PHUSER (Primer Help for USER): a novel tool for USER fusion primer design

Lars Rønn Olsen[1],*, Niels Bjørn Hansen[2], Mads Tvillinggaard Bonde[2], Hans Jasper Genee[2], Dorte Koefoed Holm[2], Simon Carlsen[2], Bjarne Gram Hansen[2], Kiran Raosaheb Patil[3], Uffe Hasbro Mortensen[2],* and Rasmus Wernersson[1],*

[1]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Building 208, [2]Center for Microbial Biotechnology, Department of Systems Biology, Technical University of Denmark, Building 223, DK-2800 Lyngby, Denmark and [3]Structural and Computational Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117, Heidelberg, Germany

ABSTRACT

Uracil-Specific Exision Reagent (USER) fusion is a recently developed technique that allows for assembly of multiple DNA fragments in a few simple steps. However, designing primers for USER fusion is both tedious and time consuming. Here, we present the Primer Help for USER (PHUSER) software, a novel tool for designing primers specifically for USER fusion and USER cloning applications. We also present proof-of-concept experimental validation of its functionality. PHUSER offers quick and easy design of PCR optimized primers ensuring directionally correct fusion of fragments into a plasmid containing a customizable USER cassette. Designing primers using PHUSER ensures that the primers have similar annealing temperature (T_m), which is essential for efficient PCR. PHUSER also avoids identical overhangs, thereby ensuring correct order of assembly of DNA fragments. All possible primers are individually analysed in terms of GC content, presence of GC clamp at 3′-end, the risk of primer dimer formation, the risk of intra-primer complementarity (secondary structures) and the presence of polyN stretches. Furthermore, PHUSER offers the option to insert linkers between DNA fragments, as well as highly flexible cassette options. PHUSER is publicly available at http://www.cbs.dtu.dk/services/phuser/.

