

- and D.J. Lockhart. 1999. High density synthetic oligonucleotide arrays. *Nat. Genet. Supp.* 1:20-24.
17. Malicka, J., I. Gryczynski, J. Fang, J. Kusba, and J.R. Lakowicz. Photostability of Cy3 and Cy5-labeled DNA in the presence of metallic silver particles. *J. Fluorescence* (In Press.)
18. Ni, F. and T.M. Cotton. 1986. Chemical procedure for preparing surface-enhanced Raman scattering active silver films. *Anal. Chem.* 58:3159-3163.
19. Nie, S. and S.R. Emory. 1997. Probing single molecules and single nanoparticles by surface-enhanced Raman scattering. *Science* 275:1102-1106.
20. Soper, S.A., H.L. Nutter, R.A. Keller, L.M. Davis, and E.B. Shera. 1993. The photophysical constants of several fluorescent dyes pertaining to ultrasensitive fluorescence spectroscopy. *Photochem. Photobiol.* 57:972-977.
21. Vo-Dinh, T. 1998. Surface-enhanced Raman spectroscopy using metallic nanostructures. *Trends Anal. Chem.* 17:557-582.
22. Wokaun, A. 1984. Surface-enhanced electromagnetic processes, pp 330. In H. Ehrenreich and D. Turnbull (Eds.), *Solid State Physics: Advances in Research and Application*. Vol. 38. Academic Press, New York.

This work was supported by the National Institutes of Health grant no. EB-00682 and the National Center for Research Resources grant no. RR-08119 to J.R.L. Address correspondence to Dr. Joseph Lakowicz, University of Maryland School of Medicine, Center for Fluorescence Spectroscopy, Department of Biochemistry and Molecular Biology, 725 West Lombard Street, Baltimore, MD 21201, USA. e-mail: lakowicz@cfs.umbi.umd.edu

Received 21 June 2002; accepted 16 October 2002.

**Joseph R. Lakowicz, Joanna Malicka, and Ignacy Gryczynski**

*University of Maryland School of Medicine  
Baltimore, MD, USA*

For reprints of this or any other article, contact [Reprints@BioTechniques.com](mailto:Reprints@BioTechniques.com)

## Molecular Beacon Sequence Design Algorithm

*BioTechniques* 34:68-73 (January 2003)

### ABSTRACT

*A method based on Web-based tools is presented to design optimally functioning molecular beacons. Molecular beacons, fluorogenic hybridization probes, are a powerful tool for the rapid and specific detection of a particular nucleic acid sequence. However, their synthesis costs can be considerable. Since molecular beacon performance is based on its sequence, it is imperative to rationally design an optimal sequence before synthesis. The algorithm presented here uses simple Microsoft® Excel® formulas and macros to rank candidate sequences. This analysis is carried out using mfold structural predictions along with other free Web-based tools. For smaller laboratories where molecular beacons are not the focus of research, the public domain algorithm described here may be usefully employed to aid in molecular beacon design.*

### INTRODUCTION

This report describes a method to aid the design of molecular beacon sequences. Molecular beacons are a new optical tool that can be used to detect the presence of a specific DNA sequence in a mixture of targets. First demonstrated by Tyagi and Kramer (4), molecular beacons are fluorogenic oligonucleotide probes that signal hybridization with a complementary nucleic acid target (3). These DNA oligonucleotides contain a 5' fluorophore, a 3' quenching group, and 4-6 complementary bases on the 3' and 5' "stem" ends so that, unless the inner "loop" region hybridizes to a complementary nucleic acid, the fluorescence of the beacon in its hairpin configuration is quenched (Figure 1). When hybridized with a complementary oligonucleotide, the hairpin structure linearizes, distancing the fluorophore and quencher to yield fluorescence. The usefulness of molecular beacons to signal a hybridization event depends on

their molecular structure. Therefore, each molecular beacon must be designed such that its sequence confers the correct conformation in the hairpin unhybridized state and the linear form when hybridized to the intended target.

While designing molecular beacons that function properly is more complex than the design of simple linear probes, the advantages of molecular beacons outweigh the difficulties in their proper design. Traditionally, linear oligonucleotide probes used to identify hybridization of a specific nucleic acid sequence are tagged with fluorescent or radioactive labels and hybridized to a sample. Unhybridized probes are washed away if the sample is anchored to a solid surface (e.g., blot or tissue), or they are digested if the target is in solution (e.g., PCR or FISH). Remaining probe is assumed to be hybridized to the target and is quantified. The post-processing steps in these methods can alter the true conditions of hybridization and thus prevent the technique from being used in real-time and in vivo.

