mRNA splicing

DOI: 10.1002/anie.201202932

# An RNA Splicing Enhancer that Does Not Act by Looping**

*Helen Lewis, Andrew J. Perrett, Glenn A. Burley,* and Ian C. Eperon**

Mammalian pre-mRNA splicing exhibits an abundance of alternative sites and permissible combinations. This expands the coding possibilities of most genes by several-fold. The splice site signals are poorly conserved and often weak, but their use is augmented by interactions with proteins bound to additional sequences in the introns or exons. These sequences are known as splicing enhancers, and they can be found at distances up to several hundred nucleotides (nt) from the target splice sites.[1–4]

Most well-characterized exonic splicing enhancers (ESEs) are bound by SR proteins. These contain RNA-binding domains and a C-terminal domain rich in arginine-serine dipeptides (RS domain). They stabilize the binding of components that recognize the three canonical splicing signals: U1 snRNPs,[5–7] which base-pair to 5′ splice sites, U2AF protein,[8–10] which binds to 3′ splice sites, and U2 snRNPs,[11,12] which base-pair to branch points. The accepted model for the action of ESEs is that the RS domain encounters the target protein or RNA duplex at a 5′ or 3′ splice site by 3D diffusion and forms a protein-bridged loop in the intervening RNA (Figure 1 a). However, although this model is around 20 years old, it has not been possible to test it definitively. It is supported by two lines of evidence: 1) the rate of splicing ($r$) of a model substrate with an RS domain tethered to an ESE appeared to be related to the number of nt ($n$) between the splice site and the ESE, as predicted for sites interacting by 3D diffusion ($r \propto n^{-3/2}$);[3] 2) an ESE-tethered RS domain could be cross-linked by UV light to RNA near a splice site, demonstrating close proximity.[13] Neither of these results is conclusive. Our analysis of the rate data[3] suggests that $r \propto n^{-5/2}$ or $r \propto e^{-kn}$, where $k$ is an arbitrary constant, neither of which supports the diffusion model for free RNA. Moreover, if entire SR proteins can bind the ESE then the effects of the length $n$ are much reduced.[14] The cross-linking results do not exclude the possibility that a number of SR proteins were bound along the RNA, in a process initiated by the ESE[15] (for example, as in Figure 1 b).

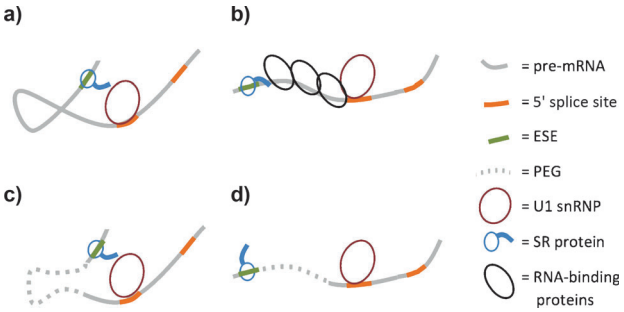


***Figure 1.*** Diagrams of possible mechanisms of action of exonic splicing enhancers (ESEs). a) Direct interactions by looping. An SR protein (blue circle with an arm representing the arginine/serine-rich domain) bound to an ESE (green) could interact directly with proteins bound to a splice site (orange) through 3D diffusion. b) Indirect effects transmitted by RNA. Proteins bound to the RNA might propagate from the ESE to the target protein bound at the splice site. c) Looping should not be prohibited by an intervening flexible linker, for example, PEG. d) Effects transmitted along the RNA could be prevented by a non-natural linker such as PEG.

The looping hypothesis necessarily entails the existence of a flexible chain connecting an ESE and the target site, but the chain need not be RNA. A definitive test would be to insert a flexible non-RNA linker between two sites, which would permit direct encounters by looping (Figure 1 c) but block indirect actions transmitted along the RNA (Figure 1 d). This test was used previously to test whether the contact of splice sites across an intron involved looping. In this case, the introduction of a poly(ethylenegylcol) (PEG) linker into the intron did not inhibit splicing.[16] To apply this test to the action of an ESE, we would need to insert a PEG linker between the ESE and a splice site. The method used previously incorporated DNA sequences flanking the PEG linker, which could compromise the interpretation if the actions of an ESE were inhibited by the DNA, since the cell contains many DNA-binding proteins. To overcome this limitation, we used click chemistry[17,18] to incorporate a PEG linker into RNA.

