SoftwareX 16 (2021) 100881

Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx

Original software publication

REMP software to introduce a screening REstriction site in site-directed Mutagenesis Primer

Madhavi Latha Yadav Bangaru (PhD) [a,1], Ravi Kumar Medabalimi [b,1], Sobhan Babu (PhD) [b,*], Nidhanapati K. Raghavendra (PhD) [a,*]

[a] Department of Biotechnology, Indian Institute of Technology Hyderabad, Kandi, Telangana, 502285, India
[b] Department of Computer Science and Engineering, Indian Institute of Technology Hyderabad, Kandi, Telangana, 502285, India

ARTICLE INFO

Article history:
Received 29 April 2021
Received in revised form 14 October 2021
Accepted 26 October 2021

Keywords:
Site-directed mutagenesis
Python 3
Silent mutation
Restriction site
DNA mutant primer

ABSTRACT

Insertion of silent mutations allowing for restriction site modification aids in the screening of successful mutants during site directed mutagenesis. Introducing a new restriction site requires the analysis of degenerate sequences within mutant primer. As the total number of degenerate codons increases, the analysis becomes increasingly laborious, time-consuming and prone to errors. Towards this, a software named as 'REMP' (for REstriction site in Mutant Primer) was developed and described here. From the input sequence, REMP instantaneously generates degenerate sequences having restriction sites that are 6–8 base pairs in length. The output sequences are arranged based on the number of bases changed compared to the input sequence. REMP software can be installed and run as a stand-alone program on different operating systems. Any user of REMP can edit the list of restriction sites to be considered by the software, without a need for writing a computer code or knowing a program language.

Code metadata

Current code version	*v.1.1*
Permanent link to code/repository used of this code version	https://github.com/ElsevierSoftwareX/SOFTX-D-21-00085
Code Ocean compute capsule	*none*
Legal Code License	MIT
Code versioning system used	*none*
Software code languages, tools, and services used	Python 3.7.6, tkinter, itertools
Compilation requirements, operating environments & dependencies	*Any operating system that can run Python 3.7.6*
If available Link to developer documentation/manual	none
Support email for questions	sobhan@cse.iith.ac.in; raghunk@bt.iith.ac.in

1. Motivation and significance

Site-directed mutagenesis (SDM) is alteration of nucleotides at a specific site on DNA. The alteration in nucleotide (nt) sequence can result in a change of an amino acid coded by the DNA (a missense mutation), or, code for the same amino acid (a silent mutation), or, may not code for any amino acid (a nonsense mutation). SDM causing missense mutation of a particular amino acid to a desired amino acid is used in the structure-function analysis of proteins, and the mutation is termed as a desired mutation. SDM can be experimentally performed using a variety of protocols [1–4], all of which invariably use a mutant DNA primer (20–40 nt in length) carrying a desired mutation. In addition to a desired mutation, a mutant primer invariably contains a coupled screening mutation. A screening mutation is incorporated into a mutant primer to generate a restriction digestion pattern that can easily distinguish a DNA having a desired mutation from one that does not. A change in restriction digestion pattern of a DNA results from either losing an existing restriction site or gaining a new restriction site (a palindromic sequence). Rarely, during designing a mutant primer for SDM, one finds that a desired mutation by itself results in a loss or gain of a restriction site,

Abbreviations: SDM, site-directed mutagenesis; nt, nucleotide; bp, base pair; Tm, melting temperature

* Corresponding authors.
E-mail addresses: madhavib@iith.ac.in (Madhavi Latha Yadav Bangaru), cs18mtech11028@iith.ac.in (Ravi Kumar Medabalimi), sobhan@cse.iith.ac.in (Sobhan Babu), raghunk@bt.iith.ac.in (Nidhanapati K. Raghavendra).

[1] Equal contribution.

https://doi.org/10.1016/j.softx.2021.100881

in which case, the desired and screening mutation are one and the same. More often than not, a silent mutation of amino acids in the vicinity of a desired mutation is necessary to generate a screening mutation.

For a silent mutation to be used as a screening mutation, it has to result in a change in restriction digestion pattern of the DNA. A silent mutation involves substituting an existing codon (comprised of 3 nt) of an amino acid with a degenerate codon of the same amino acid. Different amino acids have different numbers (one to six) of degenerate codons. For a DNA primer of 21 nt in length and coding for seven amino acids, the total number of degenerate codons can range from seven to forty two. The analysis of codons for designing a screening mutation is different in case of losing an existing restriction site and creating a new restriction site. Losing an existing restriction site is relatively straightforward, substitution of a codon that is part of the restriction site with a degenerate codon of the same amino acid almost always serves the purpose. Upon identifying a restriction site in a DNA primer sequence, one can manually replace a degenerate codon within it and ensure that the site is lost. However, creating a new restriction site requires generating a palindromic sequence using a combination of degenerate codons of 2 to 4 adjacent amino acids. For a DNA primer that is 21 nt long and based on the amino acids it codes for, the total number of degenerate sequences to be generated and analyzed can range from 2 to 279936.

