DNA Nanotechnology

International Edition: DOI: 10.1002/anie.201510313
German Edition: DOI: 10.1002/ange.201510313

Orthogonal Protein Assembly on DNA Nanostructures Using Relaxases

Sandra Sagredo, Tobias Pirzer, Ali Aghebat Rafat, Marisa A. Goetzfried, Gabriel Moncalian, Friedrich C. Simmel, and Fernando de la Cruz**

Abstract: *DNA-binding proteins are promising reagents for the sequence-specific modification of DNA-based nanostructures. Here, we investigate the utility of a series of relaxase proteins—TrwC, TraI, and MobA—for nanofunctionalization. Relaxases are involved in the conjugative transfer of plasmids between bacteria, and bind to their DNA target sites via a covalent phosphotyrosine linkage. We study the binding of the relaxases to two standard DNA origami structures—rodlike six-helix bundles and flat rectangular origami sheets. We find highly orthogonal binding of the proteins with binding yields of 40–50 % per binding site, which is comparable to other functionalization methods. The yields differ for the two origami structures and also depend on the position of the binding sites. Due to their specificity for a single-stranded DNA target, their orthogonality, and their binding properties, relaxases are a uniquely useful addition to the toolbox available for the modification of DNA nanostructures with proteins.*

Recent advances in DNA nanotechnology have resulted in methodologies that enable the generation of molecular objects made from DNA with almost any user-defined shape,[1] and these structures also facilitate the arrangement of functional molecular components or nanoparticles (e.g., proteins) into three-dimensional configurations with nanometer precision. One of the most popular methods is the DNA origami technique, which utilizes hundreds of short DNA oligomers, called "staples" to fold a long DNA scaffold into a desired shape.[1a,c]

A variety of approaches were previously utilized for the functionalization of DNA nanostructures with proteins. An often used modification is streptavidin, which binds to biotin with a dissociation constant of $K_D = 40$ fM,[2] constituting one of the strongest noncovalent bonds known in nature. In order to facilitate the binding of streptavidin to DNA nanostructures, biotinylated oligonucleotide staples are incorporated into the structures at the desired locations. The effective yield for streptavidin-binding on origami structures is roughly 85 % per binding site.[3] Naturally occurring DNA-binding proteins typically show dissociation constants in the nanomolar range and thus could also be used for modification of DNA nanostructures. A good example are zinc finger proteins, which bind to DNA in the presence of Zn^{2+} and result in effective binding yields of 30–70 % for DNA nanostructures.[4] Other proteins that have been used in DNA origami assemblies are DNA–antibody conjugates[5] and engineered proteins containing His-,[6] Snap-, and Halo-tags.[7] Other covalent conjugation techniques such as coupling by transglutaminase[8] or methyltransferases,[9] expressed protein ligation, and enzymatic ligation[10] also provide site-specific conjugation methods with controlled stoichiometry. Alternatively, chemical crosslinkers containing maleimide or *N*-hydroxysuccinimide groups are extensively used for nonspecific covalent coupling (for a more complete overview see Ref. [11]). The disadvantage of most of the aforementioned methods is that they are laborious and that the structure and thereby the activity of the proteins is compromised by the introduction of nonnative chemical tags.

In this work we present a novel strategy for DNA coupling based on relaxase proteins. Relaxases are involved in DNA processing for bacterial conjugation,[12] the horizontal transfer of genomic material among bacteria. They are single-strand (ss) DNA cleaving endonucleases, which naturally remain covalently bound to DNA after cleavage. Thus, relaxases offer a potentially superior alternative to existing methods for protein conjugation to DNA nanostructures, due to their sequence specificity, orthogonality and also their natural capacity to form covalent bonds with standard oligonucleotides. Relaxases belong to the HUH protein superfamily,[13] and the three-dimensional structure of some relaxases is known.[14] Their overall shape resembles a left hand, where the β-sheet "palm" holds the HUH motif (histidine–hydrophobic amino acid–histidine) responsible for metal coordination, and the α-helix "thumb" contains the catalytic tyrosine. These conserved residues form the active center of each relaxase and allow cleavage of a phosphodiester bond located at a specific plasmid DNA sequence called the *nic* site. Each plasmid contains a particular relaxase, recognizing a different