The test substrate that we used is a pre-mRNA with two possible 5′ splice sites (site 1 and 2; Figure 2). ESEs favor the nearest 5′ splice site.[19–22] We incorporated a GGA-rich ESE[23,24] at the 5′ end of an adenovirus-based pre-mRNA with two alternative 5′ splice sites,[25] and an intervening non-RNA linker was introduced using click chemistry (**4**, Figure 2).

[*] H. Lewis, A. J. Perrett
Department of Chemistry, University of Leicester
University Road, Leicester LE1 7RH (UK)
Dr. G. A. Burley
Department of Pure & Applied Chemistry, University of Strathclyde
295 Cathedral Street, Glasgow G1 1XL (UK)
E-mail: glenn.burley@strath.ac.uk
Homepage: http://www.burleylabs.co.uk
Prof. Dr. I. C. Eperon
Department of Biochemistry, University of Leicester
Lancaster Road, Leicester LE1 9HN (UK)
E-mail: eci@le.ac.uk
Homepage: http://www2.le.ac.uk/departments/biochemistry/staff/eperon

[**] This work was supported by the Leverhulme Trust (F00212Y). G.A.B. thanks the EPSRC for an Advanced Fellowship (EP/E055095/1).
Supporting information for this article (experimental details) is available on the WWW under http://dx.doi.org/10.1002/anie.201202932.

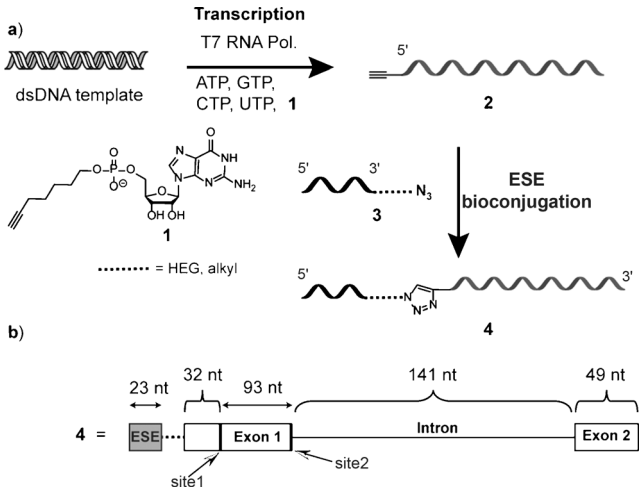


***Figure 2.*** a) Preparation of tripartite RNA transcripts with a general structure **4**. Transcription using the G-initiator **1** affords 5′ alkyne-modified RNA transcripts with the general structure **2**. The tripartite constructs **4** are produced using click chemistry ligation between **2** and the 3′ azido-modified ESE sequences **3** prepared by solid-phase synthesis. b) Scheme of the tripartite transcripts. HEG = hexa(ethyleneglycol).

The synthesis procedure was optimized using a 44 nt model transcript. A 5′ alkyne group was incorporated by initiating transcription with a novel alkyne-conjugated nucleotide **1**. The alkyne substituent enables covalent attachment of an azido-modified ESE **3** to the alkyne-modified transcript using a copper-catalyzed Huisgen [3+2] cycloaddition (click chemistry) to form a triazole linkage between the ESE and the transcript. Click chemistry was chosen because both of the functional groups used (alkyne and azide) are small and bio-orthogonal, thereby eliminating cross-reactivity problems associated with traditional amine–NHS and thiol–maleimide couplings. A drawback to this approach was that the RNA-based ESEs contain phosphorothioate diester linkages, which have not been used previously for click reactions and might inhibit catalysis by copper ions. Alternative bio-orthogonal conjugation reactions are available that do not require copper-catalysis,[26–28] but they require bulky and hydrophobic functional groups that are more likely to significantly perturb splicing.

