Inducible Genetic Code Expansion in Eukaryotes

Christine Koehler,[a, b, c, d] Gemma Estrada Girona,[c] Christopher D. Reinkemeier,[a, b, c] and Edward A. Lemke*[a, b, c]

Genetic code expansion (GCE) is a versatile tool to site-specifically incorporate a noncanonical amino acid (ncAA) into a protein, for example, to perform fluorescent labeling inside living cells. To this end, an orthogonal aminoacyl-tRNA-synthetase/tRNA (RS/tRNA) pair is used to insert the ncAA in response to an amber stop codon in the protein of interest. One of the drawbacks of this system is that, in order to achieve maximum efficiency, high levels of the orthogonal tRNA are required, and this could interfere with host cell functionality. To minimize the adverse effects on the host, we have developed an inducible GCE system that enables us to switch on tRNA or RS expression when needed. In particular, we tested different promotors in the context of the T-REx or Tet-On systems to control expression of the desired orthogonal tRNA and/or RS. We discuss our result with respect to the control of GCE components as well as efficiency. We found that only the T-REx system enables simultaneous control of tRNA and RS expression.

Genetic code expansion (GCE) is a powerful method to site-specifically introduce noncanonical amino acids (ncAAs) into proteins *in vivo*. In order to achieve this, most commonly an orthogonal aminoacyl-tRNA-synthetase/tRNA (RS/tRNA) pair is used. Usually, the anticodon of the orthogonal tRNA is chosen to recognize the amber stop codon (TAG). The orthogonal RS aminoacylates the tRNA with the ncAA, and the tRNA then suppresses the amber codon to site-specifically incorporate the ncAA into the growing peptide chain.[1] The archaea derived PylRS/tRNA^{Pyl} pairs are among the most popular systems as it is orthogonal in both *Escherichia coli* and eukaryotic systems and over 100 different ncAAs have become available for incorporation with this RS/tRNA pair.[2]

To control the expression of synthetase and tRNA, we developed an inducible amber suppression system that gives us the opportunity to switch on GCE technology on demand. We used the T-REx and Tet-On systems to regulate not only the synthetase gene, but also different kinds of tRNA constructs.

RNA polymerase II (Pol II) transcribes protein-coding genes into mRNAs and hence controls the expression levels of the synthetase. tRNA molecules, instead, are transcribed by RNA polymerase III (Pol III). Eukaryotic tRNAs contain internal promoter regions (i.e. A and B boxes); however, the archaeal tRNA^{Pyl} lacks these and therefore requires an external Pol III-specific promoter (e.g., U6 or H1) or a bicistronic expression cassette (derived from Val or Arg tRNA) for efficient transcription.[2b,3] Major components of a U6 or an H1 promoter are, besides the TATA box, the distal and proximal sequence elements (DSE and PSE), which are important for gene expression.[4] It is assumed that high levels of tRNAs are needed for efficient amber suppression.[5] Accumulation of RS and/or suppressor tRNA could have adverse effects on the host machinery, as this can lead to mischarging of tRNA and/or read-through of natural stop codons.[6]