A restriction site that can be used for screening a desired mutation has to be at least 6 base pairs (bp) in length. Restriction sites that are less than 6 bp in length or those having non-specific bases (R, Y, X) in them, generate an undesirably large number of DNA fragments of varied lengths and are ineffective in producing a restriction digestion pattern that can distinguish DNA carrying a desired mutation from one that does not. A restriction site that is 6 base pair (bp) in length (a hexamer) can span across a minimum of two (6 nt) and maximum of three codons (9 nt), while a restriction site that is 8 bp in length (an octamer) spans across a minimum of three (9 nt) and maximum of four codons (12 nt). Creating a 6 bp long restriction site by manually generating and analyzing all degenerate sequences of a 21 nt long DNA primer can be challenging, time-consuming and requires sufficient experience to avoid errors. A software that can use degenerate codons to instantaneously generate 6–8 bp long restriction sites from an input sequence, will assist biologists in designing screening mutations in a DNA primer carrying a desired mutation. Such software can make the whole exercise simple, less time-consuming and more efficient, regardless of primer design experience.

Several software tools [5–8] are available on the internet to introduce silent mutations in a given nucleotide sequence. These tools analyze a given nucleotide sequence in all three open reading frames (+1, +2 and +3 ORF) and generate restriction sites ranging in length from 4 to 8 bp. Consequently, the output includes a large number of sequences that: are out-of-frame with respect to ORF coding for protein, have recognition sites that are less than 6 bp long and carry screening-mutations far away from a desired mutation. In order to efficiently design a mutant DNA primer carrying both desired and screening mutations, a biologist requires a software that generates palindromic hexamer or octamer restriction sites within 20 nt on either side of a desired mutation and without altering the amino acids coded by the primer sequence. The generated palindromic sequence has to be a specific recognition site for a standard type II restriction enzyme that cleaves within its recognition site [9]. Such a software will allow the user to design an ideal mutant DNA primer having a desired mutation towards the middle of the primer for stable binding of primer to template, and, a screening mutation sufficiently away from the ends of a primer for efficient extension by a DNA polymerase.

REMP software program described here is developed for biologists to assist in designing ideal mutant DNA primers. Taking advantage of the user's knowledge of ORF of an input primer sequence carrying a desired mutation towards the middle of the sequence, REMP considers the first nt of an input sequence as the first position of a codon (+1 ORF). Hence, the output sequences from REMP have same ORF as that of the input sequence. REMP generates degenerate sequences of an input primer sequence using degenerate codons of amino acids dictated by +1 ORF as specified by the user. Within the degenerate sequences, a sub-set of sequences having palindromes that are identical in sequence to 6 and 8 bp long recognition sites of commercially available restriction enzymes are displayed in the output, along with the name of the restriction enzyme and sequence of its recognition site. The output sequences are sorted according to the number of nt changes made to generate that sequence compared to the input sequence. The output sequences having a least number of nucleotide changes are displayed at the top of the list, and followed in an increasing order of number of nucleotide changes towards the bottom of the list. To design a screening mutation in a mutant DNA primer having a desired mutation, the user can choose an output sequence that best meets all the three criteria of (1) least nucleotide changes compared to an input sequence, (2) a screening mutation far from the ends of a primer, and, (3) a restriction site that generates a restriction digestion pattern which can be distinguished from that of input sequence.

2. Software description

REMP is developed in Python 3.7, along with itertools and tkinter (https://docs.python.org/3/library/) packages. REMP can be run in Windows and Mac operating systems that are most commonly used by biologists, by clicking on the respective application file (generated using Pyinstaller library), without a need for Python 3 compiler. The Python script of REMP (REMP.py) can also be run on other operating systems through Python 3 using command prompt/terminal (see Launch section in REMP link for details). The functioning of REMP is described in the following steps:

1. The user submits mutant DNA primer sequence having a desired mutation towards its middle. The submitted sequence has a complete amino acid codon at the 5′ end of the primer.
2. Upon submitting an input nucleotide sequence, REMP considers the first nucleotide at the 5′ end (left-hand side of user) as the first position of the codon and translates the input sequence to generate the sequence of amino acids coded by the DNA. Incomplete codon at the 3′ end (right-hand side of user) of an input sequence, whose length is not in multiples of three, is not considered for analysis by REMP.
3. In order to create degenerate sequences of the input, REMP considers four consecutive codons as one window. Of the overlapping windows, the first window contains initial four codons (1 to 4), second window contains next four codons (2 to 5), and third window has codons 3 to 6 and so on. Since the maximum number of codons, a restriction site of 8 bp length can span is four, the degenerate sequences obtained will be sufficient to analyze for restriction sites of 6 and 8 bp in length.