G-initiators of transcription have been reported using other functional groups,[29–33] but we used a terminal alkyne to minimize any steric perturbation. G-initiator **1** was prepared by solid-phase synthesis using a standard phosphoramidite coupling protocol. This afforded **1** in 26% yield, after purification by reverse-phase HPLC (see Supporting Information).

The rate of incorporation of **1** by the RNA polymerase was highly sensitive to its concentration relative to the concentration of rNTPs. The optimal incorporation of **1** was seen when it was included at 0.4 mM, at which concentration it was incorporated into 72% of the transcripts according to quantitation by gel electrophoresis. Both lower and higher concentrations produced markedly reduced yields (Supporting Information, Figure S1). The presence of an alkyne on the 5′-end of this 44 nt transcript was confirmed by click labeling with fluorescein azide and 3′-azido-modified ESE sequences (Supporting Information, Figure S2 and S3). These optimal conditions were used to transcribe a 5′-truncated version of Ad1WW (transcript defined as A; Figure 3) for the splicing assays.

The 3′-azido-modified ESEs were prepared by solid-phase synthesis using 3′-amino-modified solid supports. After aminolytic cleavage from the controlled-pore glass (CPG) support, ESE1-ESE8 (Table 1) were obtained by coupling with an NHS azide (Supporting Information) and purification by gel electrophoresis. These ESEs varied with respect to: 1) the nucleotide sugar structure (i.e. 2′-*O*-methyl (2′-OMe) or 2′-hydroxy (2′-OH)), 2) the nature of the phosphate backbone (i.e. phosphodiester or phosphorothioate), and 3) the non-RNA linker (i.e. an alkyl or HEG chain).

To prepare the tripartite splicing constructs, the ESEs were ligated by click chemistry to the 5′-alkyne-modified transcript A (Table 2). This proceeded smoothly, albeit in lower overall yields compared to the shorter 44 nt transcript (Supporting Information, Figure S3). The tripartite products were separated by gel electrophoresis from the starting

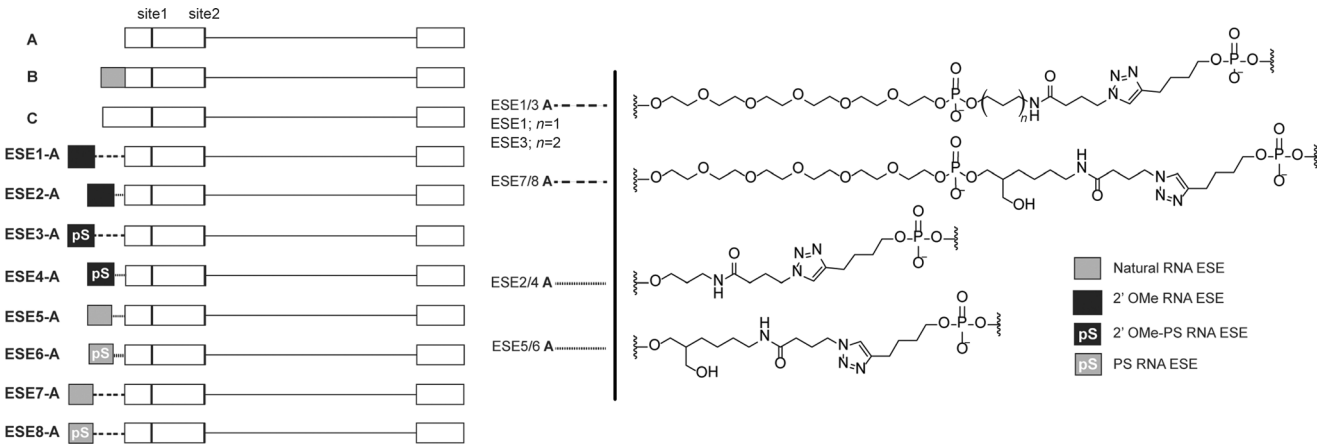


***Figure 3.*** Schematic representation of the various modified transcripts used in the splicing assays. 2′-OMe = 2′-*O*-methyl RNA; pS = phosphorothioate linkage.

