W522–W526 *Nucleic Acids Research, 2015, Vol. 43, Web Server issue* *Published online 20 May 2015*
doi: 10.1093/nar/gkv538

Primerize: automated primer assembly for transcribing non-coding RNA domains

Siqi Tian[1], Joseph D. Yesselman[1], Pablo Cordero[2] and Rhiju Das[1,2,3,*]

[1]Departments of Biochemistry, Stanford University, Stanford CA 94305, USA, [2]Program in Biomedical Informatics, Stanford University, Stanford CA 94305, USA and [3]Department of Physics, Stanford University, Stanford CA 94305, USA

Received March 11, 2015; Revised May 11, 2015; Accepted May 11, 2015

ABSTRACT

Customized RNA synthesis is in demand for biological and biotechnological research. While chemical synthesis and gel or chromatographic purification of RNA is costly and difficult for sequences longer than tens of nucleotides, a pipeline of primer assembly of DNA templates, *in vitro* transcription by T7 RNA polymerase and kit-based purification provides a cost-effective and fast alternative for preparing RNA molecules. Nevertheless, designing template primers that optimize cost and avoid mispriming during polymerase chain reaction currently requires expert inspection, downloading specialized software or both. Online servers are currently not available or maintained for the task. We report here a server named Primerize that makes available an efficient algorithm for primer design developed and experimentally tested in our laboratory for RNA domains with lengths up to 300 nucleotides. Free access: http://primerize.stanford.edu.

INTRODUCTION

Biological and biotechnology research is creating a strong demand for custom synthesis of RNA sequences to study the behavior of non-coding RNA molecules in cells and viruses and to design novel RNAs that modulate translation, genome editing, silencing and other biological processes (1,2). Compared to chemical synthesis of RNA, strategies that leverage primer assembly of DNA templates and subsequent *in vitro* transcription by T7 RNA polymerase are rapid and cost-effective, and RNA lengths up to hundreds of nucleotides are readily achievable (3–5). Creating RNAs via this route requires preparing DNA templates, which can be assembled at low cost from mixtures of short primers with lengths up to 60 nucleotides via the polymerase chain reaction (PCR). This problem can be challenging particularly if one wishes to avoid primer 3′ ends that might misprime into incorrect locations and be extended by DNA polymerase into undesired products and if one is not allowed to change the sequence (as is sometimes possible for gene-coding sequences, but not for non-coding RNAs) (6). There has been substantial work on developing algorithms for designing primers for PCR assembly into DNA templates, with special methods to make codon adjustments for protein synthesis (7–10), to optimize primer boundaries against incorrect hybridization of primers (4,9) and to assemble large genes (11,12). However, with the terminated support of previous web servers (4,7), automated primer design tools that optimize against mispriming still require software download, installation and time to learn.

We previously developed a dynamic programming-based algorithm (‘design_primers.m’ in the na_thermo package) to design primers that can be PCR-assembled into templates for high-throughput RNA synthesis and simple kit or bead-based purification (13). Given a desired DNA template sequence, this method, renamed Primerize herein, is optimized to reduce mispriming during PCR by avoiding primer boundaries that might anneal to incorrect sequences. The algorithm has been tested in the synthesis and rapid purification of numerous RNA sequences from our lab with lengths up to 300 nucleotides, including molecules that illustrated damage from standard gel purification methods (14); natural riboswitch aptamers, ribosomal domains and tRNAs (13,15,16); designs from an internet-scale RNA engineering project (17); ‘puzzle’ sequences from community-wide RNA structure prediction trials (18); and domains of human mRNAs (19). In each of these applications, the sequence and purity of the transcribed RNA was verified by reverse transcription and capillary electrophoresis methods (14–16,20,21), with particularly detailed quantitative evaluation of purity for several RNAs in ref. (14). Nevertheless, these scripts previously required MATLAB installation to run and nontrivial efforts to set up. Requests to use this algorithm and the lack of other primer design servers motivated us to prepare an online version of Primerize that should be more broadly useable by the RNA community and testable for other applications, including coding gene synthesis. This report describes the algorithm and details of the current Primerize server implementation.

