

Primer Design Tool

Leon Rösch

May 2026

1 Computergestütztes Primerdesign

1.1 Datengrundlage

Als biologische Referenz dienen kuratierte Genrecords im FASTA-Format, in denen jedes Zielgen mit 600 nt Up- und Downstream-Flanke und auf den codierenden Strang normalisiert hinterlegt ist; jede CDS wird zu Beginn der Operation auf ATG-Start, ein einzelnes 3'-Stopcodon und eine durch drei teilbare Länge geprüft. Die Vektoren pEXG2 (Allelaustausch via *sacB*-Gegenselektion [3]) und pBBR1MCS2 (Broad-Host-Range, *Plac*, KmR [4]) liegen als GenBank-Dateien vor; pro Kombination aus Vektor, Restriktionsenzym und Anwendung ist eine Tail-Konvention hinterlegt, die die 15 nt Homologiearme für die In-Fusion-Klonierung [7, 6] aus der linearisierten Vektorsequenz definiert.

1.2 Biologische Designlogik

Das Tool implementiert drei Anwendungen, die sich in Fragmentzahl und Junction-Architektur unterscheiden (Abb. 1).

Deletion (pEXG2, 3-Fragment). Im Genrecord werden N N-terminale und C C-terminale Codons der CDS erhalten und über ein neues Stopcodon zu einem in-frame Scar-ORF verbunden. Der Algorithmus durchsucht erschöpfend alle Paare (N, C) mit $N \geq 1$, $C \geq 1$ und $N + C \leq 24$ Aminosäuren und prüft jeweils, ob der resultierende Scar als gültiges Protein mit Start-Methionin und einzelner Stop translatiert. Die 30 nt Junction zwischen Upstream- (UP) und Downstream-Fragment (DN) wird hälftig auf die Tails von P2 und P3 verteilt; P1 und P4 tragen die Vektor-Homologiearme.

In-locus Tagging (pEXG2, 3-Fragment). Eine Tag-Kassette aus GGS-Linker, codon-optimiertem Tag (FLAG, His6, FLAG/His8) und neuem Stopcodon ersetzt das native Stop. Die Kassette *ist* die Junction: für 30 nt Tags wird sie 15 + 15 nt auf P2- und P3-Tail aufgeteilt, für 36 nt Tags (FLAG/His8) entsprechend 18 + 18 nt.

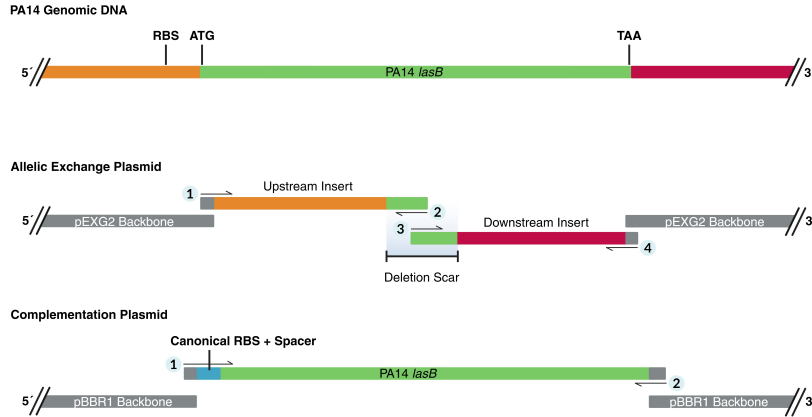


Figure 1: Schematische Architektur der zwei Designanwendungen. Oben: Ausschnitt aus dem Genom von PA14 zeigt beispielhaft das *lasB* Gen mit Up- und Downstream Sequenz. Mitte: pEXG2 Backbone und die zwei Inserts für den Allelic Exchange Vector. Position der vier prognostizierten Primer (1-4). Nach In-Fusion bleibt von dem *lasB* Gen nur die Deletionsnarbe übrig. Primer 1 und 4 liefern den Vektorüberhang. Unten: Insert mit gesamtem *lasB* Gen welches über einen synthetischen Spacer mit der kanonischen RBS (AGGAGG) verbunden ist. Primer 1 und 2 liefern zudem den Vektorüberhang.

Plasmidische Expression (pBBR1MCS2, 2-Fragment). Der P1-Tail kodiert neben dem 15 nt Vektorarm zusätzlich die Shine-Dalgarno-Sequenz AGGAGG [5] und einen 8 nt Spacer (ACTTG TTC); der RBS-ATG-Abstand wird nach Assemblierung auf 8–18 nt verifiziert. Optional kann ein N- oder C-terminaler Tag über GGS-Linker eingefügt werden.

1.3 Primer-Design-Algorithmus

Jeder Primer besteht aus einem festen *Tail*, der die Homologie für die In-Fusion-Reaktion bzw. funktionale Elemente (Scar-Boundary, Tag-Kassette, RBS) bereitstellt, und einem *Body*, der an die Template-DNA bindet und über thermodynamische Filter optimiert wird. Schmelztemperaturen werden mit dem Nearest-Neighbor-Modell nach Allawi und SantaLucia [1] berechnet, parametrisiert auf den In-Fusion-Puffer mit 50 mM Na⁺ und einer Primer-Konzentration von 500 nM.