An essential aspect of molecular beacons is that the beacon cannot have a target hybridized in the loop region and still maintain its duplex conformation in the stem region. The rigidity of the DNA helix prevents beacon-target hybrids from coexisting with stem hybrids. If designed correctly, then a perfectly matched beacon-target hybrid will be energetically more stable than the beacon hairpin conformation. In addition, a mismatched beacon-target hybrid will not open the hairpin of the stem because of its lower energetic stability. These features impart a high level of specificity for the exact target sequence.

In addition to specificity for the desired target sequence, optimal design of molecular beacons dictates that fluorescence quenching of the beacon in the absence of a target should only occur in the correct hairpin conformation so that the 5'- and 3'-ends are in close proximity. While the sequence of the loop region is constrained to be complementary to the desired nucleic acid target, the stem region sequence can contain any possible combination of nucleotides. Therefore, it is useful to predict what conformations are most likely for a molecular beacon of a given sequence.

While one could in theory synthe-

# Short Technical Reports

size a large number of molecular beacons with different stem sequences and then test each one for effectiveness, the current costs associated with synthesizing the oligonucleotide with attached quencher and fluorophore makes this unpractical. On average, commercial

synthesis of a molecular beacon on the 50-nM scale currently costs \$400 (companies polled included Integrated DNA Technologies, Biosearch Technologies, Stratagene, Genset, TriLink Biotechnologies, and Sigma Genosys). As such, it is imperative to design a

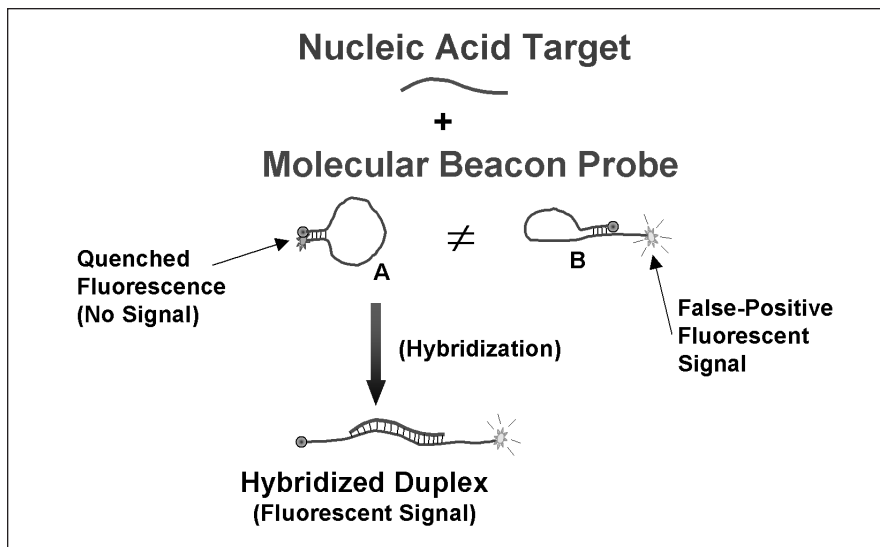
beacon that will yield optimal results before synthesis. An algorithm that can predict which sequences will be most suitable for use as beacons minimizes the costs associated with synthesizing and testing many beacons.