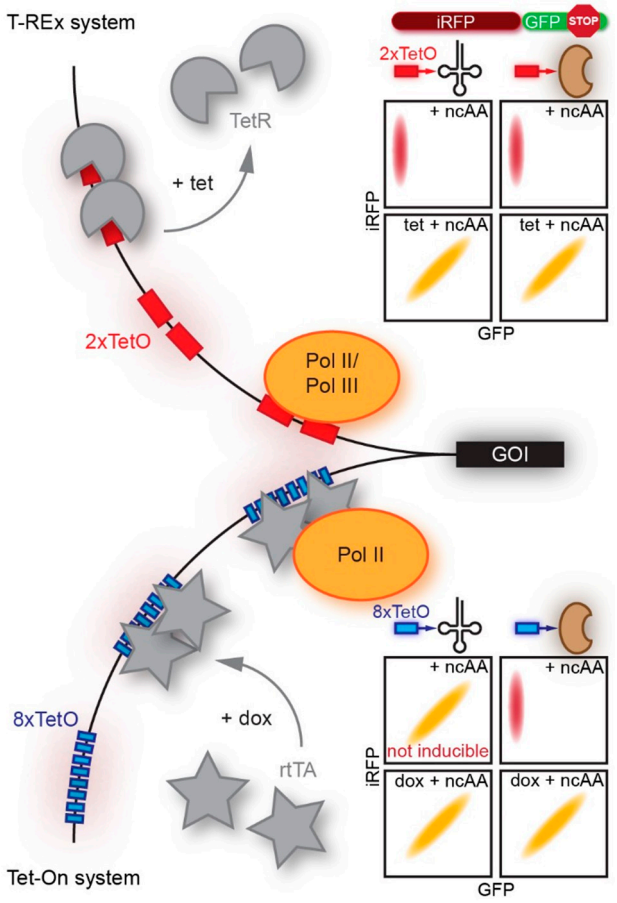
A better control of amber suppression is demanded, as it can help to minimize crosstalk with the host machinery. We developed an inducible GCE method by combining amber suppression technology with known tetracycline inducible systems (Scheme 1). Both the T-REx and Tet-On systems rely on a tet-responsive promoter (Ptet-1 based on the Tn10-specified tetracycline-resistance operon of *E. coli*[7]) in combination with a regulatory element, that is, the tetracycline repressor protein (TetR) or an evolved reverse TetR (rtTA), respectively. This promoter contains two TetO signals, O1 and O2. The TetO2 signal was further used to develop a tetracycline inducible system in eukaryotes (for simplicity we refer to this as TetO). In the T-REx system, TetR binds to the 2xTetO promoter sequence and blocks protein expression. Upon tetracycline (tet) addition TetR undergoes structural changes, unbinds the promoter and transcription can occur.[7,8]

In the Tet-On system, the TetR protein is fused to the transcription activation domain of the herpes simplex virus VP16. Introduction of several mutations lead to the reverse tetracycline-controlled transactivator (rtTA), which binds the promoter region and induces gene expression only in presence of doxycycline (dox).[9] To improve the Tet-On system, several attempts have been undertaken, for example, viral evolution

[a] C. Koehler, C. D. Reinkemeier, Prof. Dr. E. A. Lemke
Biocentre, Johannes-Gutenberg University Mainz
55128 Mainz (Germany)
E-mail: edlemke@uni-mainz.de
[b] C. Koehler, C. D. Reinkemeier, Prof. Dr. E. A. Lemke
Institute of Molecular Biology gGmbH
55128 Mainz (Germany)
[c] C. Koehler, G. Estrada Girona, C. D. Reinkemeier, Prof. Dr. E. A. Lemke
Structural and Computational Biology Unit and Cell Biology and Biophysics Unit
European Molecular Biology Laboratory
Meyerhofstraße 1, 69117 Heidelberg (Germany)
[d] C. Koehler
ARAXA Biosciences GmbH
Meyerhofstraße 1, 69117 Heidelberg (Germany)

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cbic.202000338

This article is part of a Special Collection on Xenobiology. To view the complete collection, visit our homepage



Scheme 1. Overview of T-REx and Tet-On systems in combination with GCE technology. Top: Upon addition of tetracycline (tet), the tet-inducible repressor protein (TetR) unbinds the TetO signals facilitating polymerase (Pol II and Pol III) binding to the promoter. Production of tRNA and translation of the synthetase are inducible in the T-REx system, illustrated by the flow cytometry schemes. The full-length reporter can only be expressed in the presence of tet and ncAA (yellow diagonal ellipse), but not upon addition of ncAA alone (red vertical ellipse). The promoter sequence, containing a 2xTetO signal (red box with arrow) leads to expression of the tRNA or synthetase (shown in brown). Bottom: In the Tet-On system, the reverse Tet-repressor protein (rtTA) can bind to the TetO signals in response to doxycycline (dox), enhancing the binding of Pol II; this enables inducible translation of the synthetase gene, but not a controllable production of tRNA, as illustrated by the flow cytometry schemes. The yellow ellipse represents full-length production of the reporter, whereas the red vertical ellipse illustrates the expression of iRFP alone. An 8xTetO-promoter (blue box with arrow) is positioned in front of the tRNA and synthetase gene (shown in brown).