  

***Table 1:*** Exonic splicing enhancer (ESE) sequences prepared by solid-phase synthesis and 3′-modified with azide.


| Name | 3′-Modified RNA enhancer sequences |
|---|---|
| ESE1 | $A_oG_oG_oA_oG_oG_oA_oC_oG_oG_oA_oG_oG_oA_oC_oG_oG_oA_oG_oG_oA_oC_oA_o$-HEG-$N_3$ |
| ESE2 | $A_oG_oG_oA_oG_oG_oA_oC_oG_oG_oA_oG_oG_oA_oC_oG_oG_oA_oG_oG_oA_oC_oA_o$-alkyl-$N_3$ |
| ESE3 | $A_{os}G_{os}G_{os}A_{os}G_{os}G_{os}A_{os}C_{os}G_{os}G_{os}A_{os}G_{os}G_{os}A_{os}C_{os}G_{os}G_{os}A_{os}G_{os}G_{os}A_{os}C_{os}A_{os}$-HEG-$N_3$ |
| ESE4 | $A_{os}G_{os}G_{os}A_{os}G_{os}G_{os}A_{os}C_{os}G_{os}G_{os}A_{os}G_{os}G_{os}A_{os}C_{os}G_{os}G_{os}A_{os}G_{os}G_{os}A_{os}C_{os}A_{os}$-alkyl-$N_3$ |
| ESE5 | $A_sG_sG_sA_sG_s$GACGGAGGACGGAGGACA-alkyl-$N_3$ |
| ESE6 | $A_sG_sG_sA_sG_sG_sA_sC_sG_sG_sA_sG_sG_sA_sC_sG_sG_sA_sG_sG_sA_sC_sA_s$-alkyl-$N_3$ |
| ESE7 | $A_sG_sG_sA_sG_s$GACGGAGGACGGAGGACA-HEG-$N_3$ |
| ESE8 | $A_sG_sG_sA_sG_sG_sA_sC_sG_sG_sA_sG_sG_sA_sC_sG_sG_sA_sG_sG_sA_sC_sA_s$-HEG-$N_3$ |

o = 2′-OMe; s = phosphorothioate; HEG = hexa(ethyleneglycol). See Figure 3 for linker formulas.

***Table 2:*** RNA transcripts used in this study and the lengths of the alternative mRNA products resulting from the use of 5′ splice sites 1 or 2.


| Transcript | Pre-mRNA [nt] | Site 1 mRNA [nt] | Site 2 mRNA [nt] |
|---|---|---|---|
| A | 315 | 81 | 174 |
| B | 337 | 103 | 196 |
| C | 336 | 102 | 195 |

Transcript A was the substrate used for ligation to the synthetic ESE sequences. Transcript B included the ESE sequence as an integral part of the transcribed pre-mRNA, and transcript C was the same length as B but it contained the natural sequence at the 5′ end. After splicing for 90 min, the majority of the spliced mRNA from transcripts A and C came from the use of site 2, whereas the presence of the ESE in transcript B shifted splicing substantially to site 1, closer to the ESE (Figure 4a,b). The preference for site 1 in transcript B does not depend on a 5′ cap on the transcript (Supporting Information, Figure S4). Strikingly, none of the ligated ESEs, whether based on 2′OMe or RNA chemistry, were able to enhance the use of site 1. This phenomenon was seen whether the linker between the ESE and the transcript contained short alkyl chains (ESE2, ESE4, ESE5, and ESE6) or a single HEG linker (ESE1, ESE3, ESE7, and ESE8). The ESEs containing 2′OMe nucleotides, without phosphorothioate substitution (ESE1 and ESE2), alkyne transcript.[34] Surprisingly, the phosphorothioate backbone enhanced the yield of the click ligation, as determined by quantification of the radiolabeled product bands relative to the starting alkyne transcript. For example, the yield of the click ligation between the 44-mer alkyne transcript and ESEs at 10 μM was 12% for ESE1 and 44% for ESE3 (Supporting Information, Figure S2). This is the first example of a click reaction with a phosphorothioate backbone and extends the scope of click chemistry for nucleic acids. This is particularly important for the use of modified oligonucleotides that re-direct splicing reactions for therapeutic purposes.[35] Most of the major strategies of this involve the use of oligonucleotides with phosphorothioate linkages,[24,36,37] and the ability to use click chemistry to add functionalities that facilitate nuclease resistance, targeting, and uptake would be a significant advantage.