[*] S. Sagredo, Dr. G. Moncalian, Prof. Dr. F. de la Cruz
Departamento de Biología Molecular e Instituto de Biomedicina y Biotecnología de Cantabria, Universidad de Cantabria-Consejo Superior de Investigaciones Científicas-SODERCAN
Albert Einstein 22, 39011 Santander (Spain)
E-mail: delacruz@unican.es
Dr. T. Pirzer, A. Aghebat Rafat, M. A. Goetzfried, Prof. Dr. F. C. Simmel
Physik-Department E14 and ZNN/WSI, TU Munich
Am Coulombwall 4a, 85748 Garching (Germany)
E-mail: simmel@tum.de
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under http://dx.doi.org/10.1002/anie.201510313.

nic site, so the number of relaxase/*nic* site pairs is enormous. The cleavage reaction also occurs in vitro on ssDNA containing the *nic* region.[15] The *nic* region contains an inverted repeat (IR) and an AT-rich region downstream from *nic*. Upon binding ssDNA in vitro, the relaxase bends its substrate into a U shape and fits the *nic* sequence into its active center; here, the tyrosine generates a covalent 5′-phosphotyrosine intermediate and a free 3′-OH end (cf. Figure 1). In this state, cleavage and religation of substrate are in fact in a dynamic equilibrium.[15b,16] Each relaxase recognizes its own *nic* region with a K_D between 1 nM and 400 nM, depending on the individual protein and the target sequence.

In the following, we utilize the N-terminal domain of relaxases TrwC from conjugative plasmid R388 (denoted as $TrwC_{R388}$), TraI from R100 ($TraI_{R100}$), MobA from R1162 ($MobA_{R1162}$), and TraI from pKM101 ($TraI_{pKM101}$). All relaxase domains are small monomeric proteins (20–30 kDa) with low K_D[15a,17] that can be engineered and fused to other proteins such as fluorescent proteins (Figures S6, S7, and S9). $TraI_{R100}$ and $TrwC_{R388}$ are able to cleave proficiently the *nic* region lacking the distal arm of the IR. The reported K_d values of single-stranded R100 and R388 *nic* regions are 2.4 nM[17a] and 320 nM,[17b] respectively. It was shown that $MobA_{R1162}$ effectively binds targets with the entire IR, but binding was compromised after deletion of the distal arm.[15a] The relaxase $TraI_{pKM101}$ had not been characterized yet, but we expected a similar behavior as it has 51% and 37.7% sequence identity with $TrwC_{R388}$ and $TraI_{R100}$, respectively. Moreover, the pKM101 target is 64.7% identical to the R100 target. Comparison of targets with and without a complete IR showed that in the pKM101 target a complete IR is also needed (cf. Figure S8).

DNA origami structures used in this work were long, rodlike six-helix bundles (6HB) or flat, twist-corrected rectangular origami sheets (tcRO), with dimensions of 430 nm × 5 nm and 90 nm × 60 nm, respectively. The binding sites with the target sequence were designed with a double-stranded stem as a rigid linker in close proximity to the nanostructure followed by a single-stranded region with the *nic* site (cf. Figures 1 A and S10 for further details). Along its length, each 6HB is equipped with specific binding positions for the aforementioned four relaxases. In order to increase the local concentration of binding sites, each binding *position* (see Figure 1 B and the Supporting Information) was equipped with five binding *sites* each for the specific relaxase. Due to the small size of the relaxase domains, we used fusion proteins of relaxases with fluorescent proteins mKATE or mCFP (45–62 kDa). These fused proteins resulted in increased visibility in TEM images. Each relaxase was incubated for 24 h with 6HB nanotubes containing all four binding positions. After incubation and purification, the number of occupied binding positions was determined from TEM images (Figure 2). The resulting binding yields *per position* were 93% for $TraI_{R100m}$mKATE, 84% for $TraI_{pKM101}$mCFP, 73% for $TrwC_{R388}$mCFP, and 34% for $MobA_{R1162}$mCFP.