## MATERIALS AND METHODS

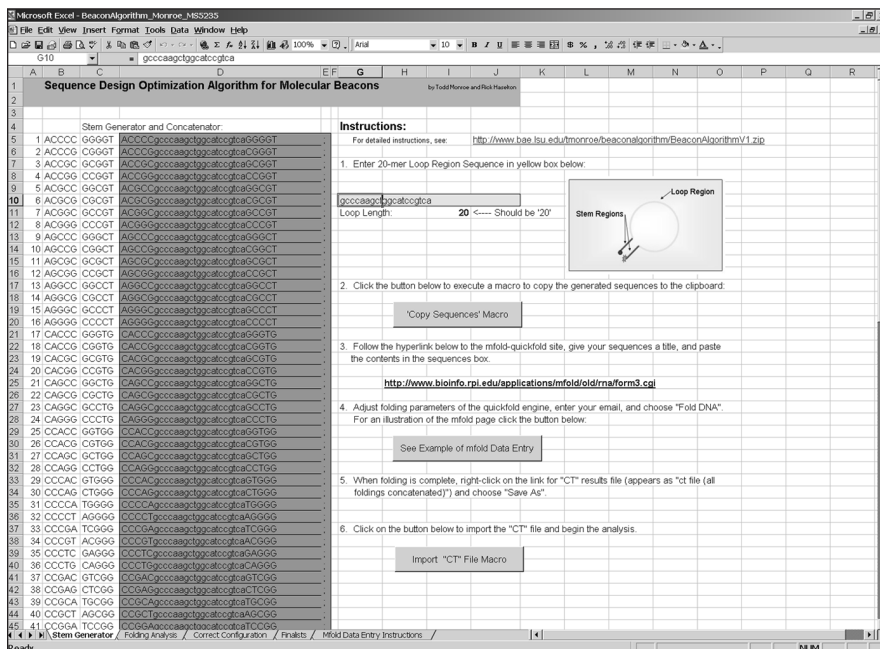
The design method proposed here determines which sequences are most likely to assume the desirable hairpin conformation for optical molecular beacon effectiveness. The algorithm embodies four components: a Microsoft® Excel® spreadsheet to generate candidate sequences based on user input, a Web site server to establish the stability of candidates, an Excel spreadsheet to rank sequences according to their suitability, and a Web site to check for dimerization. A compressed copy of the Excel example spreadsheet can be obtained from <http://www.bae.lsu.edu/tmonroe/beaconalgorithm/Beacon-AlgorithmV1.zip>. The visual basic macros contained within the Excel file have been tested for compatibility on PC-based versions of Microsoft Excel 97, 2000, and XP. As described below, the other two components are readily accessible online.

## RESULTS AND DISCUSSION

The first step in the selection process is the generation of all possible candidate sequences. This is done in the "Stem Generator" sheet of a Microsoft Excel file (Figure 2). The algorithm generates all possible stem sequences for a stem length of five bases, then concatenates them with the inner loop sequence, and formats the data set to be sent to *mfold*, a DNA folding program. *mfold* is a public domain, Web-based RNA/DNA structural prediction package designed and maintained by Dr. Michael Zuker (<http://www.bioinfo.rpi.edu/~zukerm/>). Using the thermodynamic parameters established by SantaLucia (2), the *mfold* program predicts the most likely secondary structures for an entered DNA sequence. *mfold* alone could be used to predict the structure of a single sequence, but the optimization of multiple sequences is



**Figure 1.** Schematic diagram that illustrates a hybridization assay of a nucleic acid target and complementary molecular beacon. Properly designed molecular beacons (A) in the stem-loop configuration should have little fluorescent signal. When hybridized to a complementary nucleic acid target, the beacon linearizes, distancing the quencher from the fluorophore and generating a fluorescent signal. Molecular beacons with suboptimal stem sequences can self-hybridize in conformations other than the desired stem-loop structure and give false-positive signals (B).



**Figure 2.** Screenshot of the Excel spreadsheet used in this algorithm to design optimal molecular beacon sequences. Using macros and formulas, candidate stem sequences are generated, which are then sent to the *mfold* server for folding analysis. Sequences are then ranked according to proper conformations for optimal molecular beacon performance.

# Short Technical Reports

currently a trial-and-error process. Once the *mfold* calculations are complete, our algorithm ranks the sequences according to their desirability as molecular beacons.

The first step in beacon design is to determine the “loop region” sequence of the molecular beacon. This typically 20-mer sequence will be complementary to the target of interest. While the Excel file presented here does not specifically design this portion of the beacon, this step can be readily accomplished by using a number of Web-based and freeware tools, such as Primer3<sup>®</sup> (1). These packages will determine the optimal sequence of a hybridization probe for an entire gene or other sequence of interest. The molecular beacon stem sequences that flank the loop region are then determined in the Excel file. Most molecular beacons have stem regions ranging from four to seven bases in length. Given that each nucleotide can have one of four bases, a stem *n* bases long could have  $n^4$  different sequences. We used the rules suggested by Tyagi and Kramer (4) to narrow the number of possibilities. The sequence of the stem should be designed such that the melting temperature is 7°C–10°C above the annealing temperature of the target and its complementary region (e.g., annealing temperature of primers in PCR). For most applications, the G–C content of the stem should be 70%–80% (i.e., if stem length is five bases, then four bases should be either G or C). Sequences with a 5′ guanine should be avoided, since guanine may quench the attached fluorophore (3).

