Thermodynamic Constraints on DNA-based Computing

R. Deaton and M. Garzon
The Molecular Computing Group
The University of Memphis
Memphis, TN, 38152
rjdeaton@memphis.edu, garzonm@hermes.msci.memphis.edu

September 18, 1998

Abstract

Computing with biological macromolecules, such as DNA, is fundamentally a physical/chemical process. The DNA chemistry introduces a level of complexity that makes reliable, efficient, and scalable computations a challenge. All the chemical and thermodynamic factors have to be analyzed and controlled in order for the molecular algorithm to produce the intended result. For instance, a computation based on DNA requires that the problem instance be encoded in single strands of DNA and that these strands react as planned, that molecular biology protocols, such as PCR or affinity separation, correctly extract the result, and that sufficient flexibility remains so that worthwhile computations can be done. In this paper, various thermodynamic and chemical constraints on DNA computing are enumerated. A similarity measure, based on Gibb's free energy of formation, is defined to judge the goodness of DNA encodings. Finally, the DNA computation problem for implementing molecular algorithms is defined, and it is likely that it is as difficult as the combinatorial optimization problems they are intended to solve.

1 Introduction

To solve hard combinatorial optimization (NP-complete[1]) problems, Adleman[2] introduced a method that utilized single-stranded DNA molecules (oligonucleotides) and techniques from molecular biology[3]. The method, essentially, involves three steps: 1) ENCODING: an encoding that maps a problem instance onto a collection of oligonucleotides, 2) REACTION: template matching reactions, or hybridizations, between the oligonucleotides that do the massively parallel search for a solution, along with ligation, and 3) EXTRACTION: using basic biotech techniques[3] such as polymerase chain reaction (PCR), gel electrophoresis, and affinity separtation, to make the results known.

In Adleman's solution to the Hamiltonian Path Problem (HP)[2], the vertices and edges of the graph are encoded in oligonucleotides. DNA representations of paths through the graph are formed as the vertex oligonucleotides hybridize with the edge oligonucleotides. If the proper hybridizations occur, the DNA representation of the Hamiltonian path can be extracted with affinity separation and PCR. Since Adleman's pioneering work, numerous applications and algorithms have been proposed for DNA based computing[4, 5, 6, 7, 8]. In Adleman's algorithm, the fundamental reaction is hybridization between oligonucleotides that are Watson-Crick complements[2]. Hybridization to target strands on magnetic beads was used to extract the result. Subsequent proposals have continued to rely on the mechanism of hybridization to do the computation and extraction[4, 5, 6, 7, 8]. Other key chemical processes are ligation and PCR.

The complexity in Adleman's algorithm originates in the design of a set of oligonucleotides, both their sequence and concentrations, that will hybridize in preferred alignments, and a set of molecular biology protocols that will extract the desired result. The potential for mishybridizations between oligonucleotides necessitates that sequences be designed of prevent them[9], which in turn, bounds the size of the problem. Increasing the size of the problem requires exponentially increasing amounts of DNA which soon makes implementation impractical[10]. PCR can, also, be error-prone[11], and extraction, likewise, is difficult to do error-free[12]. Each of these steps involves the optimization of chemical and thermodynamic factors in order for the molecular algorithm to function as planned.

In general, molecular algorithms using DNA will use chemical steps which require some optimization in order for the computation to take place. For example, in splicing systems[13, 14], which have also been implemented in the lab[15], the fundamental reaction is enzymatic. Restriction enzymes cut double-stranded DNA molecules at locations of specific base sequence. The molecules are reassembled through hybridization and ligation of the pieces. It has been shown that these mechanisms are capable of producing regular languages over the set of DNA molecules. The chemical difficulty in splicing systems is in matching the optimal reactions conditions of different restriction enzymes, controlling the enzymatic reactions themselves, which are in general incomplete, and in extracting the result.