studies helped to find novel rtTA variants, like Tet-On® 3G (Clontech® Laboratories, Inc.), with increased transcriptional activities and dox-sensitivity.[10] Other studies identified several variants of the promoter Ptet-1 that decrease background expression and increase induced expression levels.[11]

First, we tested the T-REx and Tet-On systems in combination with amber suppression technology, using U6- and H1-driven tRNA expression cassettes. We cloned the tRNA-synthetase, PylRS^{AF} (AF referring to a variant of PylRS 306A 384F) from *Methanosarcina mazei* (Mm) downstream of a nuclear export signal (NES),[6f] under the control of a CMV tet-inducible promoter for the T-REx system (2xTetO) or the CMV dox-inducible promoter for the Tet-On system (8xTetO). To investigate both systems, we followed the expression of a reporter gene, which contains a nuclear localization signal (NLS) upstream of iRFP fused to GFP, harboring the amber stop codon (iRFP-GFP^{Y39TAG}) with fluorescence flow cytometry (FFC). In this reporter, an iRFP signal reports on transfection, whereas full-length GFP and thus green fluorescence is only produced upon successful amber codon suppression by incorporation of a noncanonical amino acid such as *N*-(*tert*-Butoxycarbonyl)-L-lysine (BocK). We designed six different U6- or H1-driven tRNA expression cassettes[12] (Figure 1a). The first set carries TetO sequences surrounding the TATA box for both promoters, and we call these U6-TetO and H1-TetO. The 8xTetO-U6 and 8xTetO-H1 constructs have an additional stretch of 8xTetO sequences upstream of the DSE but no TetO sequences flanking the TATA box. Lastly, the 8xTetO-U6-TetO or 8xTetO-H1-TetO contain both the 8xTetO sequences as well as the two TetO sequences flanking the TATA box.

We tested the inducibility of these constructs via functional amber suppression readout by FFC in the Flp-In™ T-REx™ 293 cell line (Invitrogen), which harbors a stably integrated TetR gene, as well as in HEK293T cells transfected with the Tet-On system using a noninducible NES-PylRS^{AF} as a control and our reporter iRFP-GFP^{Y39TAG}. In the case of the T-REx system, we observe that all constructs containing the TetO sequences flanking the TATA box are inducible by tetracycline (U6-TetO, 8xTetO-U6-TetO, H1-TetO and 8xTetO-H1-TetO). By calculating the ratio of the geometric mean of GFP to iRFP, we show that for these constructs the expression of the full-length reporter, iRFP-GFP^{Y39TAG}, can only be achieved in the presence of tet and BocK, but not with the addition of only BocK. However, if the two TetO sequences surrounding the TATA box are not present (8xTetO-U6 and 8xTetO-H1), tRNA expression is always on and tetracycline independent (Figure 1b). Further controls comparing the data to conventional U6 promoters without any TetO signal are shown in Figure S1 in the Supporting Information. For these constructs, the reporter construct can also be expressed by only adding BocK. In general, higher amber suppression efficiency was observed for U6 versus H1 promoters in line with the known expression strength of those promotors.[13] Next, we tested whether we could also induce the expression of the synthetase gene, and therefore cloned NES-PylRS^{AF} into a plasmid harboring a CMV promoter with 2xTetO sequences (Figure 1c and d). We analyzed the expression of the reporter gene by FFC and, as expected, observed that the synthetase gene is also controllable by tet with this system (Figure S2).