[*]To whom correspondence should be addressed. Tel: +1 650 723 5976; Fax: +1 650 723 7310; Email: rhiju@stanford.edu

METHOD OVERVIEW

Primerize takes as input a sense-strand DNA sequence. By default, the Primerize server checks for the presence of the T7 RNA polymerase transcription promoter to help avoid the mistake of leaving out this promoter in ordered templates. This check can be turned off for applications that involve different transcription promoters or that seek fragments for gene assembly.

The algorithm designs forward (sense strand) and reverse (anti-sense strand) primers that minimize the total length and therefore the total synthesis cost, of the oligonucleotides. The algorithm obeys a constraint that the hybridization segments between primers have predicted annealing temperatures (T_{m}) above a user-adjustable cutoff (22,23) (60°C, by default) and that primers have lengths between a minimum and maximum length (L_{min} and L_{max}) adjustable by the user (15 and 60 nucleotides, respectively, by default, matching constraints from current DNA synthesis companies). There can be gaps between successive primers for the forward strand or for primers on the reverse strand. Although developed independently, Primerize is a special case of the general 'Gapped Oligo Design' algorithm described and mathematically analyzed in detail by Thachuk and Condon (6), optimizing total primer length summed with a mispriming score (see below) instead of T_{m} (Figure 1B). For completeness, we give a brief description of the method here.

Figure 1A–C illustrates the method. In the first step of Primerize, the T_{m} of every possible overlapping region is pre-calculated; only primers whose overlap give calculated T_{m} above the user-defined cutoff are accepted. In the second step, a heuristic mispriming score (see below) for each possible 3′ end of a primer is also pre-calculated. In the last step, the primers are designed through a recursive dynamic programming algorithm, as follows. In an initial round, all optimal two-primer designs are computed for subsequences that start at $i = 1$ and end at different locations j. In these two-primer solutions, the forward primer's 5′ end is at nucleotide $i = 1$ and the reverse primer's 5′ end is at j (see Figure 1A); the primers' optimal 3′ ends (p and q, respectively) are computed and stored. The calculation is enumerative but fast due to the pre-calculations above and to caching of scores associated with primer end positions p and q. In general, solutions are not found except for subsequences near the beginning of the desired template, due to length constraints on the primers; however these solutions are used in subsequent rounds. In the next round, the optimal four-primer designs that end at each j' are calculated, enumerating over the stored two-primer solutions ending at $j < j'$ and optimizing over endpoints of the two new primers i' .. p' and j' .. q' on forward and reverse strands, respectively. The calculation is continued up to $2m$-primer assemblies, making use of the solutions of the $(m - 1)$-th round. Unless the number of primers is specified, Primerize uses a maximum number of primers of $2\ (N/L_{\mathrm{min}})$, where N is the length of the desired sequence and returns the assembly among all the rounds that ends at $j = N$ (a full coverage of the template) and has the best score. The optimization has a running time that scales quadratically with N, as is checked below.