For any computation using biological macromolecules or enzymes, an issue is the feasibility of achieving the computation within the limits imposed by the thermodynamics and chemistry of the underlying molecular system. Examples of these physical constraints are:

1. Reaction Conditions (*e. g.* time, temperature, solution)
2. Unwanted reactions (interactions) between oligonucleotides, enzymes, or other chemical constituents of the system.
3. Concentrations of the chemical constituents of the system.
4. Fidelity of the molecular biology protocols.

Taking all these into account, along with the original goal of producing a computational result, leads to what might be called the DNA Computation Problem. In what follows, the focus for discussion will be Adleman's molecular algorithm. Chemical and thermodynamic constraints on this algorithm will be discussed. A similarity measure based on the Gibb's free energy of hybrid formation is defined and suggested as an appropriate measure of the goodness of DNA encodings of problem instances. Finally, the DNA computation problem is defined and its computational complexity discussed.

2 Thermodynamic and Chemical Constraints

The pool of DNA oligonucleotides that represent the instance of the problem must meet several conditions. In a hybridization reaction, individual base pairs (bp) hydrogen bond in a highly specific way, with the purine base adenine (A) binding with the pyrimidine thymine (T), and the purine guanine (G) binding with the pyrimidine cytosine (C). These base pairs are the Watson-Crick complements of each other, and ideally, oligonucleotide hybridizations occur only between Watson-Crick complements. Depending upon several factors, however, base pairs that are not Watson-Crick complements can occur in hybridized DNA molecules[16]. The effect of the hybridization reaction conditions on the potential for non-Watson-Crick base pairs in a hybridized molecule is to introduce a possibility of error in both the reaction and extraction steps. In addition, oligonucleotides can shift out of their designed alignments and hybridize. Though these shifted

alignments would not produce a false positive, they do use up oligonucleotide in wasteful reactions, and if enough occur, could produce a false negative.

Factors that influence the stringency of hybridization include the base sequences of the hybridizing oligonucleotides, the location of potential mismatches, the concentrations of the reactant oligos, the temperature of the reaction, the length of the oligonucleotides, and solvent concentrations[17]. These factors can be summarized in the melting temperature parameter, T_m. The melting temperature is the temperature at which 50% of the oligos are melted or single stranded. As reaction temperature increases, an increasing percentage of hybrids melt. For oligos in solution, the melting temperature is given by[18],

$$T_m = \frac{\Delta H^\circ}{\Delta S^\circ + R \ln([C_t]/4)}, \tag{1}$$

where ΔH° is the enthalpy, ΔS° is the entropy, R is the gas constant, and $[C_t]$ is the total molar strand concentration. Melting curves measured by UV absorbance techniques. Single-stranded oligos will absorb UV radiation, and therefore, as the temperature is increased, an increase in absorbance indicates melting. The width of the melting curve for equimolar complementary oligonucleotides is

$$\Delta T = \frac{6RT_m^2}{\Delta H^\circ}. \tag{2}$$

If the reaction temperature for the hybridization, T_r, is greater than $T_m + \Delta T/2$, then, the oligos involved in that reaction will not hybridize, and no errors from that mismatch are possible. Therefore, to prevent unwanted hybridizations, their $T_m + \Delta T/2$ should be less than the reaction temperature, T_r. This condition should hold for all possible binding configurations between oligos that are not wanted. In addition, we want the desired hybridizations to occur, and in sufficient number to enable an efficient and detectable computation. Therefore, for a desired hybridization, its T_m should be greater than the reaction temperature.