Auf jeden Body-Kandidaten werden die in Tabelle 1 aufgeführten Hard-Filter angewendet.

Die Primer-Suche ist anwendungsspezifisch ankerbasiert: P1 wird in der Upstream-Flanke mit einem 5'-Offset von 0–150 nt verankert, P4 spiegelbildlich in der Downstream-Flanke; P2 und P3 sind an der Scar- bzw. Kassetten-Grenze

Table 1: Akzeptanzkriterien der Body-Filter im Primer-Design.

Kriterium	Akzeptanzbereich
Länge	18–28 nt
GC-Gehalt	40–70 %
Schmelztemperatur Body	58–66 °C (Soll 60–64 °C)
3'-Endbase	G oder C (Clamp)
G/C in letzten 5 nt	1–3
Homopolymer-Run	< 4 nt
3'-Selbstdimer	< 5 nt komplementär
ΔT_m aller Primer eines Sets	≤ 3 °C (Warnung > 2 °C)

fixiert und werden in der Länge variiert. Findet die Erstsuche keinen Kandidatenpool, relaxiert das Tool die Filter in fester Hierarchie (Homopolymer \rightarrow GC-Gehalt $\rightarrow T_m$ -Fenster \rightarrow 3'-Clamp), während Längen- und Dimer-Filter unverändert bleiben. Aus den verbleibenden Pools wird jenes Quadrupel (bzw. Paar für Expression) ausgewählt, das den minimalen T_m -Spread aufweist; bei mehreren Kandidaten mit gleichem Spread entscheidet die Nähe an die Soll- T_m von 62 °C.

1.4 In-silico-Verifikation

Vor der Ausgabe wird das gesamte Genom des gewählten Isolats gegen alle Primer auf Off-Target-Bindung gescannt: ein 10 nt-3'-Anker erlaubt maximal eine Fehlpaarung, der vollständige Body maximal vier. Aus allen Treffern werden sämtliche möglichen PCR-Produkte enumeriert und gegen die erwarteten Paare validiert (P1+P2 \rightarrow UP, P3+P4 \rightarrow DN, P1+P4 \rightarrow WT-Locus für Kolonie-PCR, P3+P2 muss leer sein; bei Expression nur P1+P2). Zusätzlich wird das in silico assemblierte Plasmid auf exakte Junction-Identität, korrekte Translation des Scars bzw. der Tag-Fusion und auf die erwartete Anzahl der Restriktionsschnittstellen (0 für pEXG2, 1 regeneriert für pBBR1MCS2) geprüft; nur Designs, die alle Hard-Checks bestehen, werden ausgeliefert.

1.5 Implementierung und Ausgabe

Das Frontend ist als Astro-Single-Page-Anwendung umgesetzt und übergibt die Designparameter an eine Python-Serverless-API, in der die gesamte Pipeline (Biopython [2], ReportLab) ausgeführt wird. Pro Anfrage werden ein FASTA mit Primern, Amplikons und assembliertem Plasmid, eine annotierte GenBank-Datei, ein PDF-Protokoll mit PCR-Bedingungen sowie ein JSON-Manifest zur Reproduzierbarkeit zurückgegeben.

References

- [1] Hatim T. Allawi and John SantaLucia. Thermodynamics and NMR of internal G·T mismatches in DNA. *Biochemistry*, 36(34):10581–10594, 1997.
- [2] Peter J. A. Cock, Tiago Antao, Jeffrey T. Chang, Brad A. Chapman, Cymon J. Cox, Andrew Dalke, Iddo Friedberg, Thomas Hamelryck, Frank Kauff, Bartek Wilczynski, and Michiel J. L. de Hoon. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11):1422–1423, 2009.
- [3] Laura R. Hmelo, Bradley R. Borlee, Henrik Almlblad, Michael E. Love, Trevor E. Randall, Boo Shan Tseng, Chuyang Lin, Yasuhiko Irie, Kelly M. Storek, Jaeun Jane Yang, Richard J. Siehnell, P. Lynne Howell, Pradeep K. Singh, Tim Tolker-Nielsen, Matthew R. Parsek, Herbert P. Schweizer, and Joe J. Harrison. Precision-engineering the *pseudomonas aeruginosa* genome with two-step allelic exchange. *Nature Protocols*, 10(11):1820–1841, 2015.
- [4] Michael E. Kovach, Philip H. Elzer, D. Steven Hill, Gregory T. Robertson, Martin A. Farris, R. Martin Roop, and Kenneth M. Peterson. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene*, 166(1):175–176, 1995.
- [5] John Shine and Lynn Dalgarno. The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proceedings of the National Academy of Sciences USA*, 71(4):1342–1346, 1974.
- [6] Takara Bio USA. In-Fusion Snap Assembly User Manual. Technical manual, Takara Bio USA, Inc., 2023.
- [7] Bingliang Zhu, Gang Cai, Eric O. Hall, and Gary J. Freeman. In-Fusion assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations. *BioTechniques*, 43(3):354–359, 2007.