INTRODUCTION

The assembly of several DNA fragments is a common task in molecular biology and traditionally this has been achieved by merging matching ends generated by restriction endonucleases. However, this rarely allows for a seamless fusion of two fragments, which is often desirable. Moreover, the need for unique restriction sites at the site of desired fusion can render the process practically unfeasible for large DNA segments; and if possible, the process is time consuming and tedious. A popular way to bypass these obstacles is to merge DNA fragments by fusion PCR (also known as overlapping PCR or overlap extension PCR) (1). While this technique is suitable for achieving seamless fusion, it requires PCR optimization and the amplification rate drops dramatically with an increase in the number of fragments to be fused and/or when the final product size is >4 kb (2,3). USER fusion, which allows rapid and efficient assembly of multiple PCR fragments in a one-step approach (4,5), offers a number of advantages over traditional cloning procedures and fusion PCR: (i) it is independent of restriction sites; (ii) it does not require DNA ligase; (iii) it eliminates the need for amplifying very large fragments and (iv) the final chimerical DNA fragment, which is inserted into the destination vector, is generated from sequences that have only been PCR amplified once, hence, reducing the amounts of PCR generated errors that often compromise fragments fused by PCR. The only requirement of USER fusion is the presence of an A and a T separated by 8–20 bp's in the region that one wishes to fuse (see Figure 1 for an overview of the process). This region is hereafter referred to as segment. The design of the primers for
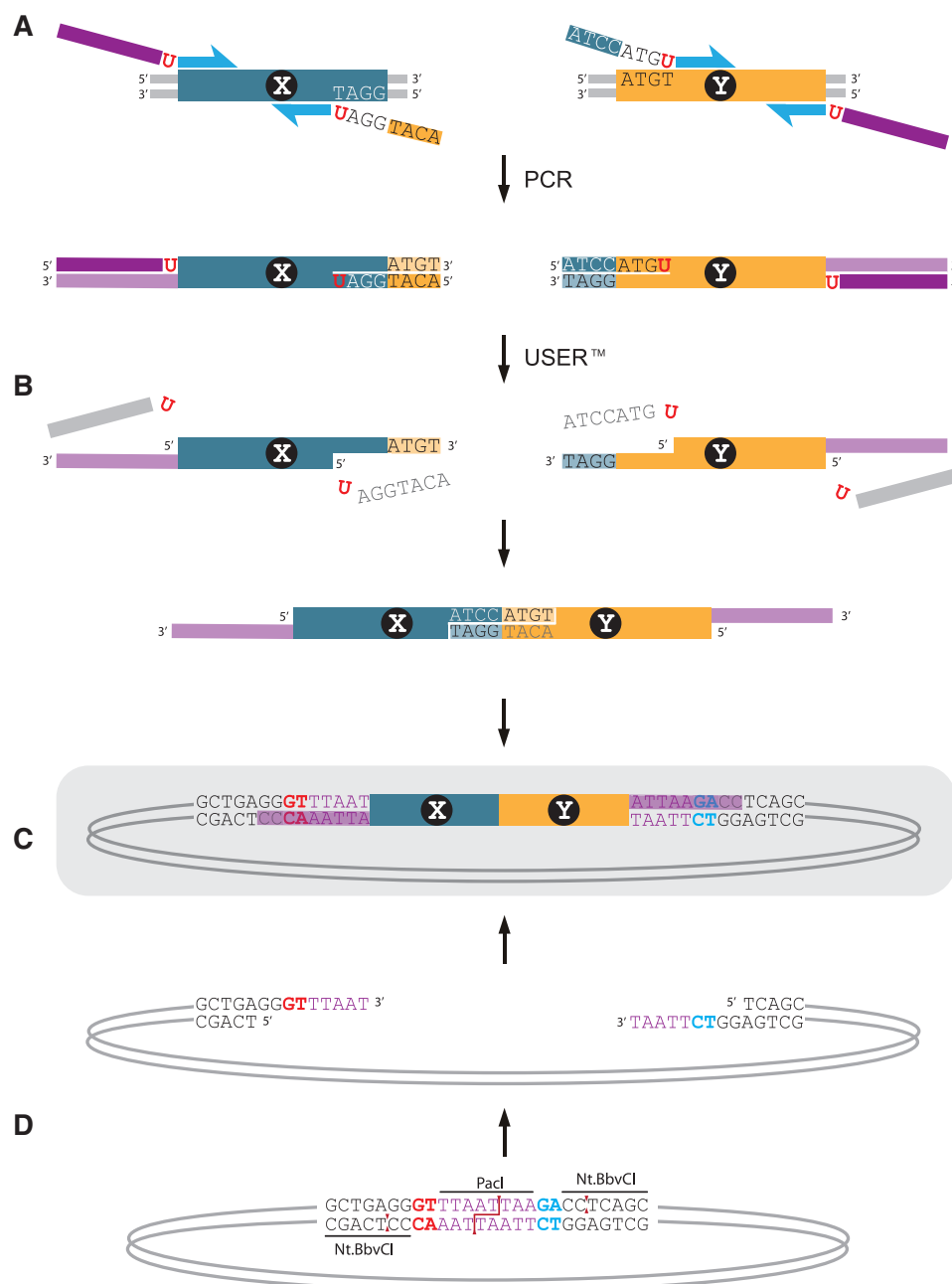
Figure 1. Overview of the USER fusion process. (**A**) PCR amplification of target DNA using uracil containing primers and a proofreading polymerase which does stall at uracil during PCR. (**B**) The PCR products are treated with the uracil DNA glycosylase and DNA glycosylase–lyase Endo VIII enzyme mix (USER^{TM}), creating 3′ extensions. The primer design allows complementary hybridization of overhangs between fragments. (**C**) The overhangs (purple) match extensions of a prepared USER compatible vector. Complementary base paring facilitates self-assembly of the fragments and the vector. The construct is now ready for transformation in *E. coli*. (**D**) Preparation of USER compatible vector is done by digestion of the USER cassette with restriction and nicking enzymes.

PCR amplification of the DNA fragments is based on these segments, and after PCR amplification the DNA fragments are treated with USER^{TM} reagent and will self-assemble in a pretreated plasmid containing a USER cassette. The independence from restriction sites, along with the fact that USER fusion does not require subsequent rounds of PCR (as fusion PCR does), makes the method particularly favourable.