In addition, the oligonucleotide pool has to meet criteria associated with the reaction kinetics and thermodynamics of hybridization. As pointed out in [19], because of chemical effects, there is a possibility that certain graphs will be more difficult to solve with DNA than others. The essential reaction in a DNA computation is the hybridization reaction, as expressed by

$$a + b \rightleftharpoons c, \tag{3}$$

where a and b are oligonucleotides, c is the double-stranded hybridization product, and the $\rightleftharpoons$ indicates that the reactions are reversible. Assuming constant volume, the thermodynamic parameter that describes the state of the DNA computer at equilibrium is the Gibb's free energy, G[20]. The change in G for a small change in the chemical components is

$$dG = \sum_{i=1}^{k} \left(\frac{\partial G}{\partial n_i} \right)_{T,P,n_{j \neq i}} dn_i, \tag{4}$$

where n_i is the mole number of the ith of k components in the system. The chemical potential of component i is defined as

$$\mu_i \equiv \left(\frac{\partial G}{\partial n_i} \right)_{T,P,n_{j \neq i}}. \tag{5}$$

Reactions can be exothermic or endothermic. Exothermic reactions produce heat or the capacity for work, $dG < 0$. Endothermic reactions require heat or work input to proceed, and thus, $dG > 0$. The sign of dG determines whether the reaction can be spontaneous or not, and its direction (Table 1). Therefore, dG is the driving force for the reaction, and ultimately, will determined what reaction products are formed, and in

what concentrations they occur. The reaction will be driven towards chemical equilibrium, where the rates to the left and right of $\rightleftharpoons$ in Eq. 3 are equal, and $dG = 0$. The condition for chemical equilibrium is

$$\sum_i \mu_i dn_i = 0. \tag{6}$$

For a general chemical reaction of species, A_i[20],

$$0 \rightleftharpoons \sum_i \nu_i A_i, \tag{7}$$

where the ν_i are the stoichiometric coefficients, which are negative for reactants and positive for products. The change in the amount of a species, A_i, in the reaction is proportional to the ν_i, $dn_i = \nu_i d\xi$, where the constant of proportionally is called the extent of the reaction, ξ. Substitution for dn_i into Eq. 6 produces another condition on chemical equilibrium,

$$\sum_i \nu_i \mu_i = 0. \tag{8}$$

The change in G with the extent of the reaction is

$$\frac{dG}{d\xi} = \sum_i \nu_i \mu_i, \tag{9}$$

which is zero at equilibrium.

ΔG	$A + B \rightleftharpoons AB$
< 0	Reaction proceeds to right
> 0	Reaction proceeds to left
$= 0$	Equilibrium

Table 1: Direction of reaction according to sign of change in free energy.

The free energy change can be written as[20]

$$\frac{dG}{d\xi} = \Delta G^\circ + RT \log Q, \tag{10}$$

where ΔG° is the free energy change under ideally dilute standard conditions, and $RT \log Q$ is a correction term for nonstandard conditions, with T the temperature and R the gas constant. The correction term for the solution of reacting oligos is

$$Q = \prod_i a_i^{\nu_i}, \tag{11}$$

where a_i is the activity of component i. The activity is defined as

$$a_i = \exp\left[\frac{(\mu_i - \mu_i^\circ)}{RT}\right], \tag{12}$$

where μ_i° is the chemical potential in the standard state of an ideally dilute solution[20]. Next, an activity coefficient is defined that is the ratio of the activity to the mole fraction,

$$\gamma_i = a_i / \chi_i, \tag{13}$$

where the mole fraction is $\chi_i = (n_i)/(\sum_i n_i)$, and $\sum_i^P \chi_i = 1$ for P species. Through Eq. 10 and 13, the free energy change is related to the concentrations of the reaction components. At equilibrium, $dG/d\xi = 0$, and the equilibrium constant is

$$K^{eq} = \prod_i (a_{(i,eq)})^{\nu_i} = \exp(-\Delta G^\circ / RT), \tag{14}$$

where $a_{(i,eq)}$ are the equilibrium activities. For an ideally dilute solution, $\gamma_i = 1$, $a_i = \chi_i$, and the thermodynamic parameters are expressed directly in terms of the mole fractions. In what follows, an ideally dilute solution is assumed.