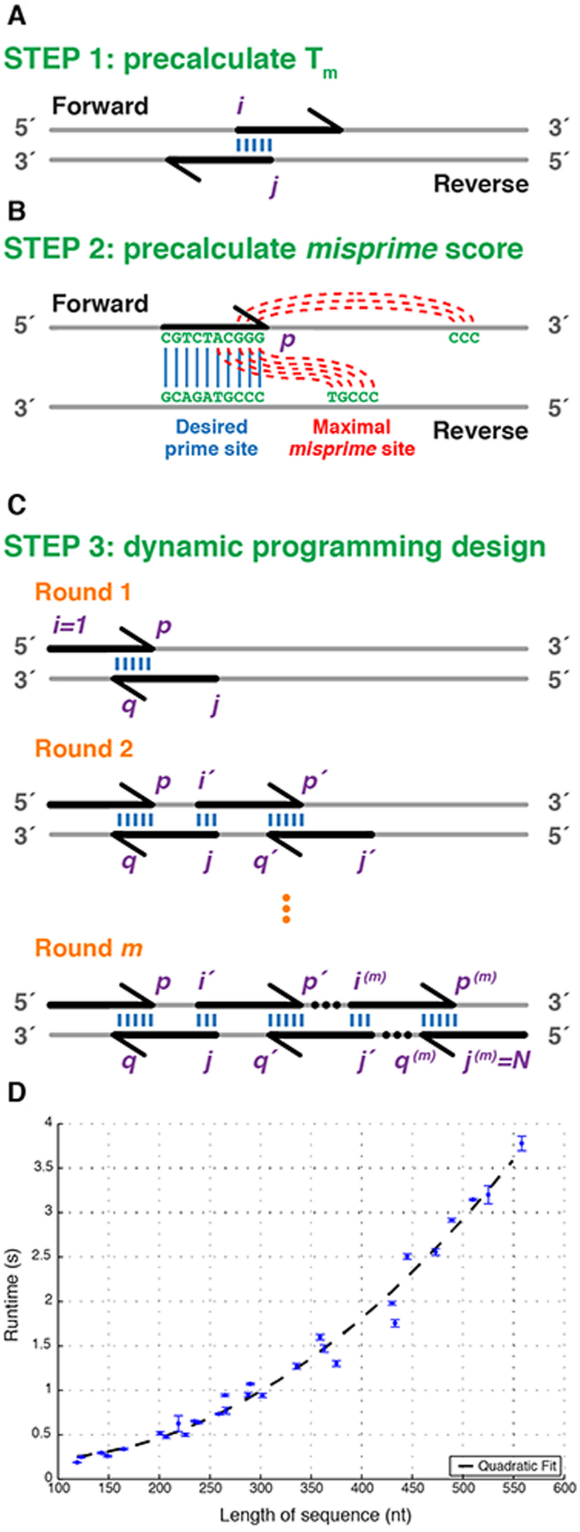
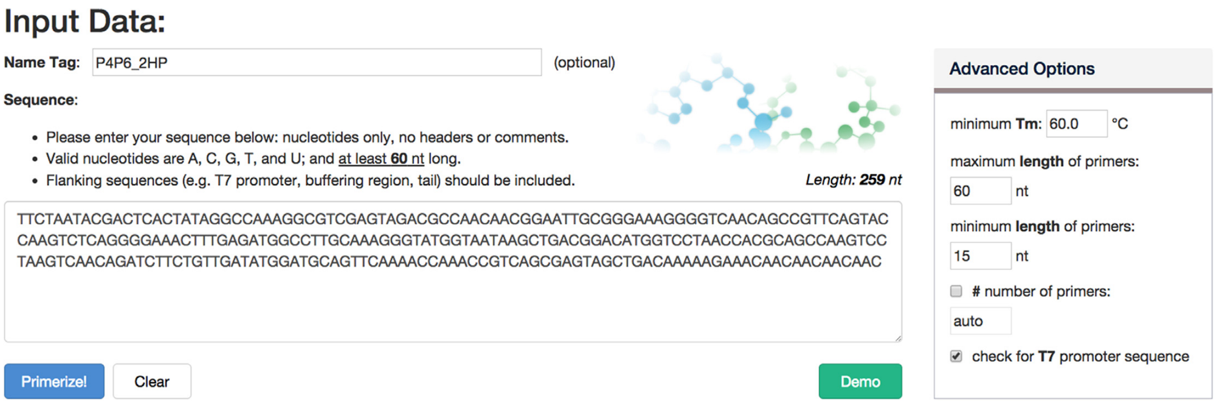


Figure 1. Schematic and runtime of the Primerize algorithm. (**A**–**C**). Schematic of the Primerize algorithm. T_{m} (STEP 1) and *misprime* matrices (STEP 2) are pre-calculated for the dynamic programming assembly. Primerize optimizes the score based on *misprime* and returns the best solution among a range of number of primers (STEP 3). (**D**). Runtime of Primerize on DNA sequences of length between 100 and 600 nucleotides. Each data point is an average of five recorded runtimes of the same sequence. Error bars are standard deviation. A quadratic fit of run time to length of sequence is shown (coefficient of determination R^2 is 0.9850).



Input Data:

Name Tag: P4P6_2HP (optional)

Sequence:

- Please enter your sequence below: nucleotides only, no headers or comments.
- Valid nucleotides are A, C, G, T, and U; and at least **60** nt long.
- Flanking sequences (e.g. T7 promoter, buffering region, tail) should be included.