Since the 6HB contained all four binding positions, quantitative values for the specificity of each relaxase to its target could be measured (see Figure 2). Proteins $TraI_{R100}$mKATE, $TrwC_{R388}$mCFP, and $MobA_{R1162}$mCFP show

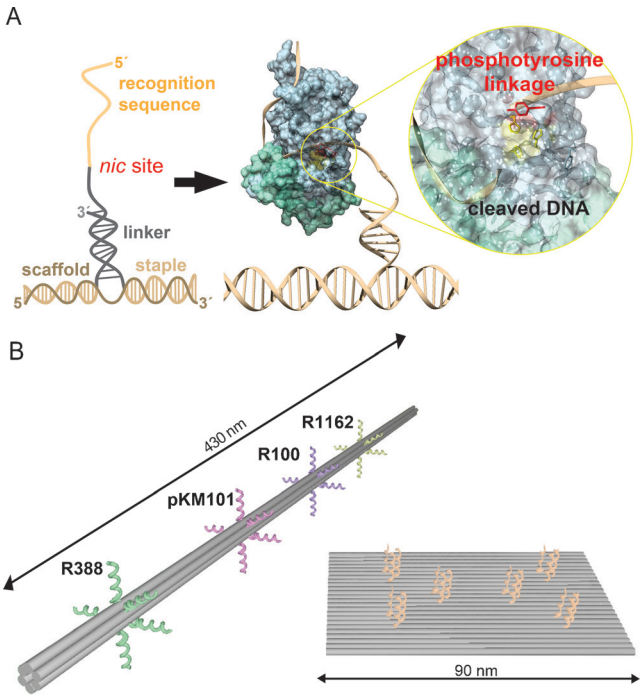

Figure 1. Strategy for the coupling of relaxases onto DNA nanostructures. A) Extended staples contain the recognition sequence of a relaxase and several base pairs that allow the target to protrude from the nanostructure. Binding of the relaxase to its cognate site involves bending of the ssDNA to fit the *nic* site into the relaxase active center. Inset: the catalytic tyrosine carries out a transesterification reaction that links it covalently to the 5′-phosphate between the scissile nucleotides. As a result, the relaxase becomes covalently bound to the extended staple. B) DNA nanostructures used in this approach to study the coupling of four relaxases.

highly selective binding abilities, with unspecific binding yields ranging from 0% to 7.1% (cf. Figure 2). Despite its high yield for its specific target, $TraI_{pKM101}$mCFP shows significant nonspecific binding and therefore was not used simultaneously with the other relaxases in the following.

In further experiments we used the twist-corrected rectangular origami sheet tcRO, for which we designed six binding positions with up to three redundant binding sites each (cf. Figure 3 A). We first compared the binding of the orthogonal relaxases $TraI_{R100}$mKATE and $TrwC_{R388}$mCFP onto the origami sheets. To test their binding yields, only position 1 for $TraI_{R100}$mKATE and position 5 for $TrwC_{R388}$mCFP, both equivalently at the rim of the tcRO next to the corners (cf. Figure 3), were used. Each protein was incubated for 24 h with the tcRO, each sheet containing two binding sites per position for each protein, and purified afterwards (see the Methods section in the Supporting Information for details). For visualization of the bound proteins we used atomic force microscopy (AFM) imaging (Figure 3). The binding yield of $TraI_{R100}$mKATE was found to be 70%, and 48% for $TrwC_{R388}$mCFP. Nonspecific binding to the unwanted position was 1% for $TraI_{R100}$mKATE and 3% for $TrwC_{R388}$mCFP.

We then checked for a potential position dependence of relaxase binding. To this end, the tcRO was equipped with all