Our algorithm generates all possible 1024 sequences for a beacon with five stem-forming bases on the 5′-end and also generates the complementary bases on the 3′-end that will form the stem. However, applying the Tyagi rules mentioned previously, this number of candidate sequences is reduced to 96. Using simple formulas in an Excel spreadsheet, the algorithm then concatenates these 5′ and 3′ sequences with the entered loop region sequence and generates a dataset in the format to be analyzed further.

This dataset is then copied and pasted into a form at the DNA *quick-fold* server, <http://www.bioinfo.rpi.edu/>

[applications/mfold/old/rna/form3.cgi](http://www.bioinfo.rpi.edu/applications/mfold/old/rna/form3.cgi), which can accept batch entries, as long as the sequences are less than 200 bases in length. The *mfold* engine then calculates the most likely configurations for each sequence and estimates the negative free energy for each possible configuration. The results can be obtained in several formats, both graphical and ASCII text formats. The graphical formats are striking, but analysis of a large number of sequences by image analysis can be tedious. We have found that the ASCII files generated are easier to manipulate for optimizing the best candidate sequences. The ASCII datasets (“CT” file in which all results are concatenated) are readily imported into the Excel spreadsheet using a macro and filtered so that only configurations with the correct stem alignment in the hybridized state remain. As seen in Figure 3, the records of the “CT” file can be used to determine which sequences have bases hybridized in the correct configurations (5).

Using simple, editable formulas within Excel, candidate hairpin sequences are evaluated for the number of the hybridized bases and that the correct bases are hybridized. For an *n*-base beacon with a five-nucleotide stem length, only five bases should hybridize in the correct hairpin. In addition, base #1 and base #*n* should hybridize, as should #5 and #*n*-4. In this approach, an additional stipulation that we em-

ploy is that a given sequence should have only one possible conformation. If a beacon is synthesized that has several possible conformations, then the likelihood of mixtures of correct and incorrect hairpins increases, leading to false-positive signals from the beacons in the absence of hybridization targets. If a candidate sequence returns several possible conformations, then it is omitted from the optimal sequence list to ensure that the only likely conformation is the correct hairpin. It is important to note that the user can easily modify the rules or add additional stipulations should a particular hybridization application require a higher stringency.

Once all sequences that did not meet these requirements are filtered out, the remaining sequences are ranked by stem stability based on the negative free energy obtained from *mfold*. While this ranking should predict the stability of the correct stem-loop conformation, it does not account for the possibility that the beacons can form dimers with each other. A Web-based application for designing PCR primers was used to analyze the most promising candidate sequences from the previous step for their dimer conformations. The top 10 candidate sequences with the lowest free energies from the previous ranking were individually analyzed using Integrated DNA Technologies’ Web-based *Oligo Analyzer 2.5*<sup>®</sup> (<http://www.idtdna.com/program/oligocalc/oligocalc.asp>). In ad-

Calculated dG				Sequence Name_Number	
# of bases					
	30	dG =	-3.9	BeaconA_1	
1	A	0	2	30	1
2	C	1	3	29	2
3	C	2	4	28	3
4	C	3	5	27	4
5	C	4	6	26	5
6	G	5	7	17	6
7	C	6	8	16	7
8	C	7	9	15	8
9	C	8	10	0	9
10	A	9	11	0	10
(a)	(b)	(c)	(d)	(e)	(f)

**Figure 3.** Example of the ASCII “CT” file output from *mfold* for the folding of a typical molecular beacon sequence. The first record in this text file displays the sequence length, predicted free energy change, and sequence name. The subsequent records contain (a) the index, (b) the base of that index, (c) the index of the 5′-connecting base, (d) the index of the 3′-connecting base, (e) the index of the hybridized base, and (f) the historical numbering of the base index in the original sequence. When there is no hybridization between given bases, a zero is returned for (e) (5).

dition to the negative free energies for hairpin conformations, this applet calculates free energies for the self-dimerization of two oligonucleotide sequences.