The RNA splicing preferences of the tripartite ESE-conjugated pre-mRNAs (Figure 3) were tested by incubating them in a HeLa cell nuclear extract for 90 minutes. Figure 4 shows the result of a splicing assay using 2′-OMe-based ESE constructs (ESE1–ESE4) and 2′-OH/pS-based constructs (ESE5–ESE8).

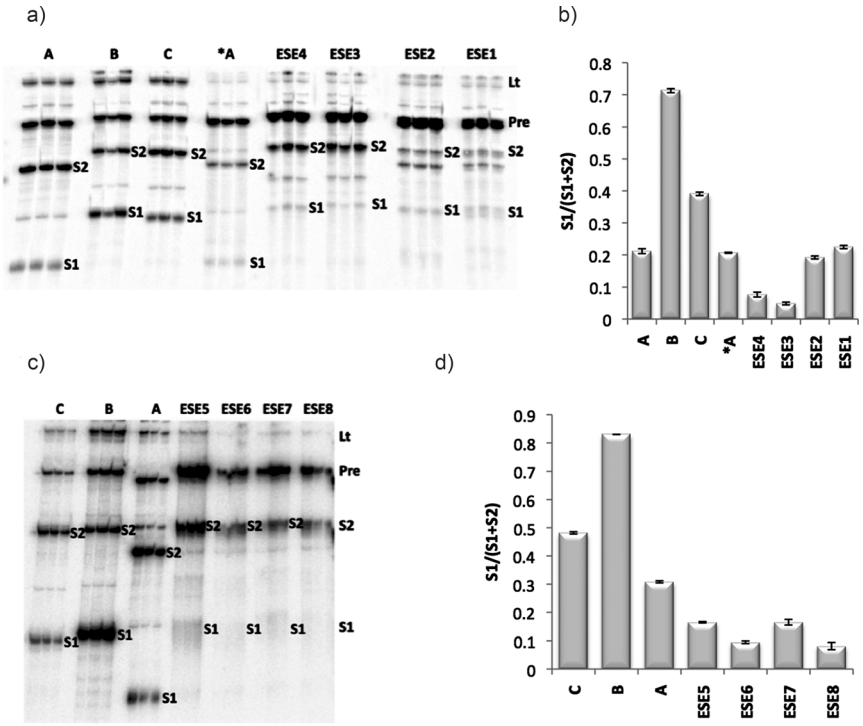


***Figure 4.*** a) Polyacrylamide gel electrophoresis of radiolabeled RNA after splicing reactions done in triplicate for 90 minutes. A, B, and C are unconjugated transcripts; ESE1-ESE4 are tripartite molecules with transcript A conjugated to the 2′OMe ESE sequences in Table 1. *A is transcript A conjugated to tri(ethyleneglycol) (*O*-(2-Azidoethyl)-*O*′-methyl-triethylene glycol) but not an ESE. S1 and S2 show product mRNAs formed using splicing to sites 1 and 2, respectively; Pre is unspliced pre-mRNA; Lt shows lariat by-products of splicing. b) Means and standard deviations for the proportion of mRNA spliced to site 1. c) Same conditions as in (a) for the RNA-based ESE5–ESE8. d) Means and standard deviations for the proportion of mRNA spliced to site 1 in (c).

were susceptible to degradation, giving rise to double bands of product mRNA. In previous work, we have shown that phosphorothioate substitution in the residues near the 5′ end increases the activity of the ESE.[23] However, although the RNA-based ESEs with phosphorothioates (as in ESE5 and ESE7) showed no degradation, they did not enhance the use of site 1.