4. REMP considers each window at a time and generates degenerate sequences having all degenerate codons (Table 1) of the four codons within the window, while keeping the rest of the input sequence unaltered. This allows the analysis to be completed with a maximum of 5184 degenerate sequences rather than 279936 sequences, for a primer of 21 nt in length.
5. Within the degenerate sequences generated by REMP using a sliding window technique [11] and also the input sequence, REMP identifies palindromic hexamer or octamer sequences that are identical in nucleotide sequence to the restriction sites included in Table 2 and displays those degenerate sequences in the output.
6. For each output sequence, the name(s) of the restriction enzyme(s) along with the nucleotide sequence of the recognition site(s) is displayed to the right of the sequence. In case, no restriction site is found in an input and its degenerate sequences generated by REMP, the word 'None' is displayed in the output.
7. The sequences in the output are arranged in the order of increasing number of bases changed compared to the input sequence.

3. Illustrative example

Following example illustrates the functioning of REMP (Fig. 1). The input nucleotide sequence 'TCAGACCCATTAAAAA GCGCCGACGT' of a DNA primer of 26 nt in length and carrying a desired mutation is provided in the input window having the header 'Mutant Primer'. The 'SUBMIT' button below the input window starts the analysis by REMP. As a first step, REMP translates the input sequence in +1 ORF using the first nt T at the 5′ end as the first position of the codon TCA and based on the standard codon Table 1 [10] generates the amino acid sequence from the codons 'TCA GAC CCA TTA AAA AGC GCC GAC' as 'SDPLKSAD'. Hence, it is mandatory to input complete codon of an amino acid coded by the 5′ end of an input DNA sequence. As a reminder to the user, this is also mentioned above the input window as '[Input complete codon of amino acid at 5′ end]'. The nt 'GT' at the 3′ end of the above input sequence which forms an incomplete codon is not considered for analysis by REMP, hence, although it is not mandatory, it is advantageous to maintain the length of an input DNA sequence as multiples of three. Following the steps described in Section 2, the output sequences with restriction sites generated via silent mutations are displayed in the output window having the header 'Mutant Primer with a Restriction Site'. In the current example, the recognition sites of eight restriction enzymes, namely, AflII, AfeI, BamHI, DraI, EagI, PstI, PvuII and SacII are generated by REMP for the input sequence. The order of display of output sequences is sorted based on the number bases changed in the output sequence with respect to the input sequence. The output sequence with a restriction site: DraI has one base change close to 5′ end, AfeI has one base change close to 3′ end, BamHI has 2 base change close to 5′ end, AflII has 3 base change towards middle, EagI has 3 base change close to 3′ end, SacII has 3 base change closer to 3′ end, PvuII and PstI have 4 base change close to 3′ end. In this example, there are no 8 bp long restriction sites in either the input or the degenerate sequences generated by REMP. Clicking on 'INSTRUCTIONS' opens a separate window having a PDF file that illustrates the best practice for input of a DNA sequence carrying a desired mutation and describes the features of an ideal DNA primer to guide the user for choosing the best output sequence for performing SDM. The 'CLEAR' button erases the input sequence and allows user to input a new sequence for analysis.

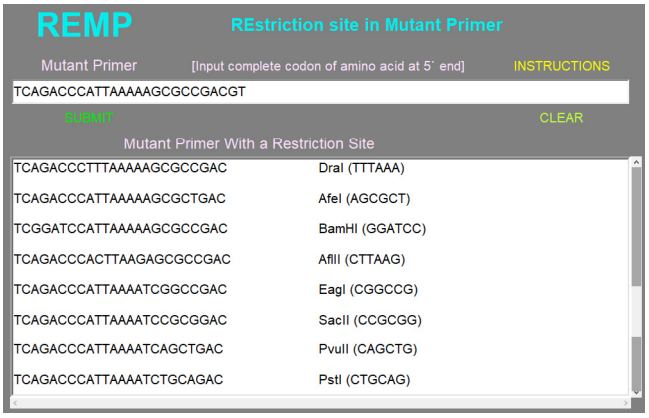

Fig. 1. Screenshot of REMP. The upper part shows REMP header, space for input sequence and command buttons. The lower part shows the output sequences having degenerate sequences generated by REMP and analyzed for presence of restriction sites. The name of the restriction enzyme and the palindromic sequence it recognizes are indicate to the right of the output sequence. The output sequences are arranged according to the increasing number of bases changed with respect to the input sequence.