Given that USER fusion is based on a single-step cloning approach, the design of good primers is essential

for ensuring correct and efficient fusions. Designing primers for USER applications is a tedious and time consuming task to perform manually, due to the number of factors (e.g. defining the USER tails, making compatible ends that can be joined in an orderly fashion, ensuring uniform T_m, avoiding hairpins and primer dimer formation), which must be taken into account. To address this problem, we have developed the PHUSER (Primer Help for USER) primer design software that employs a simple, yet powerful algorithm to determine the best primers for

USER fusion, evaluated based on a number of parameters described in the 'Results page' section of this article. PHUSER is user friendly and significantly simplifies the primer design step of USER fusion of multiple DNA fragments.

SOFTWARE FEATURES

PHUSER is available as an easy-to-use web server aimed at molecular biologists, as well as a downloadable UNIX command-line tool aimed at more technically oriented users to whom a scriptable interface to the program is desirable. The web server has been tested in Internet Explorer, Firefox and Safari on Mac OSX, Linux and Windows. The command-line tool is written in Perl with a Python wrapper, and will thus be compatible with most existing computer platforms.

Primer design

The input data for the program are any number of DNA fragments in FASTA format, and a USER cassette consisting of a restriction site and a nicking site, which matches the cassette to be used in the destination vector. Based on this input, PHUSER will identify all segments in proximity of the joining region [i.e. the junction between two DNA fragments (4)] of the DNA fragments containing a dTTP located 8–20 bases downstream of an dATP. This typically gives rise to a number of joining regions, which will produce overhang in the length of 8–20 bases, also called fusion tails. When fusing more than two DNA fragments, unique fusion tails must be used to ensure that the fusion of fragments occurs in the correct order. There is no computational limit as to how many fragments can be processed using PHUSER, although the number of unique fusion tails may be exhausted if multiple fragments are submitted. Six fragments have been reported successfully fused in a single process (6), but the practical experimental limitations have yet to be established.

The part of the primer that binds to the target DNA is here referred to as the binding region. The desired length of a primer binding region is 18–24 bp (7), leading to seven potential primer lengths for both the forward and reverse strand. The selection of the pair of primers is based on matching the annealing temperature, T_m; the T_m is calculated for the binding region part of the primer only. For a given primer pair, the primers' T_m should be within 2°C of each other, in order to achieve efficient PCR amplification (8). The T_m for every possible forward and reverse primer is calculated using the nearest neighbor method (9), and pairs that obey the 2°C rule are stored. If no primers obey the rule, the pairs that have the lowest T_m difference are processed further. In the event that fusion tails are identical, PHUSER compares all the primer pairs, disqualifies pairs with identical fusion tails and selects the next best T_m match as the best primer pair.

Cassette options

PHUSER features a choice of five predefined USER cassettes: (i) PacI/Nt.BbvCI (10); (ii) XbaI(2)/Nt.

BbvCI (http://www.neb.com/nebecomm/ManualFiles/manualE5500.pdf) (11); (iii) AsiSI/Nb.BsmI (12); (iv) PmeI/Nb.BbvCI; and (v) AsiSI/Nb.BtsI (12). Linear assembly (12) is also supported, when PHUSER is used without selecting a USER cassette. Furthermore, users have the option to tailor their own cassettes (see 'Query page' section).

Linker insertion

Lastly, PHUSER enables the insertion of linkers between the DNA fragments. This is done simply by including a FASTA entry between the fragments with the header '>linker'.

WEB INTERFACE FOR PHUSER 1.0

Query page

The PHUSER query page is designed to be simple and intuitive with a tabbed-based interface separating basic from advanced options. DNA sequences can be submitted in FASTA format either by copy/pasting or uploading a text file containing the sequences. PHUSER has a built-in feature to divide large fragments into several shorter fragments for PCR. The desired maximum length of the PCR products can be adjusted in the drop-down menu titled 'Automatic adjustment of maximum PCR product length'. If this option is used, PHUSER will automatically split the input sequence into a number of subsequences each below the defined threshold and design fusion primers for them all. This makes it possible to input a single large sequence (which would be considered too long for efficient PCR) and let the program take care of the entire process of designing primers for USER fusion for this particular sequence.