Returning to Eq. 3, the equilibrium constant in terms of the mole fractions of the reactant oligonucleotides and the hybridization product is

$$K^{eq} = \frac{[c]}{[a][b]} = \exp\left(\frac{-\Delta G^\circ}{RT}\right), \tag{15}$$

where [ ] indicates the mole fraction. Therefore, the free energy change for the hybridization of Eq. 3 will determine the concentrations of the product formed, whether it is a desired hybridization product, a mismatched hybridization, or a hybridization with a shifted alignment. For a proper encoding of a problem instance, the free energy change for desired hybridizations should be maximized (more negative), and the free energy change for undesired hybridizations minimized (more positive). In addition, since $\Delta G^\circ = \Delta S^\circ - T\Delta S^\circ$, the melting temperature is very closely related to the free energy, and therefore, the free energy seems a likely candidate for characterizing the strength of a hybridization attraction between two oligonucleotides.

Another factor which can affect the results of a DNA computation is coupling among a set of hybridization reactions. In Adleman's algorithm[2], paths through the graph are represented by DNA molecules that are formed by successive hybridization reactions. The equilibrium constant for a series of hybridizations is proportional to the product of the individual hybridization equilibrium constants. Therefore, unfavorable Reactions can be driven forward by coupling to favorable reactions. For instance, let

$$a + b \rightleftharpoons c, \tag{16}$$

$$c + d \rightleftharpoons e, \tag{17}$$

be a pair of coupled hybridization reactions. Therefore, the total equilibrium constant for the final product e is,

$$K_e^{eq} \propto K_{ab}^{eq} \times K_{cd}^{eq}. \tag{18}$$

This means that mishybridizations can be coupled with favorable hybridization to produce unforseen hybridization products. In addition, in certain graphs, if the molecule representing the Hamiltonian path is less favorable energetically than other paths, then,they will be formed preferentially over the Hamiltonian path. In fact, their relative concentrations will be

$$\frac{[\text{other path}]}{[\text{HP}]} \propto \frac{K^{eq}_{\text{other path}}}{K^{eq}_{\text{HP}}}. \tag{19}$$

Therefore, the encoding must account for these type of effects as well. In particular, relative concentrations of reaction components are potentially important for extraction operations.

Therefore, the requirements on a "good" encoding are:

1. The encoding adequately represents the problem instance.
2. The encoding enables extraction of the result.

3. Designed hybridizations are energetically favorable, while unplanned hybridizations are energetically unfavorable.

4. The encoding takes into account factors related to chemical effects, such as free energy coupling and kinetics, that are produced by the specific problem instance.

A pool of oligonucleotides for DNA computing should fulfill these criteria. Some of these constraints, particularly those associated with the last item, would have to be accounted for by most DNA computing schemes.

3 A Similarity Measure for Encodings

Various criteria have been proposed for the prevention of unwanted hybridization errors by a DNA encoding. In [21, 9], the Hamming distance between oligonucleotides was proposed to preclude the possibility of errors. A new metric[22] which considered errors from shifting oligonucleotides relative to each other has been developed. The problem with these distances is that they fail to capture all the complexity and criteria associated with the DNA chemistry. Therefore, in what follows, it is proposed that the Gibb's free energy is the appropriate measure for the strength of a hybridization attraction. The free energy is the driving force for the reaction, and takes into account or is related to most of the chemical requirements on an encoding. These are the melting temperature, coupled reactions, and equilibrium concentrations of hybridization products.

Similarity measures have been applied to problems in molecular biology[23, 24]. The application that has the most relevance to the encoding problem is that of alignment of two sequences. An alignment is an insertion of spaces in arbitrary locations so that two sequences, which may be of different initial length, are the same final length[23]. For each possible alignment between two sequences, a score is computed based on whether each column in the alignment contains a match, a mismatch, or a space. The similarity between two sequences in the maximum score over all alignments. The number of alignments between two sequences is exponential[24]. Nevertheless, dynamic programming[24] can be used to compute the similarity with quadratic complexity[23]. This alignment problem from molecular biology has much in common with the problem of determining a good encoding for a DNA computation. For an oligonucleotide encoding of a problem, we want to check all possible alignments between the oligonucleotides for hybridization. Unlike the traditional alignment problem, however, to measure the strength of a hybridization potential, we have to measure similarity under Watson-Crick complementation, and incorporate into the similarity measure the thermodynamic cost associated with hybridization of a given alignment.