In contrast, the Tet-On system does not allow for inducible U6- or H1-driven tRNA expression (Figures 2 and S3). None of the tested tRNA expression cassettes showed a dox-dependent inducibility but rather a negative influence on the expression level can be seen when dox is present. Differences in the associated mechanisms can explain this lack of inducibility. In the T-REx system, TetR is binding to the promoter region blocking the transcription only when tet is absent. However, in the case of the Tet-On system, rtTA binds the promoter region,

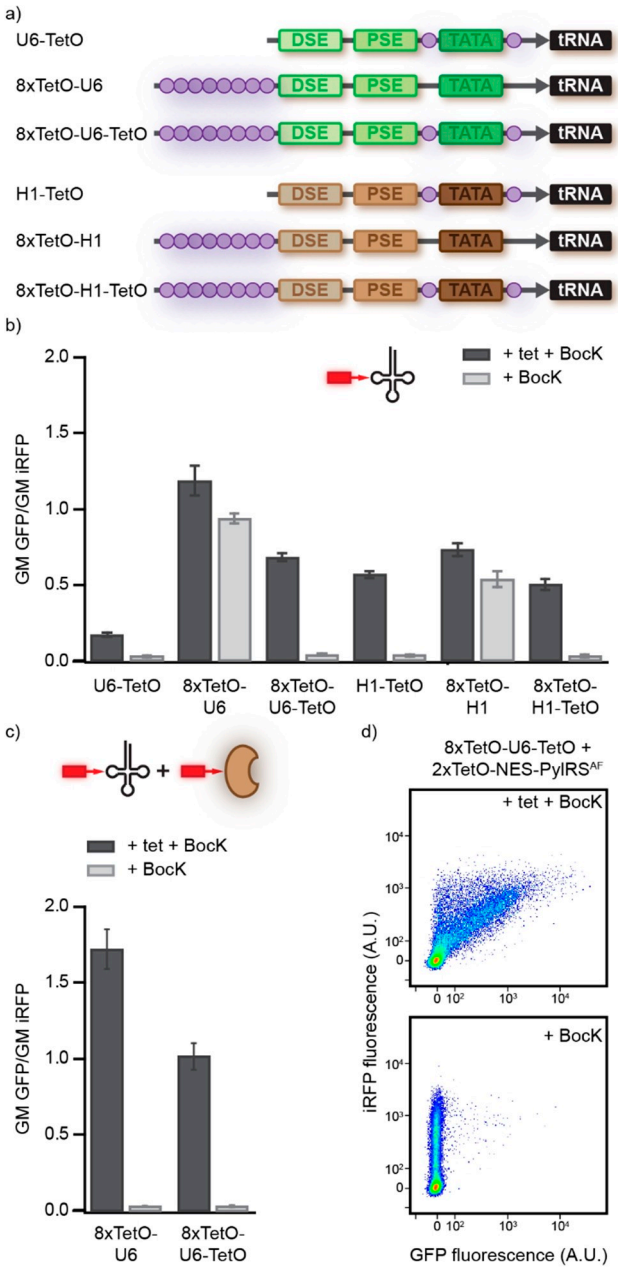


Figure 1. The inducible T-REx system. a) Six different Pol III promoter sequences based on the U6 or H1 promoter are shown. TetO signals (violet circles) are added to specific sites in the promoter. U6-TetO contains two TetO signals, one before and one after the TATA box. 8xTetO-U6 harbors eight TetO signals in front of the distal and proximal sequence elements (DSE and PSE). The 8xTetO-U6-TetO includes the 8xTetO signals before the DSE element and the two TetO signals framing the TATA box. The H1 sequences contain the same pattern as the U6 sequences, but the backbone originating from the H1 promoter. b)–d) HEK Flp-In T-REx 293 cells expressing reporter protein iRFP-GFP^{Y39ncAA} using the T-REx system. b) Measuring the geometric mean (GM) of iRFP and GFP by fluorescence flow cytometry (FFC) using the reporter gene (iRFP-GFP^{Y39ncAA}), the synthetase under CMV promoter in combination with the different Pol III promoters (U6 and H1) with tet and BocK (dark gray) or only with BocK (light gray). The bar plot visualizes how inducible the different Pol III promoters are. c) Bar plot resulting from FFC measurements of the reporter protein using the inducible synthetase gene with and without an inducible tRNA gene. The error bars in b) and c) indicate the standard error of the mean and are calculated from at least three independent measurements. d) FFC data of reporter gene expressed with inducible synthetase construct (2xTetO-CMV-NES-PylRS^{AF}) together with noninducible 8xTetO-U6-TetO-tRNA construct. The upper panel shows expression with tet and BocK, the lower just with BocK.