Designing a mutant DNA primer is a multistep process that requires decision-making at each step. The first step involves choosing a DNA sequence with a desired mutation to use as a primer. REMP performs the second step of introducing restriction sites in a mutant primer, via silent mutations. In the next step, from the restriction sites created by REMP, a restriction site yielding digestion pattern that distinguishes DNA having the desired mutation can be selected based on analysis using NEBCutter software [13]. In the final step, the length of the output sequence of REMP having the selected restriction site can be adjusted by analyzing the sequence in OligoCalc software [14], for obtaining a melting temperature (Tm) that increases specificity of binding to template DNA. By a combined use of REMP, NEBCutter and OligoCalc softwares, one can design ideal mutant primers for SDM, faster and more efficiently [15].

4. Impact

SDM is one of the most powerful protocol for structure-function analysis of proteins. SDM invariably uses a mutant primer for generating a mutant DNA fragment. REMP software assists biologists by instantaneously displaying degenerate sequences, of an input sequence, having restriction sites of 6–8 bp length. Based on the requirement, the user can edit the downloaded Table 1 to accommodate codon usage of the target organism. Similarly, the downloaded Table 2 can be edited by either deleting the restriction sites that are not to be considered or adding new sites, in the same text format. Compared to earlier software for generating silent mutations, REMP provides output sequences with restriction sites that are only 6–8 bp in length and avoids unwanted out-of-frame sequences being shown. As least base changes are preferred during mutant primer designing, REMP arranges the output sequences according to the number of bases changed compared to the input sequence.

5. Conclusions

REMP is a tool to generate silent mutations yielding restriction sites of 6–8 bp in length in a mutant DNA primer. REMP is easy to install and run as a stand-alone program on different operating systems. REMP is specifically developed to assist even a biologist with no prior practical knowledge in designing ideal mutant

Table 1
Degenerate codons of amino acids [10] considered by REMP for generating degenerate sequences from input DNA sequence.

Codons	Amino acid	Codons	Amino acid
GCT, GCC, GCA, GCG	Ala (A)	TTA, TTG, CTT, CTC, CTA, CTG	Leu (L)
CGT, CGC, CGA, CGG, AGA, AGG	Arg (R)	AAA, AAG	Lys (K)
AAT, AAC	Asn (N)	TTT, TTC	Phe (F)
GAT, GAC	Asp (D)	CCT, CCC, CCA, CCG	Pro (P)
TGT, TGC	Cys (C)	TCT, TCC, TCA, TCG, AGT, AGC	Ser (S)
CAA, CAG	Gln (Q)	ACT, ACC, ACA, ACG	Thr (T)
GAA, GAG	Glu (E)	TGG	Trp (W)
GGT, GGC, GGA, GGG	Gly (G)	TAT, TAC	Tyr (Y)
CAT, CAC	His (H)	GTT, GTC, GTA, GTG	Val (V)
ATT, ATC, ATA	Ile (I)	TAA, TGA, TAG	Stop
ATG	Met (M)		

Table 2
List of 6 and 8 bp long specific recognition sites of commercially available restriction enzymes [12] included in the REMP software.