The similarity measure is defined more precisely as follows[23, 24]. u, t are two sequences over a given alphabet Σ. By inserting spaces in u and t, a pair of new sequences $u^{'}$ and $t^{'}$ is obtained. An alignment $\alpha(u^{'}, t^{'})$ between u and t must satisfy the following properties:

- $|u^{'}| = |t^{'}|$.
- u is obtained by removing all the spaces from $u^{'}$.
- t is obtained by removing all the spaces from $t^{'}$.
- Spaces cannot be stacked atop one another.

Next, we define an additive scoring system for an alignment. Typically, the scoring system, (p, g), consists of a function $p : \Sigma \times \Sigma \rightarrow \mathbf{R}$, which assigns a score to each pair of symbols in an alignment, and a space

penalty, g (typically < 0) which penalizes spaces. The similarity, then, is

$$s(u,t) = \max_{\alpha \in \mathcal{A}(u,t)} \text{score}(\alpha), \tag{20}$$

where $\mathcal{A}(u,t)$ is the set of all possible alignments.

The primary energetic factor for hybridization is not the energy of the hydrogen bonding between nucleotide bases, but is the nearest neighbor base stacking energies[18]. These base stacking energies must be measured, and are not unique. Nevertheless, from an energetic point of view, they are the parameters of choice to determine the potential for hybridiztion between oligonucleotides. Recently measured values for these parameters are given in Table 2[25].

Sequence	ΔH° (kcal/mol)	ΔS° (eu)	ΔG°_{37} (kcal/mol)
AA/TT	-8.4	-23.6	-1.02
AT/TA	-6.5	-18.8	-0.73
TA/AT	-6.3	-18.5	-0.60
CA/GT	-7.4	-19.3	-1.38
GT/CA	-8.6	-23.0	-1.43
CT/GA	-6.1	-16.1	-1.16
GA/CT	-7.7	-20.3	-1.46
CG/GC	-10.1	-25.5	-2.09
GC/CG	-11.1	-28.4	-2.28
GG/CC	-6.7	-15.6	-1.77

Table 2: Thermodynamic parameters for DNA helix initiation and propagation in 1 M NaCl. Sequences are given $5' \rightarrow 3'/3' \rightarrow 5'$. [25].

To define an appropriate similarity measure for hybridization, we take as our alphabet all possible nearest neighbor pairs of nucleotide bases. Then, the cost measure is associated, in the case of perfect Watson-Crick complements, with the thermodynamic parameters (ΔG°) of Table 2. Mismatched base pairs could also be assigned a thermodynamic cost[17, 26, 27]. Spaces at the end of strings would be associated with the thermodynamic parameters associated with dangling ends[17, 26]. Spaces in the middle of the string would be bulge loops, with an associated thermodynamic penalty[17, 26] All the factors are added to produce a total free energy of hybridization for a particular alignment. Therefore, with this similarity measure, each alignment would have a cost associated with it, this cost would be the Gibb's free energy of formation at a standard temperature, and the similarity would be the maximum free energy, and therefore, would represent the most energetically favorable alignment for hybridization.

Typically[24], a similarity measure has the following properties:

$$p(a,a) > 0 \forall\ a \in \Sigma \tag{21}$$

$$p(a,b) < 0 \text{ for some } a, b \in \Sigma. \tag{22}$$

The purpose is to reward similar symbols from the alphabet, and penalize different symbols from the alphabet. For the encoding, however, the purpose is to reward Watson-Crick complements and penalize non-Watson-Crick complement sequences. Therefore, the proposed cost operator between two symbols is

$$p(a, b^r) = \Delta G^\circ_{a/b^r}, \tag{23}$$

where b^r indicates the reverse Watson-Crick complement.

