \tag{3.1}$$

where f is the fraction of strands in the double-stranded state and $[C_T]$ is the total molar strand concentration.

But the equilibrium constant can also be expressed in thermodynamic terms as

$$K = \exp(-\frac{\Delta G^\circ}{RT}) = \exp(-\frac{\Delta H^\circ - T\Delta S^\circ}{RT}) \tag{3.2}$$

where R is Boltzmann's constant and T is the temperature in Kelvin. Since the melting temperature, T_m, is defined as the temperature at which half of the strands are in the double-stranded state, it follows that $f = 1/2$ when $T = T_m$. Setting $f = 1/2$ and setting the two expressions for K in equations 3.1 and 3.2 equal to one another, we get an equation relating the melting temperature, T_m, to the total molar strand concentration and the enthalpy and entropy of the forward state transition:

$$T_m = \frac{\Delta H^\circ}{\Delta S^\circ + R\ln([C_T]/4)} \tag{3.3}$$

In order to adjust for $[Na^+]$ concentrations different from 1M, the salt adjustment term[2] introduced by Wetmur [**13**] is appended to the right hand side of equation 3.3, yielding equation 3.4:

$$T_m = \frac{\Delta H^\circ}{\Delta S^\circ + R\ln([C_T]/4)} + 16.6\log_{10}\left(\frac{[Na^+]}{1+0.7[Na^+]}\right) - 269.3 \tag{3.4}$$

The nearest neighbor stacking model allows the total enthalpy and entropy of the transition to be expressed as

$$\Delta H^\circ = \Delta H^\circ_{ends} + \Delta H^\circ_{init} + \sum_{k\in\{stacks\}} \Delta H^\circ_k \tag{3.5}$$

$$\Delta S^\circ = \Delta S^\circ_{ends} + \Delta S^\circ_{init} + \sum_{k\in\{stacks\}} \Delta S^\circ_k \tag{3.6}$$

Equations 3.4, 3.5, and 3.6 are similar to the ones reported in Wetmur [**13**], but they make use of the entropy, ΔS°, rather than the free energy, ΔG°, which is temperature dependent. Since temperature is the variable of interest in this calculation, the expression for melting temperature which does not include free energy terms is preferred.

[2]Note that when $[Na^+]$ is 1M, the term drops out.

If divalent cations like Mg^{++} are present in solution, their effect on duplex stability can be expressed in terms of an $[Na^+]$ equivalent using the conversion [**14**]:

$$[Na^+]_{equiv} = 4\sqrt{[Mg^{++}]} \tag{3.7}$$

For self-complementary oligonucleotides, the $[C_T]/4$ term is replaced by $[C_T]/2$: the equilibrium constant K changes by a factor of 4, but because of symmetric entropic considerations, another factor of 2 must be introduced in the other direction, partially offsetting the first factor [**7, 14**]. If the strands are not in equimolar concentration, but one strand is present in gross excess over the other, the $[C_T]/4$ term becomes $[C_T]$. Intermediate cases can also be handled by modifying this term appropriately.

The thermodynamic parameters used in calculating the enthalpy and entropy values in equations 3.5 and 3.6 are taken from the literature. The BIND simulator's modular design enables it to use multiple thermodynamic basis sets, so it was tested on two different basis sets. The enthalpy and entropy values for nearest neighbor stacks were taken first from Quartin and Wetmur [**9**] and then from SantaLucia, Allawi, and Seneviratne [**11**]. The former paper uses enthalpy values previously determined by Breslauer, *et al.* [**5**] but provides new entropy values.

Neither of these basis sets accounts for mismatched base pairs, so mismatches were handled by treating them as "virtual stacks", as described by Werntges, *et al.* [**12**]. The simulator uses the thermodynamic parameters associated with virtual stacks reported in their paper, but a correction is incorporated to compensate for the specific nucleotide context in which their mismatches occurred. Since it is possible for multiple mismatched base pairs to occur consecutively and produce a large internal loop, BIND treats such a loop as a single large virtual stack, adding the average of all the mismatch enthalpies and entropies for the bases in the loop, rather than adding the mismatch enthalpy and entropy for each.