These results show that a flexible non-RNA linker between the ESE and a target site is unable support ESE activity. This means either that contiguous RNA is required, or that the triazole group itself is somehow intrinsically inhibitory. We have shown elsewhere[40] that ESE2 can replace the corresponding contiguous 2'OMe sequence in a bifunctional oligonucleotide enhancer with no loss of activity; the insertion of the alkyl linker between the enhancer and the portion of the oligonucleotide complementary to an exon was not inhibitory. Indeed, the insertion of a HEG linker (as in ESE1 but with $n=2$ in Figure 3) produced a notable increase in activity. In these cases, the ESE is attached amidst proteins on the exon and direct interactions with neighboring proteins would be possible. This shows that the composition of the linker does not intrinsically inhibit the action of the ESE. Hence, we conclude from our present results that the stimulation of splicing at the proximal 5′ splice site by an ESE does not involve RNA looping, that is, a direct encounter through three-dimensional diffusion, but contiguous RNA is required.

One feasible mechanism involving contiguous RNA is that proteins propagate along the exon from the ESE. This is consistent with early observations that exons appeared more particulate than introns when viewed by electron microscopy.[38] We have proposed recently that tightly bound U1 snRNPs also trigger the propagation of proteins along the exon, making it more rigid.[39] SR proteins and U1 snRNPs may trigger the same event, initiated by whichever one binds most stably. The effects on splicing rates of the distance between an ESE and a target splice site[3] could be attributed to a constant attenuation of propagation along the RNA. The methods we described herein can be applied to other ESEs, such as those in 3′ exons and within internal exons, and to a wide range of other functional interactions between sites in RNA to form the first general approach to assess the functional importance of specific sequences within the context of alternative pre-mRNA splicing.

Received: April 17, 2012
Revised: July 10, 2012
Published online: August 31, 2012