Length: ***259*** *nt*

```
TTCTAATACGACTCACTATAGGCCAAAGGCGTCGAGTAGACGCCAACAACGGAATTGCGGGAAAGGGGTCAACAGCCGTTCAGTAC
CAAGTCTCAGGGGAAACTTTGAGATGGCCTTGCAAAGGGTATGGTAATAAGCTGACGGACATGGTCCTAACCACGCAGCCAAGTCC
TAAGTCAACAGATCTTCTGTTGATATGGATGCAGTTCAAAACCAAACCGTCAGCGAGTAGCTGACAAAAAGAAACAACAACAACAAC
```

Advanced Options

minimum **Tm**: 60.0 °C

maximum **length** of primers: 60 nt

minimum **length** of primers: 15 nt

□ # number of primers: auto

☑ check for **T7** promoter sequence

Figure 2. Input interface of the Primerize server. Primerize takes a sense-strand DNA template sequence as input. Advanced options, including minimum T_m, maximum and minimum lengths of primers, number of primers, and T7 promoter checking are available for customization.

An important factor governing the success of primer design methods is the optimization function, which we chose originally to be the sum of primer lengths but then revised based on experimental feedback. In initial tests on the *Escherichia coli* 5S ribosomal RNA and the P4–P6 domain of the *Tetrahymena* ribozyme, a method solely optimizing sequence length produced significant fractions of incorrect products that, when isolated and sequenced, corresponded to products due to the 3′ ends of primers 'touching down' on short reverse complementary segments outside their desired location (Figure 1C). We reasoned that such hybridizations would occasionally occur and be recognized by DNA polymerase and since the primer's 3′-most nucleotide would be bound to the complement, the products could be extended by the polymerase. If the products were shorter than the desired full-length product, they would be selectively amplified. Modeling the full process of primer hybridization and extension though multiple PCR cycles is complex, but we reasoned that it would best to penalize the possibility of these events. (We note that this model of mispriming is different than that assumed in (6), which was based on estimating stability of hybridization of primers (T_m) without special consideration of the 3′ ends.)

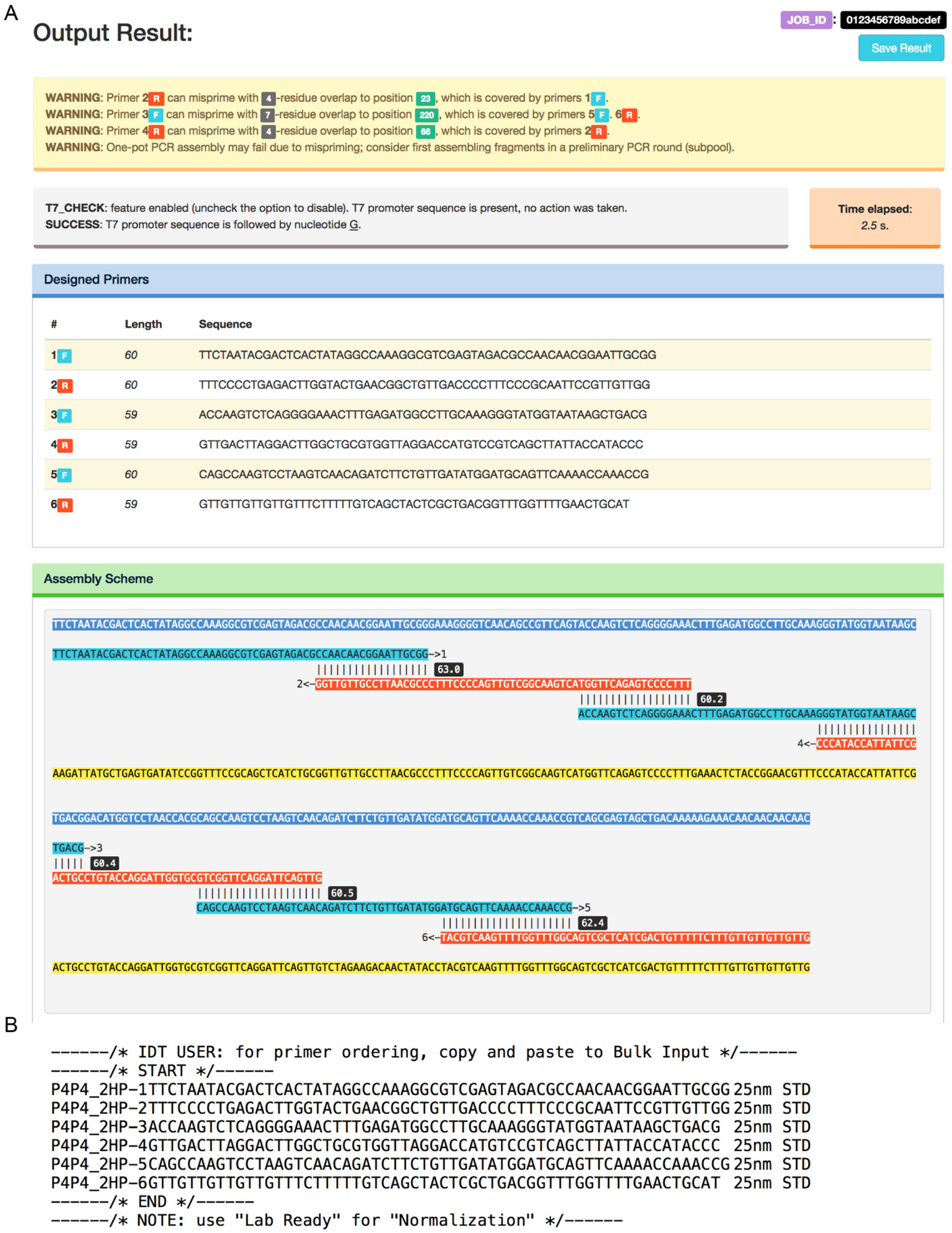
We introduced and experimentally tested a score term *misprime*, whose purpose was to reduce artifacts from mispriming during PCR assemblies. A heuristic score was defined based on the number of contiguous reverse complement matches of the 3′-most n nucleotides of the primer on a possible hybridization site; the score was incremented by 10 for A/T pairings and 12.5 for G/C pairings for each match. We also considered an alternative numerical form of this score based on nearest-neighbor parameters for the DNA base pairs; however, that form appeared less effective at excluding A/T-rich mis-hybridization sites observed in our initial experimental measurements. The scoring system was not further optimized after successful primer assembly in a range of applications. For each potential forward primer 3′ endpoint p, the highest value of this penalty for mispriming to any other segment in the forward or reverse sequence was pre-calculated and stored as $misprime_{\mathrm{forward}}(p)$. An analogous score $misprime_{\mathrm{reverse}}(q)$ was pre-calculated for each possible reverse primer 3′ endpoint q. The design algorithm minimizes the sums of these *misprime* scores for each primer end position p and q plus the sum of forward and backward primer lengths L_p and L_q.