Another advantage of the similarity measure defined above is that it is related to a distance metric[23, 24]. The distance, $d(u,t)$, and similarity, $s(u,t)$ are related by

$$s(u,t) + d(u,t) = \frac{M}{2}(m+n), \tag{24}$$

where u and t are sequences of length m and n, respectively, and M is an arbitrary constant.

4 The DNA Computation Problem

The goal of DNA computing is to implement algorithms in reactions among biological molecules. In so doing, the hope is to tap the generative power that is evident in the biological machinery of life. This is a thread that runs from cellular automata[28, 29] to genetic algorithms[30] and artificial neural networks[31]. At the very least, one would want to utilize the massive number of molecules for quicker solution of very difficult problems. In a DNA computer, a molecular solution will be reached. The important question is if that solution *in vitro* corresponds to the desired algorithmic solution. A similar situation exists when algorithmic descriptions of certain natural phenomena are attempted. For instance, computation of the folding of proteins[32], or the minimum energy state of a spin glass[33] are known to be NP-complete. Nevertheless, the actual protein folds in a fraction of the time that an algorithm would require to compute it. Therefore, given the chemical and thermodynamic factors that influence and constrain a DNA computation, a question occurs about how difficult it really is to implement algorithms in molecular systems. So far, it seems that the molecular implementation has differed for each problem[5, 6, 8], and when one says "molecular implementation," the actual laboratory procedures that produced a valid solution are meant, not the theoretical molecular algorithms. Therefore, in this section, the complexity of implementing algorithms in biomolecular systems is discussed. The process of converting an algorithm into a biomolecular systems is called the DNA Computation Problem.

DNA Computation Problem (DCP)

Instance 1 *A DNA computer $D = (S, M, f)$, where S and M are finite sets of DNA molecules, and $f : S \to M$ is a mapping that represents a set of biomolecular operations (enzymatic reactions, PCR, gel electrophoresis, etc ...), and an algorithm, A, problem Π, and problem instance I.*

Question 1 *Is there a D that implements the application of A to I?*

According to Church's thesis, there are any number of equivalent approaches to the notion of an algorithm, or finite procedure, *i. e.* Turing Machines, λ-calculus. For the current discussion, the model is restricted to finite automata. While not as powerful as Turing Machines *et al.*, DNA implementations of finite automaton have been based on both Adleman's algorithm and splicing systems[13, 34, 35, 36].

As part of the DNA implementation of the finite automata, it would have to be determined if the finite automata and its DNA implementation recognized the same languages. Determining the inequivalence of finite state automata is a problem that is known to be NP-hard[1]. Therefore, DCP contains as a subset NP-hard problems, and therefore, is itself NP-hard.

Other aspects of DCP could involve difficult problems. Most significantly, the set of molecules M represent the reaction products produced by the chemical steps represented by f. Many chemical and physical interactions are possible between the chemical constituents of a DNA computer, as outlined above. Because of this, a DNA computer is similar to a frustrated physical system, like a spin glass[37]. Implementation

of an algorithm might require that the minimum energy state be computed for the collection of chemical constituents, and if the DNA computer is a frustrated system, this problem would more than likely be NP-complete, as it is for spin glasses[33] and protein folding[32].

5 Conclusion

Given the difficulty of implementing DNA algorithms, what are the prospects for DNA based computing. Three possibilities present themselves: 1) Applications that do not require encodings have been suggested[38]. These applications might involve solution of problems in molecular biology and biotechnology with DNA computations. 2) DNA computing can continue down the track it has followed till now, and rely upon less than optimum implementations. It could be that implementations that are good enough for specific applications could be found with less effort. 3) Applications, such as artificial immune systems, evolutionary programs, and associative memories could be implemented in a DNA computer. These applications take advantage of the fuzziness of DNA chemistry to produce variation, fault tolerance, pattern recognition, and associative capabilities into the computation.