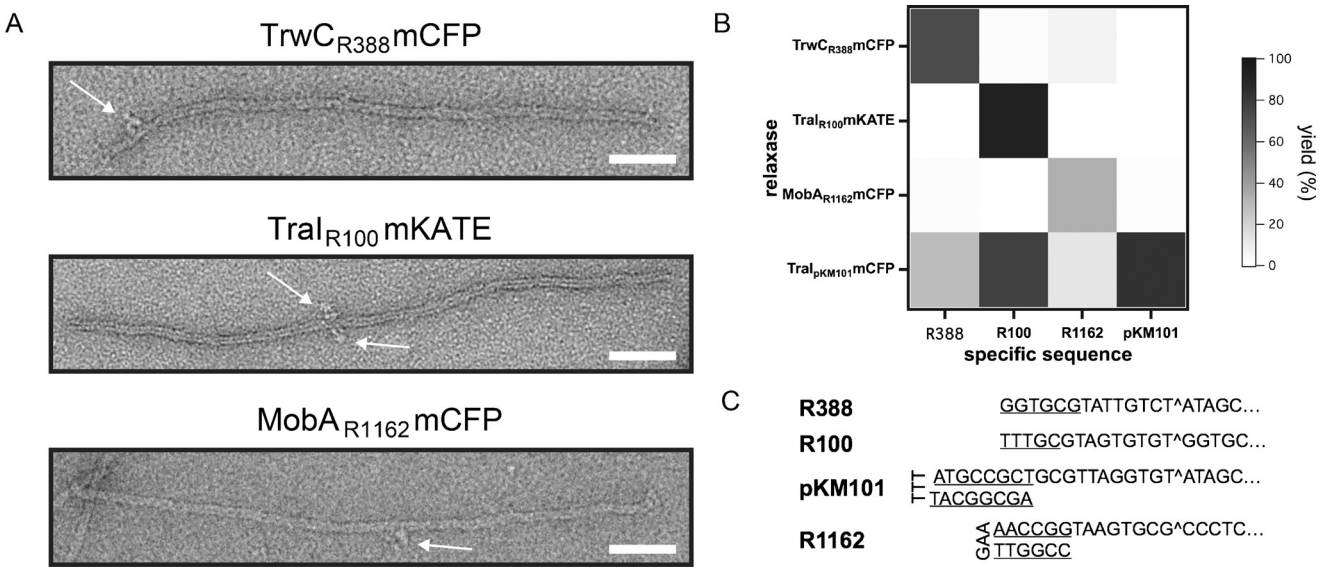

Figure 2. Orthogonal binding of relaxases to six-helix bundles (6HB). A) TEM images of the relaxases TrwC$_{R388}$mCFP, TraI$_{R100}$mKATE, and MobA$_{R1162}$mCFP bound to 6HB. Scale bar: 50 nm. B) Graphical representation of total binding yields and their orthogonality. For details on the statistical analysis see the Supporting Information. C) Recognition sequences of the relaxases used in this study. The *nic* site is depicted with ^ and the proximal arm of the inverted repeat is underlined.

designed positions containing only R388 binding sites. From these experiments, the binding yield at the tcRO rim (positions 1, 2, 5, and 6 in Figure 3A) was found to range between 40% and 60%, whereas it appeared to be significantly higher (between 60% and 90%) in the tcRO center (positions 3 and 4, cf. Figure 3C). In general, binding yields varied considerably from experiment to experiment (±10%). Surprisingly, within the range of this variation, we found no clear dependence of the binding yield on the number of binding sites per binding position. When only a single binding site for R388 was used, binding yields were as high as for multiple binding sites (between 40–50%, see the Supporting Information).

Overall, the binding yields obtained by AFM were significantly lower than that expected based on the law of mass action (see the Supporting Information for details). For instance, in previous experiments with single-stranded targets containing the *nic* site, the K_D for the binding of TrwC$_{R388}$ to its recognition sequence was determined to be 320 nM.[17b] This should translate into an *equilibrium* binding yield of 86% for a single binding site and even 99.7% for three binding sites under our incubation conditions, where TrwC$_{R388}$ and tcRO concentrations were ca. 2 μM and 100 nM, respectively. As AFM imaging is performed with purified samples in buffer (i.e., without relaxases), however, re-equilibration of the sample will result in unbinding of the proteins from the target and thus a reduced apparent binding yield. This is essentially unavoidable for relatively high K_D values in the nanomolar range.

Since the covalent phosphotyrosine linkage of a relaxase to its *nic* site is in dynamic equilibrium with the unbound state, we also used phosphorothioate suicide oligonucleotides to shift the equilibrium towards the covalent bond (and thus achieve a much lower effective K_D). In suicide oligonucleotides, the 3′ oxygen is replaced by sulfur at the scissile phosphodiester bond, preventing the religation reaction.[18] Surprisingly, the binding yield did not improve with these oligonucleotides (see the Supporting Information for details). The strong coupling theoretically expected for a covalent bond is thus not reflected in the actual binding yields.