3.2. Calibration. In order to calibrate the BIND simulator, the melting temperatures for ninety-three oligonucleotide sequences were calculated and compared with experimental melting temperatures reported in the literature. Each sequence and concentration was provided as input to the simulator which computed the melting temperatures for all the sequences using each of the two thermodynamic parameter basis sets, with the primary intention of ensuring BIND's accuracy and with the secondary intention of determining which basis set's parameters yielded better predictions of the oligonucleotide duplex melting temperature.

The calibration process was relatively *ad hoc* in that it was only used to verify that the parameter basis sets were reasonable predictors of melting temperature for a vast range of oligonucleotides. The results of this process should be taken with a grain of salt. For example, there is a dependence issue arising from the fact that forty-four of the oligonucleotide sequences tested were also used by SantaLucia, *et al.* [**11**] in the determination of their thermodynamic parameters so one would expect their basis set to perform quite well on those sequences. Similarly, the Quartin and Wetmur [**9**] thermodynamic parameter basis set was calculated using some of the other test sequences, again invalidating independence assumptions. Consequently, these results should not be misinterpreted as conclusive proof that one basis set is superior to the other.

TABLE 1. Results of BIND calibration tests

		Quartin & Wetmur		SantaLucia		Weighted 3:8	
Test	Number	*av* ΔT	*av* $\lvert\Delta T\rvert$	*av* ΔT	*av* $\lvert\Delta T\rvert$	*av* ΔT	*av* $\lvert\Delta T\rvert$
Quartin & Wetmur	12	-1.0	2.1	4.2	4.2	2.7	2.8
SantaLucia	44	-2.5	4.5	0.3	1.8	-0.5	1.9
Aboul-ela Mismatches	16	-1.4	3.3	-2.1	3.4	-1.9	3.3
Werntges Mismatches	16	1.6	1.8	1.8	1.9	1.8	1.8
Salt Concentration	5	-5.2	5.2	0.5	0.9	-1.1	1.1
Total	93	-1.6	3.5	0.6	2.3	0.0	2.2

av ΔT indicates the average difference between the calculated T_m and the experimental T_m.
av $\lvert\Delta T\rvert$ indicates the average absolute value of the difference between the calculated T_m and the experimental T_m.

Thirty-two mismatched sequences from work by Aboul-ela, *et al.* [**1**] and Werntges, *et al.* [**12**] were also tested to verify that the simulator is predicting melting temperatures for sequences with mismatched base pairs reasonably well. Additionally, a few sequences from Quartin and Wetmur [**9**] were used to test the simulator's ability to handle sequences with dangling ends. Finally, a single sequence from work by Braunlin and Bloomfield [**4**] was tested at five different salt concentrations to verify that the melting temperature is being predicted correctly as the ionic environment changes.

A summary of the results of these test runs is presented in table 1. As expected, the Quartin and Wetmur basis set outperformed the SantaLucia basis set on its own sequences and *vice versa*. Both sets of parameters performed equally well on the sequences containing mismatches. On the whole, the average absolute deviation from experimentally determined melting temperature was 3.5° C for the Quartin and Wetmur basis set and 2.3° C for the SantaLucia basis set. The former tended to underpredict T_m by 1.6° C, while the latter tended to overpredict T_m by 0.6° C. When the predicted temperatures from the two basis sets were averaged with weights of 3 and 8 to compensate for the under- and overprediction, the average absolute deviation fell to 2.2° C.

It should be emphasized that no outlying data points were discarded, the test sequences were not culled for their simplicity, a single algorithm is being used for all the sequences, whether they contain mismatches or dangling ends or exist in an environment with low salt concentration, and no changes were made to the algorithm or underlying parameters on the basis of these test sequences. In fact, absolutely no optimization of the algorithm was performed *ex post facto*. Therefore, it is expected that this performance could be improved somewhat if a new set of thermodynamic parameters were determined (especially parameters describing the thermodynamics of base pair mismatches) or a few adjustments were made to the algorithm in light of these test results.

