Paradigms for Computational Nucleic Acid Design

Robert M. Dirks[1], Milo Lin[2], Erik Winfree[3] and Niles A. Pierce[4]

Keywords: DNA, RNA, secondary structure, positive design, negative design, affinity, specificity, partition function

Abstract

The design of DNA and RNA sequences is critical for many endeavors, from DNA nanotechnology, to PCR-based applications, to DNA hybridization arrays. Results in the literature rely on a wide variety of design criteria adapted to the particular requirements of each application. Using an extensively-studied thermodynamic model, we perform a detailed study of several criteria for designing sequences intended to adopt a target secondary structure [1]. We conclude that superior design methods should explicitly implement both a positive design paradigm (optimize affinity for the target structure) and a negative design paradigm (optimize specificity for the target structure). The commonly used approaches of sequence symmetry minimization and minimum free energy satisfaction primarily implement negative design and can be strengthened by introducing a positive design component. Surprisingly, our findings hold for a wide range of secondary structures and are robust to modest perturbation of the thermodynamic parameters used for evaluating sequence quality, suggesting the feasibility and ongoing utility of a unified approach to nucleic acid design as parameter sets are further refined. Finally, we observe that designing for thermodynamic stability does not determine folding kinetics, emphasizing the opportunity for extending design criteria to target kinetic features of the energy landscape.

Introduction

A fundamental design problem consists of selecting the sequence of a nucleic acid strand that will adopt a target secondary structure [7]. As depicted in Figure 1a, this is the inverse of the more famous folding problem of determining the structure (and folding mechanism) for a given sequence. To attempt the rational design of novel nucleic acid structures, we require both an approximate empirical physical model [6, 4] and a search algorithm for selecting promising sequences based on this model. Experimental feedback on the quality of the design and the performance of the design algorithm can then be obtained by folding the molecule *in vitro*. Alternatively, if this feedback loop can be closed computationally by folding the molecule *in silico*, the quality of sequence designs could be rapidly assessed and improved before attempting laboratory validation.

In designing nucleic acid sequences, we consider the two principal paradigms illustrated in Figure 1b. *Positive design* methods attempt to select for a desired outcome by optimizing sequence affinity for the target structure. *Negative design* methods attempt to select against unwanted outcomes by optimizing sequence specificity for the target structure. A successful design must exhibit both high affinity and high specificity [8], so useful design algorithms must satisfy the objectives of both paradigms, even if they explicitly implement only one.

[1]Chemistry, Caltech. E-mail: dirks@caltech.edu
[2]Undergraduate, Caltech. E-mail: miloiq@its.caltech.edu
[3]Computer Science and Computation & Neural Systems, Caltech. E-mail: winfree@caltech.edu
[4]Applied & Computational Mathematics, Bioengineering, Caltech. E-mail: niles@caltech.edu

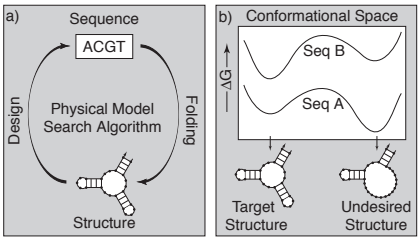

Figure 1: a) Feedback loop for evaluating nucleic acid sequence designs and methodologies. b) Positive and negative design paradigms. Two sequences are evaluated using an empirical potential on both the desired target structure and an undesired structure. Using a positive design paradigm, sequence A would be selected since it exhibits a stronger affinity than sequence B for the target structure (i.e. lower ΔG). Using a negative design paradigm, sequence B would be selected since it exhibits specificity for the target structure while sequence A exhibits specificity for the undesired structure.

For some applications, it may be desirable to supplement these thermodynamic design considerations with additional kinetic requirements. For example, in designing molecular machines [10], it may be crucial to select sequences that fold or assemble quickly. Alternatively, it may be important to design interactions with intentionally frustrated folding kinetics in order to control fuel delivery during the work cycle [9].

The present study uses efficient partition function algorithms [5, 2] and stochastic kinetics simulations [3] to examine the thermodynamic and kinetic properties of sequences designed using seven methods that capture aspects of the positive and negative design paradigms. Although several of these design criteria have been widely used, we are not aware of any previous attempt to assess their relative performance. Evaluated based on thermodynamic considerations, we consistently observe that sequence selection methods that implement both positive and negative design paradigms outperform methods that implement either paradigm alone. This trend appears to be robust to changes in both the target secondary structure and the parameters in the physical model, and to the choice of either RNA or DNA as the design material. The trend does not hold when the design criteria are judged based on kinetic considerations, as favorable thermodynamic properties do not ensure fast folding.

References

[1] R.M. Dirks, M. Lin, E. Winfree, and N. A. Pierce. Paradigms for computational nucleic acid design. *Nucleic Acids Res.*, 2004. In press.

[2] R.M. Dirks and N. A. Pierce. A partition function algorithm for nucleic acid secondary structure including pseudoknots. *J. Comput. Chem.*, 24:1664–1677, 2003.

[3] C. Flamm, W. Fontana, I.L. Hofacker, and P. Schuster. RNA folding at elementary step resolution. *RNA*, 6:325–338, 2000.

[4] D.H. Mathews, J. Sabina, M. Zuker, and D.H. Turner. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, 288:911–940, 1999.

[5] J.S. McCaskill. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, 29:1105–1119, 1990.

[6] J. SantaLucia, Jr. Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, 35:3555–3562, 1996.

[7] N. C. Seeman. Nucleic acid junctions and lattices. *J. Theor. Biol.*, 99:237–247, 1982.

[8] N. C. Seeman and R.K. Kallenbach. Design of immobile nucleic acid junctions. *Biophys. J.*, 44:201–209, 1983.

[9] A. J. Turberfield, J.C. Mitchell, B. Yurke, Jr. Mills, A. P., M.I. Blakey, and F.C. Simmel. DNA fuel for free-running nanomachines. *Phys. Rev. Lett.*, 90(11):118102, 2003.

[10] B. Yurke, A.J. Turberfield, Jr. Mills, A.P., F.C. Simmel, and J.L. Neumann. A DNA-fuelled molecular machine made of DNA. *Nature*, 406:605–608, 2000.