Adleman[2] developed a DNA based technique for the solution of a NP-complete problem. The massive parallelism of DNA hybridizations evoked a hope that these hard problems would be tractable in a DNA based computer. Implementing DNA based algorithms, however, has turned out to be very difficult, and there is evidence that it is just as hard as the original problems that they were intended to solve. The difficulty in the implementation originates in checking and controlling the many constraints imposed by the DNA chemistry. Failure to do so can produce molecular algorithms whose results are not those intended.

References

[1] M. R. Garey and D. S. Johnson, *Computers and Intractability.* New York: Freeman, 1979.

[2] L. M. Adleman, "Molecular computation of solutions to combinatorial problems," *Science*, vol. 266, pp. 1021–1024, 1994.

[3] J. D. Watson, N. H. Hopkins, J. W. Roberts, J. A. Steitz, and A. M. Weiner, *Molecular Biology of the Gene.* Menlo Park, CA: The Benjamin/Cummings Publishing Co., Inc, fourth ed., 1987.

[4] R. J. Lipton, "DNA solution of hard computational problems," *Science*, vol. 268, pp. 542–545, 1995.

[5] DIMACS, *Proceedings of the First Annual Meeting on DNA Based Computers*, (Providence, RI), American Mathematical Society, 1996. DIMACS Proc. Series No. 27.

[6] DIMACS, *Preliminary Proceedings of the Second Annual Meeting on DNA Based Computers*, (Providence, RI), American Mathematical Society, 1997. DIMACS Proc. Series.

[7] F. Guarnieri, M. Fliss, and C. Bancroft, "Making DNA add," *Science*, vol. 273, pp. 220–223, 1996.

[8] DIMACS, *Preliminary Proceedings of the 3rd DIMACS Workshop on DNA Based Computers*, (Providence, RI), American Mathematical Society, 1997. Philadelphia, PA., June 23-27.

[9] R. Deaton, M. Garzon, J. A. Rose, D. R. Franceschetti, R. C. Murphy, and S. E. S. Jr., "Reliability and efficiency of a DNA based computation," *Phys. Rev. Lett.*, vol. 80, pp. 417–420, 1998.

[10] J. Hartmanis, "On the weight of computations," *Bulletin of the European Association for Theoretical Computer Science*, vol. 55, pp. 136–138, 1995.

[11] D. Boneh, C. Dunworth, J. Sgall, and R. J. Lipton, "Making DNA computers error resistant," in *Preliminary Proceedings of the Second Annual Meeting on DNA Based Computers* [6], pp. 102–110. DIMACS Proc. Series.

[12] R. Karp, C. Kenyon, and O. Waarts, "Error-resilient DNA computation," in *Proceedings of the 7th ACM-SIAM symposium on discrete algorithms*, pp. 458–467, ACM Press/SIAM, 1996.

[13] T. Head, "Formal language theory and DNA: An analysis of the generative capacity of specific recombination behaviors," *Bull. Math. Biology*, vol. 49, pp. 737–759, 1987.

[14] G. Paun, G. Rozenberg, and A. Salomaa, "Computing by splicing," *Theoretical Computer Science*, vol. 168, pp. 321–336, 1996.

[15] E. Laun and K. J. Reddy, "Wet splicing systems," in *Preliminary Proceedings of the 3rd DIMACS Workshop on DNA Based Computers* [8], pp. 115–126. Philadelphia, PA., June 23-27.

[16] J. Sambrook, E. F. Fritsch, and T. Maniatis, *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, second ed., 1989.

[17] J. G. Wetmur, "Physical chemistry of nucleic acid hybridization," in *Preliminary Proceedings of the 3rd DIMACS Workshop on DNA Based Computers* [8], pp. 1–25. Philadelphia, PA., June 23-27.

[18] P. N. Borer, B. Dengler, I. Tinoco, Jr., and O. C. Uhlenbeck, "Stability of ribonucleic acid double-stranded helices," *J. Mol. Biol.*, vol. 86, pp. 843–853, 1974.

[19] S. A. Kurtz, S. R. Mahaney, J. S. Royer, and J. Simon, "Active transport in biological computing," in *Preliminary Proceedings of the Second Annual Meeting on DNA Based Computers* [6], pp. 111–122. DIMACS Proc. Series.