PHUSER offers three options for cassette selection, one basic and two advanced. The tab titled (i) 'Basic: Predefined fusion cassettes' allows for selection of one of the five predefined cassettes described above. The tab titled (ii) 'Advanced: Select restriction/nicking pair' allows users to custom design a cassette by selecting restriction and nicking enzymes, and entering directional bases. The tab titled (iii) 'Advanced: User designed cassettes' allows users to enter a custom cassette with the format:

5′-GCTGAGGGTTTAAT.TAAGACC.TCAGC-3′

The first period indicates the position of the restriction site and the second period indicates the position of the nicking site. This needs to be entered for both the forward and the reverse strand of the custom cassette.

Results page

Once the desired inputs are submitted, the user will be presented with the results page (Figure 2). The results report is divided into four sections:

  (i) Overview of the primers (5′–3′),
  (ii) Graphic overview of the DNA fragments, primers and their interaction,

Figure 2. Result page of the PHUSER web-server. Here a detailed colour-coded report summarizing the main finding of the primer design is shown, and the user has the option of downloading the primers in FASTA format as well.

(iii) Primer details including assessment of the parameters presented in the 'Primer design' section and

(iv) The final construct after USER fusion.

The quick overview of the primers is written in 5′ to 3′ direction and includes the dUTP. The graphical overview

of DNA fragments and primers provides a visual presentation of where the primers anneal to the template sequences. This is followed by the primer details, showing more specific information regarding each primer. Apart from T_m, this section includes assessment of the following parameters.

```
primer details:
---------------

primer details for gene1

forward primer: GGGTTTAAUTCCACCCTATTGAGGCAT
* Tm: 56.3C - in optimal region (55-72)?            ...YES
* GC ratio is: 50.00% - in optimal region (40-60)?  ...YES
* GC clamp present at 3' end?                        ...YES
* more than 3 G/C out of last 5 bases at 3' end?     ...YES
* risk of primer dimer formation in primer pair?     ...NO
* risk of intra-primer complementarity?              ...NO
* presence of polyN stretches?                       ...NO

reverse primer: AGGCCAUAGTCAGTTAGGGACTTATCG
* Tm: 54.6C - in optimal region (55-72)?            ...NO
* GC ratio is: 45.00% - in optimal region (40-60)?  ...YES
* GC clamp present at 3' end?                        ...YES
* more than 3 G/C out of last 5 bases at 3' end?     ...NO
* risk of primer dimer formation in primer pair?     ...NO
* risk of intra-primer complementarity?              ...NO
* presence of polyN stretches?                       ...NO

Tm of primers within 2C of each other?               ...YES
```

Figure 3. Primer details from the PHUSER report. The desired properties are colour-coded in green and undesired properties are encoded in red.

GC content. The GC content (number of dGTP plus dCTP as a percentage of the total number of nucleic acid residues) should ideally be between 40% and 60%, to keep annealing temperature within a favourable range (9).

Presence of 3′-end GC clamp. Due to the stronger annealing of guanine and cytosine with complementary bases, specific binding is promoted by the presence of at least one of either of these bases in the last five bases of the 3′-end (GC-clamp). However, if more than three out of the last five bases at the 3′-end are guanine or cytosine, the binding to the DNA fragment may be too strong (8).

Risk of Inter- and intra-primer complementarity. Stretches of more than five complimentary bases between different primers used in the same PCR can lead to primer dimer formation, which can interfere with the desired hybridization. Similarly, stretches of more than three complimentary bases within a primer can lead to formation of intra-primer secondary structures, resulting in reduced PCR efficiency. With PHUSER the risk of inter- and intra-primer complementarity is assessed based on the relationship between the T_m of the primer and the complimentary bases potentially leading to undesirable dimerization or intra-primer secondary structures. If the T_m of the potential inter-primer dimerization or intra-primer secondary structure formation is at least 10°C smaller than that of the annealing temperature used in the PCR cycle, the risk of complementarity is considered theoretical, since most of the DNA which could undesirably hybridize is denatured at this higher temperature.