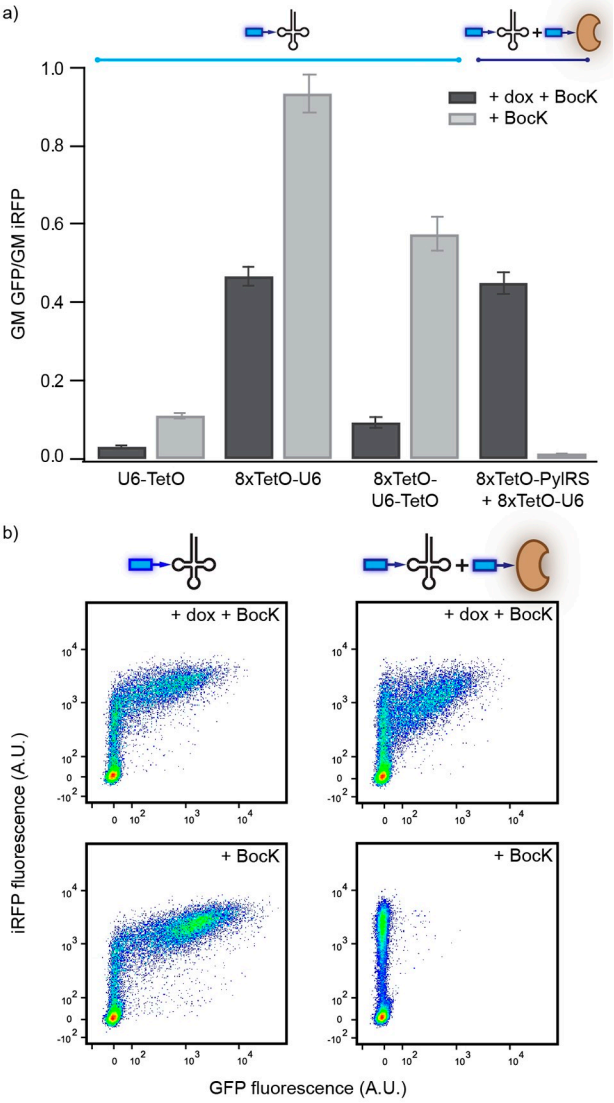


Figure 2. Inducibility of the Tet-On system in HEK293T cells. a) Bar plot illustrating the ratio between the geometric mean (GM) of GFP and iRFP resulting from FFC data measuring expression of iRFP-GFP^{Y39ncAA} with synthetase under CMV promoter in combination with the different Pol III-driven promoter constructs (U6) with and without dox in presence of BocK. The last two columns show the data for the Tet-On system in the presence of the 8xTetO-CMV-NES-PylRS^{AF} construct. None of the three U6 promoters is inducible through the Tet-On system, whereas synthetase expression can be induced by using the Tet-On system. The error bars indicate the standard error of the mean and are calculated from at least three independent measurements. b) FFC data showing expression of iRFP-GFP^{Y39ncAA}. Left: data from the 8xTetO-U6 construct with and without dox (top and bottom, respectively) in the presence of BocK. Right: the inducibility of the GCE system when using the 8xTetO-CMV-NES-PylRS^{AF} construct together with the 8xTetO-U6 construct.

resulting in an activation of transcription in the presence of dox. We speculate that the binding of rtTA to the promoter disturbs the binding of Pol III to the U6 or H1 promoter and the transcription of the tRNA gene cannot occur. As shown in Figure 2, the Tet-On system only allows for control of PylRS expression (8xTetO-CMV-NES-PylRS^{AF}, for more detail see Figure S4).