In conclusion, neither basis set is definitively superior to the other, especially in light of the dependence issues mentioned earlier. Nevertheless, because the SantaLucia basis set slightly outperformed the Quartin and Wetmur basis set, melting temperatures for the remainder of this paper will be calculated using the SantaLucia basis set.

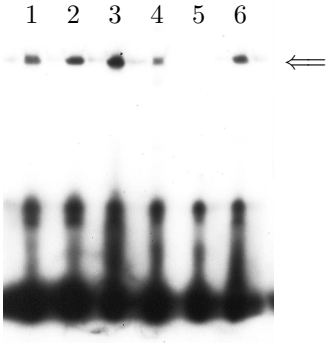

FIGURE 1. Experiment 45 gel, lanes 1-6: The six lanes are indicated by the numerals at the top, while the band of interest is indicated by the arrow at right. Descriptions of the reactions in each lane appear in the text.

3.3. Laboratory Confirmation. In addition to calibrating the simulator using melting temperature data reported in the literature, two experiments were carried out in our own laboratory to solidify the conclusion that BIND is useful in providing accurate melting temperature predictions.

In the first experiment, three distinct primers with differing degrees of mismatch were individually mixed with a single template strand at two different temperatures and given time to be extended by polymerase. Lanes 1-3 in the gel shown in figure 1 correspond to the three primer reactions run at 45° C, while lanes 4-6 in the gel correspond to the three primer reactions run at 55° C. The gel clearly indicates that while all three primers are successfully extended at 45° C, only the first and third primer are successfully extended at 55° C. This would suggest that the melting temperature of the primer-template duplex at the optimal binding position[3] along the template for the first and third primers would be above 55° C, but would be between 45° C and 55° C for the second primer. In fact, BIND predicts the melting temperatures to be 60.2° C, 47.7° C, and 65.6° C, respectively.

In the second experiment, two distinct primer template combinations were given time to be extended by polymerase at four different temperatures. Both 24-mer primers possessed two mismatches in the optimal binding position along their respective templates. Lanes 1-4 in the gel shown in figure 2 correspond to the first reaction being run at 45° C, 50° C, 55° C, and 60° C, respectively. Lanes 5-8 correspond to the second reaction being run at the same four temperatures. The gel indicates that while the first reaction allows polymerase extension at all four temperatures, the second reaction seems to falter between 55° C and 60° C. Indeed, BIND predicts the melting temperature of the primer template duplex to be 60.4° C for the first reaction and 57.1° C for the second, in close agreement with the conclusions drawn in the laboratory.

Additional confirmation of BIND's ability to find the optimal binding location along the template strand is given by the fact that in each of these experiments involving mismatched oligonucleotides, the optimal binding position reported by

[3]Optimal binding position simply means the position with the highest melting temperature.

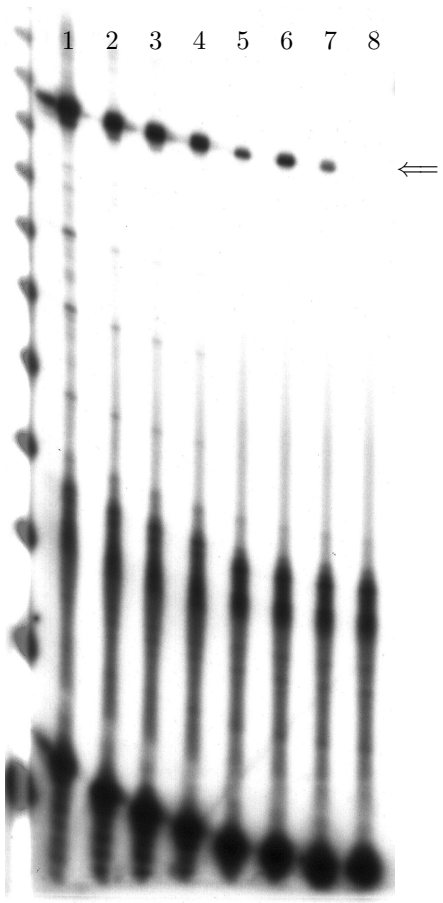

FIGURE 2. Experiment 74 gel, lanes 1-8: The eight lanes are indicated by the numerals at the top, while the band of interest is indicated by the arrow at right. Descriptions of the reactions in each lane appear in the text. This gel is slightly skewed, with the bands sloping gradually downwards moving from left to right.