Presence of polyN stretches. Sequences consisting of four or more consecutive identical nucleotides within the primer are known as polyN stretches; polyG or polyC stretches promote non-specific annealing (8), whereas polyA or polyT can cause the 'opening' of stretches of the primer template complex, referred to as 'breathing'(8).

The assessment of each parameter is highlighted in either green (signifying a positive assessment) or red (signifying a negative assessment)—see Figure 3 for an example. Note that the primer parameters used for primer design are supplied only to evaluate the efficiency of a specific amplification and not the competence of the amplification. Therefore, the lack of positive assessment in any parameter (or all parameters) does not preclude a useful PCR outcome.

The last section shows the final product, which will be the result of assembling the sequences into the cassette using USER fusion. The user also has the option to download text files of the report as shown on the screen, or only the best primers in FASTA format, which can be directly used to order the primers.

PROOF OF CONCEPT

Polyketide synthases (PKSs) are complex enzymes encoded by large genes (∼10 kb), and PCR amplification of genes of such lengths is very challenging. One way to overcome this problem is to amplify sections of PCR friendly sizes, followed by fragment reassembly to attain the original gene. USER fusion is a powerful method for assembly of such large genes. In this study, USER fusion was applied for the cloning of six genes encoding putative PKSs of known sequence from *Aspergillus niger* (Table 1). As proof of concept of PHUSER, the software was used for the design of primers (Supplementary Table S1).

According to size, the six PKS genes were amplified as 2, 3, 4 or 5 fragments generating a total of 21 PCR

Table 1. Overview of strategy and efficiency of cloning six PKSs

Model ID (JGI)	Gene annotation	No. of fragments cloned (excluding vector backbone)	Gene length (bp)	Length of PCR fragments cloned (bp)	Fraction correct clones[a]
44965	Putative PKS	2	5340	2624, 2740	2/4
56896	Putative PKS	3	6755	2250, 2170, 2260	2/4
128601	Putative PKS	5	11,657	2040, 2240, 2310, 2730, 2537	2/4
194381	Putative PKS	3	6955	2259, 2450, 2255	4/4
211885	Putative PKS	4	7996	1776, 2310, 1820, 2112	3/4
225574	*fum1*	4	8100	1620, 2310, 2030, 2027	2/4

The genes were amplified in fragments of 1.7–2.7 kb and all primers were designed with PHUSER.
[a]Result obtained by restriction analysis.

products. Prior to transformation into chemically competent *E. coli*, the fragments of the individual genes (five fragments at the most) were purified and mixed in equal quantities with the prepared cloning vector pU1111-1 and treated with USER^{TM} mix. pU1111-1 is a 12 kb expression vector for *Aspergillus nidulans*, including a gpdA promoter, an AsiSI/Nb.BtsI USER cassette, TrpC terminator and an *argB* marker flanked by UP- and DOWN-stream sequences for targeted insertion (12). Transformants were selected on LB media containing 100 µg/ml ampicillin, and the plasmids were purified and tested by restriction analysis. Despite the complexity of the six vector construction reactions, we found that in all cases, 50% or more of the transformants tested contained a plasmid with the correct configuration of inserts (Table 1). Furthermore, a selection of these plasmids was sequenced (StarSeq, Germany) confirming error-free sequence and assembly of all six genes. The false positives observed contained no insert and were therefore caused by incompletely digested pU1111-1. In other USER cloning experiments, we have shown that such false positives can be avoided by prolonged digestion with AsiSI/Nb.BtsI. For further details, please see Supplementary Data. The experiment demonstrates that the use of PHUSER for primer design results in a highly efficient cloning procedure.

Experimental details of the proof of concept study are found in the Supplementary Data.