In summary, we have designed an inducible GCE technology to control the expression level of tRNA as well as of the required synthetase.

We compared different tRNA expression cassettes, derived from U6 and H1 promoter sequences in combination with Tet-On and T-REx systems. We show that for the T-REx system inducibility of the tRNA can be achieved for four Pol III promoter sequences and for the Pol II-dependent promoter of the synthetase gene (Figure 1). On the other hand, it is not possible to induce tRNA expression with the Tet-On system with our current set of tested Pol III promoters, and we speculate that this is caused by an inability of Pol III to bind the promoter sequence as long as rtTA is bound. Only the synthetase translation can be controlled in this system because Pol II is recruited by rtTA and hence can bind to the promoter sequence and transcribe the synthetase gene (Figure 2).

With the T-REx system it is now possible to directly switch on tRNA expression in mammalian cells together with or independently of the RS whenever needed. One disadvantage of the T-REx system is the need for TetR to be present and bound to its promotor before gene expression can be controlled by tet. This system suffers from high background expression in transient transfection because all plasmids are introduced to the cells simultaneously, whereas the TetR protein should be produced first in order to bind the promoter and block Pol II- or Pol III-based transcription. Therefore, the T-REx system is only useful if the TetR protein is stably integrated into the cell line, as in HEK293T FlpIn T-REx 293 cells and not suitable in a transient transfection using for example HEK293T cells (data not shown). However, various genome engineering tools to make stable cell lines expressing proteins via Pol II have now become available. In contrast, achieving high-yielding tRNA expression in stable cell lines is still a huge challenge.[14] Methods like CRISPR and even transposons only introduce a few copies into the genome, whereas in transient transfections, easily 100s of tRNA genes are transferred into the cell.

Once this challenge is addressed, we expect the T-REx system to be useful to generate stable mammalian cell lines for amber suppression with minimized impact of the GCE machinery on housekeeping and physiological function of the cell.

Acknowledgements

The authors thank all the members of the Lemke lab for constructive discussions. We acknowledge funding support by ERC SMPFv2.0, the Gutenbergforschungskolleg and SFB1129 and SPP1623 of the DFG (German Research Foundation, Project no. 240245660, 223208190). Open access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: amber suppression · PylRS · Tet-On · T-REx · unnatural amino acid