BIND corresponded precisely with the optimal binding position as determined by the length of the polymerase-extended product appearing in the gels. Thus it seems likely that BIND is successful in locating optimal primer binding position along template strands, even in the presence of base pair mismatch.

4. Using the Simulator

Now we turn to a discussion of some examples of how BIND can be used in designing DNA-based computational systems, with reference to assumptions about binding specificity. In the following subsections, we consider a few steps in the design of a simple unary counter, much like the one being constructed in our laboratory. Throughout these subsections, the template under consideration consists

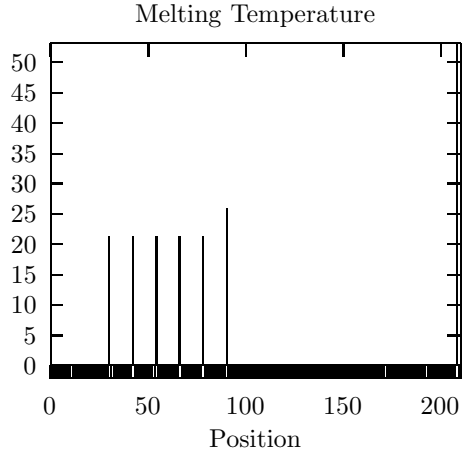

FIGURE 3. Primer binding: Melting temperature as a function of position for an outside primer with respect to the cycle 1 template. The primer should bind specifically at the end of the template but not interfere anywhere else.

of 224 nucleotides and contains within it the sequence **ZZZZZX**, as mentioned before. This sequence will henceforth be referred to as the *active site* of the template. Also as before, the first cycle rewrite rule, denoted ρ_1, is $\mathbf{X'Y'}$ and the second cycle rewrite rule, denoted ρ_2, is $\mathbf{XY}$.

4.1. Primer Specificity. As an example of BIND's usefulness, consider the choice of an outside primer for the template used in the unary counter machine. After selecting a site away from the active site for a primer to bind, the outside primer is chosen to bind perfectly to the selected site. For this primer to be safe to use, however, it must be the case that the primer does not interfere with any of the other reactions which may be taking place in the test tube simultaneously. In particular, the primer should not be able to bind in the active site of the template.

Once chosen, the template and primer are passed to BIND as input, whereupon BIND returns the information displayed in figure 3. In the rightmost position (corresponding to the $3'$ end of the template), there is a perfect match between the template and the selected 17-mer primer, yielding a melting temperature around 53° C. The primer has very unfavorable binding specificity for most of the length of the template strand, but some potential interference can be detected near the active site for temperatures in the 20-25° C range. Although this is unlikely to cause a problem for high-temperature reaction cycles, BIND has indicated a possible source of interference between the outside primer and the active site, leading a machine designer to ensure that temperatures remain well above 25° C.

4.2. Rewrite Rule Incorporation and Sequential Computation. Programmed mutagenesis relies on the sequence specificity of its rewrite rules in order to guarantee that certain steps are being performed before others. Recall that if rewrite rule ρ_i is meant to be applied to a strand of DNA representing state σ_i, producing a strand representing state σ_{i+1}, and rewrite rule ρ_{i+1} is subsequently meant to be applied to the strand of DNA representing state σ_{i+1} to produce a

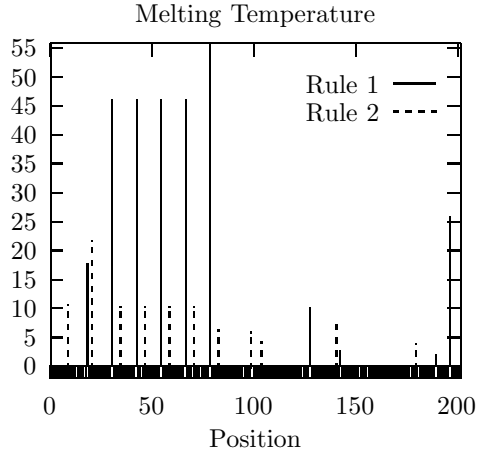

FIGURE 4. Cycle 1 binding: Melting temperature as a function of position for rewrite rules 1 and 2 with respect to the cycle 1 template. In this cycle, rule 1 should bind specifically while rule 2 remains uninvolved.