WEB SERVER

The Primerize server uses CherryPy (The CherryPy team, http://www.cherrypy.org), a framework based on Python (Python Software Foundation, https://www.python.org) for basic web services and data management. For the client-side, we coded the web pages in the HTML5 standard (World Wide Web Consortium, http://www.w3.org/TR/html5), with jQuery (The jQuery Foundation, http://jquery.com) for interactive user-interface components and Bootstrap (The Bootstrap team, http://getbootstrap.com) for styling. On the server side, we used Python to create a unique job identifier JOB_ID, to fork the MATLAB interpreter to execute the code that contains input parameters, to parse the results and to render onto a web page. Primerize supports most of the widely used web browsers including Google Chrome, Mozilla Firefox, Apple Safari and Microsoft Internet Explorer.

Figure 2 illustrates an example of Primerize input. The input is a DNA template sequence. Valid input nucleotides are A, C, G, T and U (U is automatically converted to T). An optional name tag can be specified for the sequence. Additionally, advanced options enable customization of the assembly. The minimum T_m (melting temperature) allows the user to adjust the annealing temperature of the overlapping regions; thermodynamic calculations of T_m are based on high-salt nearest-neighbor parameters for DNA (24). Maximum and minimum lengths of primers adjust the length of each primer (building block) and can be modified for long fragment assembly. An option for the number of primers can limit the total number of building blocks.

After design submission, a modal screen is displayed to report the JOB_ID and indicates the calculation is running. Once finished, the output is returned on the same web page (Figure 3A). First, a detailed table of all primers for the assembly of template sequence is returned, with their primer number and direction, length, and sequence. Next, a graphical schematic of the assembly scheme illustrates how the designed primers overlay with each other to permit PCR assembly of the full-length sequence. Primers are drawn



Output Result:

JOB_ID: 0123456789abcdef

WARNING: Primer 2 R can misprime with 4-residue overlap to position 23, which is covered by primers 1 F.
WARNING: Primer 3 F can misprime with 7-residue overlap to position 220, which is covered by primers 5 F. 6 R.
WARNING: Primer 4 R can misprime with 4-residue overlap to position 66, which is covered by primers 2 R.
WARNING: One-pot PCR assembly may fail due to mispriming; consider first assembling fragments in a preliminary PCR round (subpool).