Restriction site	Enzyme	Restriction site	Enzyme	Restriction site	Enzyme
GACGTC	AatII	TTTAAA	DraI	GGCGCC	PluTI
GGTACC	Acc65I	CGGCCG	EagI	CACGTG	PmlI
AACGTT	AclI	GAGCTC	Eco53kI	TTATAA	PsiI-v2
AGCGCT	AfeI	GAATTC	EcoRI	GGGCCC	PspOMI
CTTAAG	AflII	GATATC	EcoRV	CTGCAG	PstI
ACCGGT	AgeI	TGCGCA	FspI	CGATCG	PvuI
GGGCCC	ApaI	AAGCTT	HindIII	CAGCTG	PvuII
GTGCAC	ApaLI	GTTAAC	HpaI	GAGCTC	SacI
ATTAAT	AseI	GGCGCC	KasI	CCGCGG	SacII
CCTAGG	AvrII	GGTACC	KpnI	GTCGAC	SalI
GGATCC	BamHI	CAATTG	MfeI	AGTACT	ScaI
TGATCA	BclI	ACGCGT	MluI	GGCGCC	SfoI
AGATCT	BglII	TGGCCA	MscI	CCCGGG	SmaI
GCTAGC	BmtI	GCCGGC	NaeI	TACGTA	SnaBI
CGTACG	BsiWI	GGCGCC	NarI	ACTAGT	SpeI
TCCGGA	BspEI	CCATGG	NcoI	GCATGC	SphI
TCATGA	BspHI	CATATG	NdeI	AATATT	SspI
TGTACA	BsrGI	GCCGGC	NgoMIV	AGGCCT	StuI
GCGCGC	BssHII	GCTAGC	NheI	TCTAGA	XbaI
TTCGAA	BstBI	TCGCGA	NruI	CTCGAG	XhoI
GTATAC	BstZ17I	ATGCAT	NsiI	CCCGGG	XmaI
ATCGAT	ClaI	ACATGT	PciI	GACGTC	ZraI
GCGGCCGC	NotI	CCTGCAGG	SbfI	GGCCGGCC	FseI
GCGATCGC	AsiSI	ATTTAAAT	SwaI	GTTTAAAC	PmeI
TTAATTAA	PacI	GGCGCGCC	AscI	GCCCGGGC	SrfI

primers, with an easy-to-follow guidance from the 'Instructions' file. Built as a user-friendly software, REMP allows user to edit the list of restriction sites to be considered by the program. REMP eases the difficult step of introducing a new restriction site during design of a mutant DNA primer.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

NKR acknowledges Sruthi Sundaresan for encouraging to develop the software.

Funding

This work was supported in part by funds from Council of Scientific and Industrial Research, India, New Delhi [No. 13(8893-A)/2017-Pool] to MLYB and Department of Biotechnology (India) [grant number BT/PR35825/BRB/10/1840/2019] to NKR.

References

[1] Hutchison CA, Philips M, Edgell MH, Gillam S, Jahnke P, Smith M. Mutagenesis at a specific position in a DNA sequence. J Biol Chem 1978;253:6551–60.
[2] Stemmer WP, Morris SK. Enzymatic inverse PCR: a restriction site independent, single-frame method for high-efficiency, site-directed mutagenesis. Biotechniques 1992;13:214–20.
[3] Papworth C, Bauer JC, Braman J, Wright DA. Quikchange site-directed mutagenesis. Strategies 1996;9:3–4.
[4] Gibson DG, Young L, Chuang RY, Venter JC, 3rd HutchisonCA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods 2009;6:343–5. http://dx.doi.org/10.1038/nmeth.1318.
[5] Shankarappa B, Vijayananda K, Ehrlich GD. SILMUT: a computer program for the identification of regions suitable for silent mutagenesis to introduce restriction enzyme recognition sequences. Biotechniques 1992;12:882–4.
[6] Turchin A, Lawler JF. The primer generator: a program that facilitates the selection of oligonucleotides for site-directed mutagenesis. Biotechniques 1999;26:672–6.
[7] Evans PM, Liu C. SiteFind: a software tool for introducing a restriction site as a marker for successful site-directed mutagenesis. BMC Mol Biol 2005;6:22.
[8] Karnik A, Karnik R, Grefen C. SDM-assist software to design site-directed mutagenesis primers introducing silent restriction sites. BMC Bioinformatics 2013;14:105. http://dx.doi.org/10.1186/1471-2105-14-105.
[9] Raghavendra NK, Rao DN. Functional cooperation between exonucleases and endonucleases- basis for the evolution of restriction enzymes. Nucleic Acids Res 2003;31:1888–96.
[10] Khorana HG. Polynucleotide synthesis and the genetic code. Fed Proc 1965;24:1473–87.

[11] Datar M, Gionis A, Indyk P, Motwani R. Maintaining stream statistics over sliding windows. SIAM J Comput 2002;31:1794–813.
[12] Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE-a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 2015;43:D298–9. http://dx.doi.org/10.1093/nar/gku1046.
[13] Vincze T, Posfai J, Roberts RJ. NEBcutter: a program to cleave DNA with restriction enzymes. Nucleic Acids Res 2003;31:3688–91.
[14] Kibbe WA. OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res 2007;35:W43–6. http://dx.doi.org/10.1093/nar/gkm234.
[15] George A, N.K. Raghavendra. L368F/V408F double mutant of IBD of LEDGF/p75 retains interaction with M178I mutant of HIV-1 integrase. Biochem Biophys Res Commun 2017;490:271–5.