CONCLUSION

USER fusion is considered an attractive alternative to classical restriction–ligation-based cloning, since it offers simplicity, speed and high efficiency. However, the laborious nature of manually designing good primers is a bottleneck in creating fusion constructs (10,13–15). PHUSER makes the design of good primers quick and easy, and will reduce the time needed to get from strategy to result significantly. Consequently, PHUSER in combination with USER fusion offer means to advance high-throughput generation of constructs. This allows for, e.g. simple construction of gene expression libraries and for vectors expressing genes encoding proteins that are fused to an epitope-tag, a purification-tag or to a fluorescent protein-like GFP. We have shown that PHUSER can be used to design primers that have been

experimentally shown to be efficient, by successful assembly of up to five PCR fragments in a single reaction with a success rate of 50% or higher for the six genes tested. PHUSER also supports different usages of the USER based cloning method, such as homologous recombination for library construction (13) and linear template construction (14). All these features render PHUSER not only an efficient, but also a very flexible tool for designing primers for USER fusion.

SUPPLEMENTARY DATA

Supplementary data are available at NAR Online.

REFERENCES

1. Yon,J. and Fried,M. (1989) Precise gene fusion by PCR. *Nucleic Acids Res.*, **17**, 4895.
2. Pont-Kingdon,G. (1997) Creation of chimeric junctions, deletions, and insertions by PCR. *Methods Mol. Biol.*, **67**, 167–172.
3. Kuwayama,H., Obara,S., Morio,T., Katoh,M., Urushihara,H. and Tanaka,Y. (2002) PCR-mediated generation of a gene disruption construct without the use of DNA ligase and plasmid vectors. *Nucleic Acids Res.*, **30**, E2.
4. Geu-Flores,F., Nour-Eldin,H.H., Nielsen,M.T. and Halkier,B.A. (2007) USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. *Nucleic Acids Res.*, **35**, e55.

5. Nour-Eldin,H.H., Geu-Flores,F. and Halkier,B.A. (2010) USER cloning and USER fusion: the ideal cloning techniques for small and big laboratories. *Methods Mol. Biol.*, **643**, 185–200.
6. Bitinaite,J., Rubino,M., Varma,K.H., Schildkraut,I., Vaisvila,R. and Vaiskunaite,R. (2007) USER friendly DNA engineering and cloning method by uracil excision. *Nucleic Acids Res.*, **35**, 1992–2002.
7. Dieffenbach,C.W., Lowe,T.M. and Dveksler,G.S. (1993) General concepts for PCR primer design. *PCR Methods Appl.*, **3**, S30–37.
8. Apte,A. and Saurabha,D. (2003) PCR primer design. In Dieffenbach,C.W. and Dveksler,G.S. (eds), *PCR Primer a Laboratory Manual*, 2 edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.
9. SantaLucia,J. Jr (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci. USA*, **95**, 1460–1465.
10. Nour-Eldin,H.H., Hansen,B.G., Norholm,M.H., Jensen,J.K. and Halkier,B.A. (2006) Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res.*, **34**, e122.
11. Biolabs,N.E. (2008) USER™ Friendly Cloning Kit, (A Novel Tool for Cloning PCR Products by Uracil Excision) Catalog #E5500S, Version 1.3, 1/08.
12. Hansen,B.G., Salomonsen,B., Nielsen,M.T., Nielsen,J.B., Hansen,N.B., Nielsen,K.F., Regueira,T.B., Nielsen,J., Patil,K.R. and Mortensen,U.H. (2011) Versatile enzyme expression and characterization system for *Aspergillus nidulans*, with the Penicillium brevicompactum polyketide synthase gene from the mycophenolic acid gene cluster as a test case. *Appl. Environ. Microbiol.*, **77(9)**, 3044–3051.
13. Villiers,B.R., Stein,V. and Hollfelder,F. (2010) USER friendly DNA recombination (USERec): a simple and flexible near homology-independent method for gene library construction. *Protein Eng. Des. Sel.*, **23**, 1–8.
14. Stein,V. and Hollfelder,F. (2009) An efficient method to assemble linear DNA templates for in vitro screening and selection systems. *Nucleic Acids Res.*, **37**, e122.
15. Geu-Flores,F., Olsen,C.E. and Halkier,B.A. (2009) Towards engineering glucosinolates into non-cruciferous plants. *Planta*, **229**, 261–270.