T7_CHECK: feature enabled (uncheck the option to disable). T7 promoter sequence is present, no action was taken.
SUCCESS: T7 promoter sequence is followed by nucleotide G.

Time elapsed:
2.5 s.

Designed Primers

#	Length	Sequence
1 F	*60*	TTCTAATACGACTCACTATAGGCCAAAGGCGTCGAGTAGACGCCAACAACGGAATTGCGG
2 R	*60*	TTTCCCCTGAGACTTGGTACTGAACGGCTGTTGACCCCTTTCCCGCAATTCCGTTGTTGG
3 F	*59*	ACCAAGTCTCAGGGGAAACTTTGAGATGGCCTTGCAAAGGGTATGGTAATAAGCTGACG
4 R	*59*	GTTGACTTAGGACTTGGCTGCGTGGTTAGGACCATGTCCGTCAGCTTATTACCATACCC
5 F	*60*	CAGCCAAGTCCTAAGTCAACAGATCTTCTGTTGATATGGATGCAGTTCAAAACCAAACCG
6 R	*59*	GTTGTTGTTGTTGTTTCTTTTTGTCAGCTACTCGCTGACGGTTTGGTTTTGAACTGCAT

Assembly Scheme

```
TTCTAATACGACTCACTATAGGCCAAAGGCGTCGAGTAGACGCCAACAACGGAATTGCGGGAAAGGGGTCAACAGCCGTTCAGTACCAAGTCTCAGGGGAAACTTTGAGATGGCCTTGCAAAGGGTATGGTAATAAGC

TTCTAATACGACTCACTATAGGCCAAAGGCGTCGAGTAGACGCCAACAACGGAATTGCGG->1
                                          |||||||||||||||||| 63.0
                                       2<-GGTTGTTGCCTTAACGCCCTTTCCCCAGTTGTCGGCAAGTCATGGTTCAGAGTCCCCTTT
                                                                                    |||||||||||||||||| 60.2
                                                                                    ACCAAGTCTCAGGGGAAACTTTGAGATGGCCTTGCAAAGGGTATGGTAATAAGC
                                                                                                                          ||||||||||||||||
                                                                                                                       4<-CCCATACCATTATTCG

AAGATTATGCTGAGTGATATCCGGTTTCCGCAGCTCATCTGCGGTTGTTGCCTTAACGCCCTTTCCCCAGTTGTCGGCAAGTCATGGTTCAGAGTCCCCTTTGAAACTCTACCGGAACGTTTCCCATACCATTATTCG


TGACGGACATGGTCCTAACCACGCAGCCAAGTCCTAAGTCAACAGATCTTCTGTTGATATGGATGCAGTTCAAAACCAAACCGTCAGCGAGTAGCTGACAAAAAGAAACAACAACAACAAC

TGACG->3
||||| 60.4
ACTGCCTGTACCAGGATTGGTGCGTCGGTTCAGGATTCAGTTG
                       |||||||||||||||||||| 60.5
                       CAGCCAAGTCCTAAGTCAACAGATCTTCTGTTGATATGGATGCAGTTCAAAACCAAACCG->5
                                                              ||||||||||||||||||||| 62.4
                                                           6<-TACGTCAAGTTTTGGTTTGGCAGTCGCTCATCGACTGTTTTTCTTTGTTGTTGTTGTTG

ACTGCCTGTACCAGGATTGGTGCGTCGGTTCAGGATTCAGTTGTCTAGAAGACAACTATACCTACGTCAAGTTTTGGTTTGGCAGTCGCTCATCGACTGTTTTTCTTTGTTGTTGTTGTTG
```

```
------/* IDT USER: for primer ordering, copy and paste to Bulk Input */------
------/* START */------
P4P4_2HP-1TTCTAATACGACTCACTATAGGCCAAAGGCGTCGAGTAGACGCCAACAACGGAATTGCGG 25nm  STD
P4P4_2HP-2TTTCCCCTGAGACTTGGTACTGAACGGCTGTTGACCCCTTTCCCGCAATTCCGTTGTTGG 25nm  STD
P4P4_2HP-3ACCAAGTCTCAGGGGAAACTTTGAGATGGCCTTGCAAAGGGTATGGTAATAAGCTGACG  25nm  STD
P4P4_2HP-4GTTGACTTAGGACTTGGCTGCGTGGTTAGGACCATGTCCGTCAGCTTATTACCATACCC  25nm  STD
P4P4_2HP-5CAGCCAAGTCCTAAGTCAACAGATCTTCTGTTGATATGGATGCAGTTCAAAACCAAACCG 25nm  STD
P4P4_2HP-6GTTGTTGTTGTTGTTTCTTTTTGTCAGCTACTCGCTGACGGTTTGGTTTTGAACTGCAT  25nm  STD
------/* END */------
------/* NOTE: use "Lab Ready" for "Normalization" */------
```