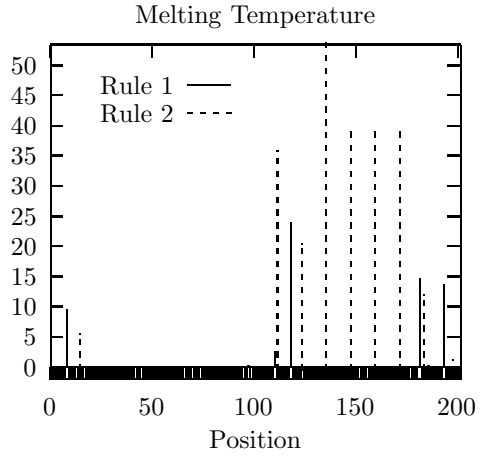

FIGURE 5. Cycle 2 binding: Melting temperature as a function of position for rewrite rules 1 and 2 with respect to the cycle 2 template. In this cycle, rule 2 should bind specifically while rule 1 remains uninvolved.

strand representing σ_{i+2}, it should be the case that ρ_{i+1} cannot be applied to σ_i and ρ_i cannot be applied to σ_{i+1}. In the context of the unary counter, we can use BIND to test that ρ_1 can bind to the template in the first cycle but not in the second, while ρ_2 can bind in the second cycle but not in the first. Additionally, we would like to be able to determine a temperature which would guarantee this binding specificity condition is met.

To test this, the same template as above can be passed to BIND, along with both rewrite rules, ρ_1 and ρ_2. The results for this first cycle reaction are shown in

figure 4. Then the template which is the product of the first cycle can be passed to BIND, again with both rewrite rules. Figure 5 displays the results for this second cycle reaction. As the figures indicate, the correct rewrite rule binds to the template in the correct cycle, with very little interference from the opposing rewrite rule. As in the case of the outside primer, if the temperature of the reaction is maintained well above 25° C, no cross-rule interference should be observed.

However, these two figures reveal that it is possible for the rewrite rules themselves to inappropriately bind to the template elsewhere in the active site, annealing to any **ZZ** subsequence in the template. In order to prevent this, reactions should be run somewhere in the 47-52° C range. In fact, these same observations were made in the laboratory before BIND was developed.

4.3. Sticker Testing. As a final sample application of BIND, an experiment was conducted to measure the likelihood of sticker interference in the sticker model proposed by Roweis, *et al.* [**10**]. A template strand of length 10,000 was generated at random. Then five stickers were chosen, of lengths 10, 15, 20, 25, and 30, respectively. Each sticker was guaranteed to match the template in at least one location because the stickers were chosen to be complementary to a subsection of the original template.

The 10,000-mer template was passed to BIND along with each sticker. The sticker of length 10 bound to its complementary subsection at a temperature of 53° C, at another location at 46° C, and at a third at 34° C. No other locations had melting temperatures above 34° C. For stickers of length 15 and 20, no "false" binding position was found with a melting temperature within 30° C of that of the optimal binding location, while for stickers of length 25 and 30, no "false" binding position was found with a melting temperature within 40° C of that of the optimal binding location. Of course, one strand constructed haphazardly does not a proof of specificity make, but it is reassuring that BIND is able to confirm that the randomly chosen stickers of length at least 15 anneal to the template strand in this example in a highly specific fashion.

This simple-minded demonstration is intended to illustrate how BIND can be used in other DNA computation contexts, such as the verification of sticker choices. Although sticker interference is unlikely at sufficiently long lengths, BIND can be used to conclusively rule out such interference or point out where such interference might be occurring. It also provides valuable information about the temperature at which sticker operations should be carried out in order to maximize binding specificity.