While kinetics suggest that unimolecular "hairpinning" hybridization is more likely than bimolecular (dimerization) hybridization, it is interesting to look at the thermodynamic predictions of a dimerization versus hairpin calculation. Using negative free energy values, the analysis of beacon dimer pairs could be important, since the dimers often form more stable conformations, which are 5–10 kcal/mol lower than the most stable hairpins for a given sequence. Dimerization of two beacons leading to overhanging ends can position the quencher too far from the fluorophore of the hybridized beacon and thus give a false-positive signal. Therefore, if dimerization is likely, it is important that the dimer conformation favor alignment without overhangs. Two molecular beacons with the same loop sequence but different stem sequences were analyzed with the algorithm described here. These sequences had similar hairpin negative free energies from *mfold* but were analyzed further for dimerization. The dimer conformation of most negative free energy for sequence 1 (5'-ACGCG-GCCCAAGCTGGCATCCGTCACGC-GT-3') is an overhanging end with hybridization between the two strands at the underlined sequence. This conformation would distance the 5' fluorophore from the 3' quencher in both strands and give a false-positive signal. In comparison, the dimer confirmation of most negative free energy for sequence 2 (5'-CGTGCGCCCAAGCTG-GCATCCGTC-GCACG-3') predicts the two strands hybridized at both ends (underlined). This conformation places the 5' fluorophore of one strand in close proximity to the 3' quencher of the opposing strand, yielding an appropriate stem hybridization and quenched fluorescent signal. Because the most likely dimer conformation for sequence 2 predicts an appropriate stem hybridization, sequence 2 would be a better choice than sequence 1 for optimal beacon performance.

Identification of sequences with the most stable hairpin, only one hairpin conformation, and the correct dimer conformation substantially narrows the

number of sequences to be synthesized and further evaluated. We have used this approach to successfully design two different molecular beacons. The validity of this approach for applications where the target sequence is variable cannot be assessed from our simple test using a small number of constrained target sequences. Since this algorithm was first developed, BeaconDesigner<sup>®</sup>, a more flexible commercial software package, has been released by PremierBioSoft (Palo Alto, CA, USA). This program allows the user to enter a long target sequence directly from GenBank<sup>®</sup> and determines optimal beacon sequences to detect that target. As in the approach detailed here, BeaconDesigner appears to rely on the *mfold* server to determine the folded conformation and negative free energies of the beacon. For laboratories that are producing large numbers of molecular beacons, the BeaconDesigner package may be an attractive option. However, for smaller laboratories where molecular beacons are not the focus of research, the public domain algorithm described here may be a better initial alternative to aid in their design efforts.

#### ACKNOWLEDGMENTS

The authors wish to acknowledge the support of a National Science Foundation Fellowship (W.T.M) and the Vanderbilt In Vivo Imaging Center, supported by the National Cancer Institute (CA86283).

#### REFERENCES

1. **Rozen, S. and H.J. Skaletsky.** 2000. Primer3 on the WWW for general users and for biologist programmers, p. 365-386. *In* S. Krawetz and S. Misener (Eds.), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ.
2. **SantaLucia, J., Jr.** 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA.* 95:1460-1465.
3. **Seidel, C.A.M., A. Schulz, and M.H.M. Sauer.** 1996. Nucleobase-specific quenching of fluorescent dyes. 1. Nucleobase one-electron redox potentials and their correlation with static and dynamic quenching efficiencies. *J. Phys. Chem.* 100:5541-5553.
4. **Tyagi, S. and F.R. Kramer.** 1996. Molecular beacons: probes that fluoresce upon hybridization. *Nat. Biotechnol.* 14:303-308.
5. **Zuker, M., D.H. Mathews, and D.H. Turner.** 1999. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide, p. 11-43. *In* J. Barciszewski and B.F.C. Clark (Eds.), *RNA Biochemistry and Biotechnology*. NATO ASI Series, Kluwer Academic Publishers, Hingham, MA.

Received 14 August 2002; accepted 9 October 2002.

*Address correspondence to Dr. W. Todd Monroe, 149 E.B. Doran Building, Department of Biological & Agricultural Engineering, Louisiana State University, Baton Rouge, LA 70803-4505, USA. e-mail: tmonroe@lsu.edu*

**W. Todd Monroe and  
Frederick R. Haselton**  
*Vanderbilt University  
Nashville, TN, USA*

**For reprints of this or  
any other article, contact  
Reprints@BioTechniques.com**