Figure 3. Output interface and plain text of the Primerize server. (**A**). Results of Primerize, including potential mispriming warnings, T7 promoter checking, run time, table of all primers and assembly scheme are returned. All results are assigned with a unique JOB_ID and are available for download as plain text. (**B**). Primer information written in a format that can be copied and pasted to IDT Bulk Ordering page.

in their directions (forward or reverse), with the T_m of each overlapping region marked. Any potential mispriming problems are reported to the user as warnings, including the primers involved and the position and length of mispriming regions. If the user's constraints for T_m or the desired number of building blocks cannot be satisfied, an informative error message is provided. The result of T7 promoter checking is also displayed on the output as a separate section. As expected, the running time of the server increased quadratically with the number of input nucleotides, with typical design times of seconds for templates with lengths of 300 nucleotides (Figure 1D).

All results of the Primerize server are made available for download. The user can retrieve a particular run result using the JOB_ID from the home page. Results are anonymously cached on the server for 3 months. We provide a link to save the web page result in plain text format. The text file contains information about the input sequence and parameters, output primer sequences and length, graphical assembly scheme and potential mispriming warnings. We also include a copy of the primer information in a format that is directly compatible with DNA primer ordering in bulk format (Figure 3B) from companies such as Integrated DNA Technologies (Coralville, IA, USA). The user can copy and paste the text into the IDT Bulk Ordering page, a convenient feature when the number of primers is large.

An automatic demonstration of primer design for the 158-nt *Tetrahymena* group-I intron P4-P6 domain, including a T7 promoter sequence, is available through a 'Demo' button on the web server as well as a detailed tutorial page. On the 'Tutorial' page, we have also demonstrated the use of the IDT Bulk Ordering page and a separate 'Protocol' page describes suggested steps and reagents to experimentally assemble the primers by PCR.

SUMMARY

We have developed and launched the Primerize web server, a straightforward tool for designing primers used in RNA synthesis by primer assembly. The underlying algorithm is optimized for minimizing primer boundaries against mispriming and has been stress-tested in several RNA studies involving hundreds of sequences. The online version enables a user-friendly interface with customizable parameters with reasonable default values, automatic checking of promoter sequences, anonymized access and return to long jobs and output formatted for both human evaluation and convenient ordering from synthesis companies. We hope the Primerize server will contribute to RNA bioscience by helping accelerate RNA synthesis.

ACKNOWLEDGEMENT

The authors thank members of the Das laboratory for extensive testing of the web server.

FUNDING

National Institutes of Health [R01 R01GM102519 to R.D.]; Stanford Graduate Fellowship (to S.T.); CONACyT Fellowship (to P.C.); Burroughs Wellcome Foundation Career Award [1007236.01 to R.D.] at the Scientific Interface.
Conflict of interest statement. None declared.