[1] a) C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413–444; b) J. W. Chin, *Annu. Rev. Biochem.* **2014**, *83*, 379–408; c) W. Wan, J. M. Tharp, W. R. Liu, *Biochim. Biophys. Acta* **2014**, *1844*, 1059–1070; d) J. S. Italia, Y. Zheng, R. E. Kelemen, S. B. Erickson, P. S. Addy, A. Chatterjee, *Biochem. Soc. Trans.* **2017**, *45*, 555–562.
[2] a) P. R. Chen, D. Groff, J. Guo, W. Ou, S. Cellitti, B. H. Geierstanger, P. G. Schultz, *Angew. Chem. Int. Ed.* **2009**, *48*, 4052–4055; *Angew. Chem.* **2009**, *121*, 4112–4115; b) T. Mukai, T. Kobayashi, N. Hino, T. Yanagisawa, K. Sakamoto, S. Yokoyama, *Biochem. Biophys. Res. Commun.* **2008**, *371*, 818–822; c) H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat. Chem. Biol.* **2008**, *4*, 232–234; d) E. A. Lemke, *ChemBioChem* **2014**, *15*, 1691–1694.
[3] S. M. Hancock, R. Uprety, A. Deiters, J. W. Chin, *J. Am. Chem. Soc.* **2010**, *132*, 14819–14824.
[4] P. Carbon, S. Murgo, J. P. Ebel, A. Krol, G. Tebb, L. W. Mattaj, *Cell* **1987**, *51*, 71–79.
[5] a) A. Garcia, G. Roy, C. Kiefer, S. Wilson, M. Marelli, *PLoS One* **2019**, *14*, e0216356; b) W. H. Schmied, S. J. Elsasser, C. Uttamapinant, J. W. Chin, *J. Am. Chem. Soc.* **2014**, *136*, 15577–15583; c) N. M. Reynolds, O. Vargas-Rodriguez, D. Soll, A. Crnkovic, *Biochim. Biophys. Acta* **2017**, *1861*, 3001–3008.
[6] a) H. J. Park, U. L. RajBhandary, *Mol. Cell. Biol.* **1998**, *18*, 4418–4425; b) J. Guo, C. E. Melancon III, H. S. Lee, D. Groff, P. G. Schultz, *Angew. Chem. Int. Ed.* **2009**, *48*, 9148–9151; *Angew. Chem.* **2009**, *121*, 9312–9315; c) H. Tian, D. Deng, J. Huang, D. Yao, X. Xu, X. Gao, *Biochimie* **2013**, *95*, 881–888; d) A. Bhattacharya, C. Kohrer, D. Mandal, U. L. RajBhandary, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 6015–6020; e) C. D. Reinkemeier, G. E. Girona, E. A. Lemke, *Science* **2019**, *363*; f) I. Nikic, G. Estrada Girona, J. H. Kang, G. Paci, S. Mikhaleva, C. Koehler, N. V. Shymanska, C. Ventura Santos, D. Spitz, E. A. Lemke, *Angew. Chem. Int. Ed.* **2016**, *55*, 16172–16176; *Angew. Chem.* **2016**, *128*, 16406–16410.
[7] M. Gossen, H. Bujard, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 5547–5551.
[8] A. T. Das, L. Tenenbaum, B. Berkhout, *Curr. Gene Ther.* **2016**, *16*, 156–167.
[9] M. Gossen, S. Freundlieb, G. Bender, G. Muller, W. Hillen, H. Bujard, *Science* **1995**, *268*, 1766–1769.
[10] a) A. T. Das, X. Zhou, M. Vink, B. Klaver, K. Verhoef, G. Marzio, B. Berkhout, *J. Biol. Chem.* **2004**, *279*, 18776–18782; b) X. Zhou, M. Vink, B. Klaver, B. Berkhout, A. T. Das, *Gene Ther.* **2006**, *13*, 1382–1390.
[11] R. Loew, N. Heinz, M. Hampf, H. Bujard, M. Gossen, *BMC Biotechnol.* **2010**, *10*, 81.
[12] H. Zhou, C. Huang, X. G. Xia, *Biochim. Biophys. Acta* **2008**, *1779*, 773–779.
[13] P. I. Makinen, J. K. Koponen, A. M. Karkkainen, T. M. Malm, K. H. Pulkkinen, J. Koistinaho, M. P. Turunen, S. Yla-Herttuala, *J. Gene Med.* **2006**, *8*, 433–441.
[14] a) S. J. Elsasser, R. J. Ernst, O. S. Walker, J. W. Chin, *Nat. Methods* **2016**, *13*, 158–164; b) S. Shao, M. Koh, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 8845–8849; c) C. Koehler, P. F. Sauter, M. Wawryszyn, G. E. Girona, K. Gupta, J. J. Landry, M. H. Fritz, K. Radic, J. E. Hoffmann, Z. A. Chen, J. Zou, P. S. Tan, B. Galik, S. Junttila, P. Stolt-Bergner, G. Pruneri, A. Gyenesei, C. Schultz, M. B. Biskup, H. Besir, V. Benes, J. Rappsilber, M. Jechlinger, J. O. Korbel, I. Berger, S. Braese, E. A. Lemke, *Nat. Methods* **2016**, *13*, 997–1000; d) B. Shen, Z. Xiang, B. Miller, G. Louie, W. Wang, J. P. Noel, F. H. Gage, L. Wang, *Stem Cells* **2011**, *29*, 1231–1240.

Manuscript received: May 28, 2020
Revised manuscript received: June 25, 2020
Accepted manuscript online: June 29, 2020
Version of record online: July 23, 2020