For this reason, the simulator seems to hold a great deal of promise, not only for use in programmed mutagenesis contexts but also in the context of many other proposed models of computation. In fact, it should prove valuable to any research effort involving site-specific annealing, whether in biological computation or molecular biology at large.

5. Extensions

The BIND simulator is being extended in a number of directions. First, because polymerization plays a large role in some proposed DNA-based computational systems, it would be helpful if BIND categorized various binding sites by the stability of their open $3'$ end. That is, BIND should verify that a primer which is intended to

be extended by polymerase does in fact bind to the template strand with sufficient stability at the primer's $3'$ end to allow for successful polymerization.

Another extension involves explicitly incorporating a time element into the simulator's calculations, in order to be able to handle more complex reaction types where more than one binding site is possible at the prevailing temperature or competition between primers arises. This dynamic version of BIND would also provide information on how long reactions would need to be run in order to achieve certain predetermined levels of binding and/or primer extension.

Of course, better thermodynamic enthalpy and entropy values would be extremely beneficial in making a more precise version of BIND. In particular, accurate parameters describing mismatched binding would be of great use. Additionally, more elaborate models of the influence of the ionic and chemical environments on deoxyoligonucleotide hybridization would allow BIND to provide accurate melting temperature predictions over a wider range of possible reactions.

Acknowledgements. The authors would like to thank Julia Khodor for her generous assistance in the provision of laboratory gels and experiment protocols. In addition, Hartemink gratefully acknowledges the support of the National Science Foundation and wishes to thank James Wetmur and John SantaLucia, Jr. for their helpful communication.

References

[1] Aboul-ela, F., Koh, D. & Tinoco, I. (1985) *Nucleic Acids Res.* **13**, 4811-4824.
[2] Adleman, L. (1994) *Science* **266**, 1021-1024.
[3] _____(1995) On Constructing a Molecular Computer. Univ. of Southern California, draft, January 8.
[4] Braunlin, W. & Bloomfield, V. (1991) *Biochemistry* **30**, 754-758.
[5] Breslauer, K., Frank, R., Blöcker, H. & Marky, L. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3746-3750.
[6] Cai, W., Condon, A., Corn, R., Glaser, E., Fei, Z., Frutos, T., Guo, Z., Lagally, M., Liu, Q., Smith. L. & Thiel, A. (1996) The Power of Surface-Based DNA Computation. Univ. of Wisconsin, draft, August 13. Available by anonymous ftp from *corninfo.chem.wisc.edu* as */Papers/powerDNA.ps*.
[7] Cantor, C. & Schimmel, P. (1980) *Biophysical Chemistry, Part III: The Behavior of Biological Macromolecules*. W. H. Freeman, San Francisco.
[8] Doktycz, M., Paner, T., Amaratunga, M. & Benight, A. (1990) *Biopolymers* **30**, 829-845.
[9] Quartin, R. & Wetmur, J. (1989) *Biochemistry* **28**, 1040-1047.
[10] Roweis, S., Winfree, E., Burgoyne, R., Chelyapov, N. V., Goodman, M. F., Rothemund, P. W. K. & Adleman, L. (1996) A Sticker Based Model for DNA Computation. *Proceedings of a DIMACS Workshop on DNA Based Computers*, May.
[11] SantaLucia, J., Allawi, H. & Seneviratne, P. A. (1996) *Biochemistry* **35**, 3555-3562.
[12] Werntges, H., Steger, G., Riesner, D. & Fritz, H. (1986) *Nucleic Acids Res.* **14**, 3773-3790.
[13] Wetmur, J. (1991) *Crit. Rev. in Biochem. and Mol. Biol.* **26**, 227-259.
[14] _____(1997) Personal communication.

Massachusetts Institute of Technology, Laboratory for Computer Science, 545 Technology Square, Cambridge, Massachusetts 02139
E-mail address: amink@mit.edu

Massachusetts Institute of Technology, Laboratory for Computer Science, 545 Technology Square, Cambridge, Massachusetts 02139
E-mail address: gifford@mit.edu