REFERENCES

1. Khalil,A.S. and Collins,J.J. (2010) Synthetic biology: applications come of age. *Nat. Rev. Genet.*, **11**, 367–379.
2. Qi,L.S. and Arkin,A.P. (2014) A versatile framework for microbial engineering using synthetic non-coding RNAs. *Nat. Rev. Micro.*, **12**, 341–354.
3. Sampson,J.R. and Uhlenbeck,O.C. (1988) Biochemical and physical characterization of an unmodified yeast phenylalanine transfer RNA transcribed in vitro. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 1033–1037.
4. Rydzanicz,R., Zhao,X.S. and Johnson,P.E. (2005) Assembly PCR oligo maker: a tool for designing oligodeoxynucleotides for constructing long DNA molecules for RNA production. *Nucleic Acids Res.*, **33**, W521–W525.
5. Gnirke,A., Melnikov,A., Maguire,J., Rogov,P., LeProust,E.M., Brockman,W., Fennell,T., Giannoukos,G., Fisher,S., Russ,C. *et al.* (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechol.*, **27**, 182–189.
6. Thachuk,C. and Condon,A. (2007) *Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering, 2007*. pp. 123–130.
7. Bode,M., Khor,S., Ye,H., Li,M.-H. and Ying,J.Y. (2009) TmPrime: fast, flexible oligonucleotide design software for gene synthesis. *Nucleic Acids Res.*, W214–W221.
8. Xiong,A.-S., Yao,Q.-H., Peng,R.-H., Duan,H., Li,X., Fan,H.-Q., Cheng,Z.-M. and Li,Y. (2006) PCR-based accurate synthesis of long DNA sequences. *Nat. Protoc.*, **1**, 791–797.
9. Hoover,D.M. and Lubkowski,J. (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.*, **30**, e43.
10. Gao,X., Yo,P., Keith,A., Ragan,T.J. and Harris,T.K. (2003) Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Res.*, **31**, e143.
11. Kosuri,S., Eroshenko,N., LeProust,E.M., Super,M., Way,J., Li,J.B. and Church,G.M. (2010) Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat. Biotechnol.*, **28**, 1295–1299.
12. Rouillard,J.-M., Lee,W., Truan,G., Gao,X., Zhou,X. and Gulari,E. (2004) Gene2Oligo: oligonucleotide design for in vitro gene synthesis. *Nucleic Acids Res.*, **32**, W176–W180.
13. Kladwang,W., VanLang,C.C., Cordero,P. and Das,R. (2011) A two-dimensional mutate-and-map strategy for non-coding RNA structure. *Nat. Chem.*, **3**, 954–962.
14. Kladwang,W., Hum,J. and Das,R. (2012) Ultraviolet shadowing of RNA can cause significant chemical damage in seconds. *Sci. Rep.*, **2**, 517.
15. Kladwang,W., Mann,T.H., Becka,A., Tian,S., Kim,H., Yoon,S. and Das,R. (2014) Standardization of RNA chemical mapping experiments. *Biochemistry*, **53**, 3063–3065.
16. Tian,S., Cordero,P., Kladwang,W. and Das,R. (2014) High-throughput mutate-map-rescue evaluates SHAPE-directed RNA structure and uncovers excited states. *RNA*, **20**, 1815–1826.
17. Lee,J., Kladwang,W., Lee,M., Cantu,D., Azizyan,M., Kim,H., Limpaecher,A., Yoon,S., Treuille,A., Das,R. *et al.* (2014) RNA design rules from a massive open laboratory. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 2122–2127.
18. Miao,Z., Adamiak,R.W., Blanchet,M.-F., Boniecki,M., Bujnicki,J.M., Chen,S.-J., Cheng,C., Chojnowski,G., Chou,F.-C., Cordero,P. *et al.* (2015)RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures. *RNA*, **21**, 1065–1084.
19. Xue,S., Tian,S., Fujii,K., Kladwang,W., Das,R. and Barna,M. (2014) RNA regulons in Hox 5[prime] UTRs confer ribosome specificity to gene regulation. *Nature*, **517**, 33–38.
20. Cordero,P., Kladwang,W., VanLang,C.C. and Das,R. (2012) Quantitative Dimethyl Sulfate Mapping for Automated RNA Secondary Structure Inference. *Biochemistry*, **51**, 7037–7039.
21. Kladwang,W., Chou,F.-C. and Das,R. (2012) Automated RNA structure prediction uncovers a Kink-Turn linker in double glycine riboswitches. *J. Am. Chem. Soc.*, **134**, 1404–1407.
22. SantaLucia,J. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 1460–1465.
23. SantaLucia,J. and Hicks,D. (2004) The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 415–440.
24. SantaLucia,J., Allawi,H.T. and Seneviratne,P.A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, **35**, 3555–3562.