**Keywords:** click chemistry · mRNA · oligonucleotides · pre-mRNA splicing · splicing enhancers

[1] Z. Wang, C. B. Burge, *RNA* **2008**, *14*, 802.
[2] A. Lavigueur, H. La Branche, A. R. Kornblihtt, B. Chabot, *Genes Dev.* **1993**, *7*, 2405.
[3] B. R. Graveley, K. J. Hertel, T. Maniatis, *EMBO J.* **1998**, *17*, 6747.
[4] M. Tian, T. Maniatis, *Genes Dev.* **1994**, *8*, 1703.
[5] S. Cho, A. Hoang, R. Sinha, X. Y. Zhong, X. D. Fu, A. R. Krainer, G. Ghosh, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 8233.
[6] I. C. Eperon, D. C. Ireland, R. A. Smith, A. Mayeda, A. R. Krainer, *EMBO J.* **1993**, *12*, 3607.
[7] J. D. Kohtz, S. F. Jamison, C. L. Will, P. Zuo, R. Luhrmann, M. A. Garcia-Blanco, J. L. Manley, *Nature* **1994**, *368*, 119.
[8] B. R. Graveley, K. J. Hertel, T. Maniatis, *RNA* **2001**, *7*, 806.
[9] Z. Wang, H. M. Hoffmann, P. J. Grabowski, *RNA* **1995**, *1*, 21.
[10] J. Y. Wu, T. Maniatis, *Cell* **1993**, *75*, 1061.
[11] M. Martins de Araujo, S. Bonnal, M. L. Hastings, A. R. Krainer, J. Valcarcel, *RNA* **2009**, *15*, 515.
[12] H. Shen, M. R. Green, *Genes Dev.* **2006**, *20*, 1755.
[13] H. Shen, J. L. Kan, M. R. Green, *Mol. Cell* **2004**, *13*, 367.
[14] K. S. Sciabica, K. J. Hertel, *Nucleic Acids Res.* **2006**, *34*, 6612.
[15] B. R. Graveley, *Mol. Cell* **2004**, *13*, 302.
[16] Z. Pasman, M. A. Garcia-Blanco, *Nucleic Acids Res.* **1996**, *24*, 1638.
[17] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem.* **2001**, *113*, 2056; *Angew. Chem. Int. Ed.* **2001**, *40*, 2004.
[18] M. Meldal, C. W. Tornøe, *Chem. Rev.* **2008**, *108*, 2952.
[19] M. Gabut, M. Mine, C. Marsac, M. Brivet, J. Tazi, J. Soret, *Mol. Cell. Biol.* **2005**, *25*, 3286.
[20] E. C. Ibrahim, T. D. Schaal, K. J. Hertel, R. Reed, T. Maniatis, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 5002.
[21] S. Spena, M. L. Tenchini, E. Buratti, *RNA* **2006**, *12*, 948.
[22] Z. Wang, X. Xiao, E. Van Nostrand, C. B. Burge, *Mol. Cell* **2006**, *23*, 61.
[23] N. Owen, H. Zhou, A. A. Malygin, J. Sangha, L. D. Smith, F. Muntoni, I. C. Eperon, *Nucleic Acids Res.* **2011**, *39*, 7194.
[24] L. A. Skordis, M. G. Dunckley, B. Yue, I. C. Eperon, F. Muntoni, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 4114.
[25] L. O'Mullane, I. C. Eperon, *Mol. Cell. Biol.* **1998**, *18*, 7510.
[26] N. K. Devaraj, R. Weissleder, *Acc. Chem. Res.* **2011**, *44*, 816.
[27] J. A. Prescher, C. R. Bertozzi, *Nat. Chem. Biol.* **2005**, *1*, 13.
[28] E. M. Sletten, C. R. Bertozzi, *Angew. Chem.* **2009**, *121*, 7108; *Angew. Chem. Int. Ed.* **2009**, *48*, 6974.
[29] E. Paredes, S. R. Das, *ChemBioChem* **2011**, *12*, 125.
[30] E. Paredes, M. Evans, S. R. Das, *Methods* **2011**, *54*, 251.
[31] S. Pfander, R. Fiammengo, S. I. Kirin, N. Metzler-Nolte, A. Jäschke, *Nucleic Acids Res.* **2007**, *35*, e25.
[32] J. C. Schlatterer, A. Jaschke, *Biochem. Biophys. Res. Commun.* **2006**, *344*, 887.
[33] J. Schoch, S. Ameta, A. Jaschke, *Chem. Commun.* **2011**, *47*, 12536.
[34] S. I. Presolski, V. Hong, S.-H. Cho, M. G. Finn, *J. Am. Chem. Soc.* **2010**, *132*, 14570.
[35] I. Eperon, *Nat. Chem. Biol.* **2012**, *8*, 507.
[36] N. M. Goemans, M. Tulinius, J. T. van den Akker, B. E. Burm, P. F. Ekhart, N. Heuvelmans, T. Holling, A. A. Janson, G. J. Platenburg, J. A. Sipkens, J. M. Sitsen, A. Aartsma-Rus, G. J. van Ommen, G. Buyse, N. Darin, J. J. Verschuuren, G. V. Campion, S. J. de Kimpe, J. C. van Deutekom, *N. Engl. J. Med.* **2011**, *364*, 1513.
[37] Y. Hua, K. Sahashi, F. Rigo, G. Hung, G. Horev, C. F. Bennett, A. R. Krainer, *Nature* **2011**, *478*, 123.
[38] A. L. Beyer, A. H. Bouton, O. L. Miller, Jr., *Cell* **1981**, *26*, 155.
[39] M. J. Hodson, A. J. Hudson, D. Cherny, I. C. Eperon, *Nucleic Acids Res.* **2012**, DOI: 10.1093/nar/gks322.
[40] A. J. Perrett, R. L. Dickinson, Z. Krpetić, M. Brust, H. Lewis, I. C. Eperon, G. A. Burley, unpublished results.

  