[20] I. N. Levine, *Physical Chemistry*. New York: McGraw–Hill Book Company, Inc., fourth ed., 1995.

[21] R. Deaton, R. C. Murphy, M. Garzon, D. R. Franceschetti, and S. E. Stevens Jr., "Good encodings for DNA-based solutions to combinatorial problems," in *Preliminary Proceedings of the Second Annual Meeting on DNA Based Computers* [6], pp. 159–171. DIMACS Proc. Series.

[22] M. Garzon, R. Deaton, P. Neathery, R. C. Murphy, S. E. Stevens Jr., and D. R. Franceschetti, "A new metric for DNA computing," in *Genetic Programming 1997: Proceedings of the Second Annual Conference*, pp. 479–490, AAAI, 1997. Stanford University, July 13-16, 1997.

[23] J. Setubal and J. Meidanis, *Introduction to Computational Molecular Biology*. Boston: PWS Publishing Co., 1997.

[24] M. S. Waterman, *Introduction to Computational Biology*. New York: Chapman and Hall, first ed., 1995.

[25] J. SantaLucia, Jr., H. T. Allawi, and P. A. Seneviratne, "Improved nearest-neighbor parameters for predicting DNA duplex stability," *Biochemistry*, vol. 35, pp. 3555–3562, 1996.

[26] C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry: Part III The Behavior of Biological Macromolecules*. New York: W. H. Freeman and Company, 1980.

[27] H. Werntges, G. Steger, D. Riesner, and H. J. Fritz, "Mismatches in DNA double strands: Thermodynamic parameters and their correlation to repair efficiencies," *Nucleic Acids Res.*, vol. 14, pp. 3773–3790, 1986.

[28] J. von Neumann, *Theory of Self-Reproducing Automata.* Urbana, IL: University of Illinois Press, third ed., 1966.

[29] S. Ulam, "On some mathematical problems connected with patterns of growth of figures," in *Essays on cellular automata* (A. E. Burks, ed.), pp. 219–231, Chicago: U. of Illinois Press, 1972.

[30] J. H. Holland, *Adaptation in Natural and Artificial Systems.* Cambridge, MA: MIT Press, 1992.

[31] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Sciences of the U.S.A.*, vol. 89, pp. 2554–2558, 1982.

[32] A. S. Fraenkel, "Complexity of protein folding," *Bull. Math. Biology*, vol. 55, pp. 1199–1210, 1993.

[33] F. Barahona, "On the computational complexity of Ising spin glass models," *J. Phys. A*, vol. 15, pp. 3241–3253, 1982.

[34] R. Freund, G. Paun, G. Rozenberg, and A. Salomaa, "Watson-Crick finite automata," in *Preliminary Proceedings of the 3rd DIMACS Workshop on DNA Based Computers* [8], pp. 305–317. Philadelphia, PA., June 23-27.

[35] Y. Gao, M. Garzon, R. C. Murphy, J. A. Rose, R. Deaton, D. R. Franceschetti, and S. E. Stevens Jr., "DNA implementation of nondeterminism," in *Preliminary Proceedings of the 3rd DIMACS Workshop on DNA Based Computers.*

[36] M. Garzon, Y. Gao, J. A. Rose, R. C. Murphy, R. Deaton, D. R. Franceschetti, and S. E. Stevens Jr., "In-vitro implementation of finite-state machines," in *Proc. 2nd Intl. Workshop on Implementing Automata WIA'97*, (Berlin), Springer-Verlag, 1998. Lecture Notes in Computer Science.

[37] M. Mezard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond.* Singapore: World Scientific, 1987.

[38] L. Landweber and R. J. Lipton, "DNA^{2}DNA computations: A potential "Killer App"," in *Preliminary Proceedings of the 3rd DIMACS Workshop on DNA Based Computers* [8], pp. 59–68. Philadelphia, PA., June 23-27.