Our findings are in line with results for other protein-modification approaches for DNA origami. Streptavidin binding to biotin showed only about 85%[3] binding yield (with a theoretical value of close to 100%) and also for covalent coupling methods only 84–90% experimental binding yields were obtained.[3] Apart from the re-equilibration issue mentioned above, binding yields on origami structures may be reduced by a variety of other factors. Some target sites may be inaccessible or simply absent (e.g., because of missing staples), and diffusion to the target may be affected by the geometry of the binding site or the overall shape and flexibility of the underlying origami structure. For instance, in our experiments with six-helix bundles, the binding sites were accessible from all sides, and we typically observed higher binding yields than for the tcRO structures.

In summary, we have introduced DNA-binding relaxase proteins as promising reagents for the sequence-addressable modification of DNA nanostructures with proteins. Binding yields on DNA origami structures ranged between 40–50% for a single binding site, and were thus comparable to those of other modification strategies. Depending on the origami geometry, the specific relaxase, and also the utilization of redundant binding sites in close proximity, in some cases apparent binding yields as high as 90% could be achieved. Upon further optimization it should be possible to fully exploit the covalent phosphotyrosine linkage between the relaxases and the DNA target site, and thus achieve even higher binding yields. At least two of the four investigated

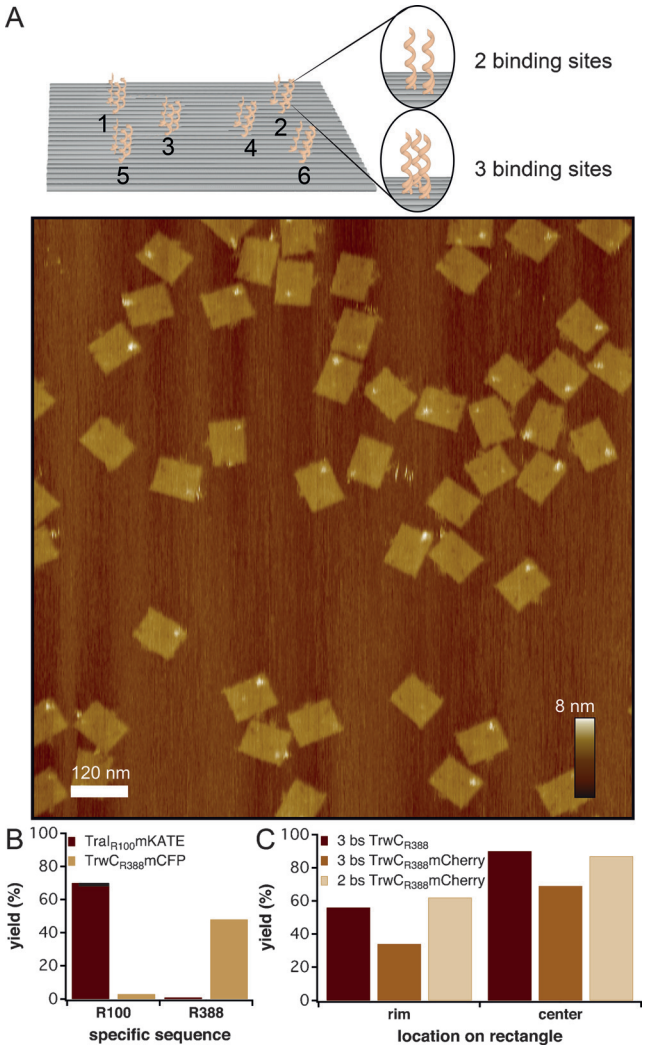

Figure 3. Position-dependent binding of relaxases to the rectangular DNA sheet tcRO. A) Top: Illustration of the six binding positions with two or three binding sites each. Bottom: AFM image of the nanostructures decorated with TraIR100mKATE at position 1. B) Binding yields of TrwCR388mCFP and TraIR100mKATE to available sites. The relaxases bind preferably to their specifc target. C) Binding yield of TrwCR388 to tcRO equipped with two and three targets per position. Binding positions at the rim of the tcRO (1, 2, 5, and 6) show lower yields than in the center (3 and 4). For details on the statistical analysis see the Supporting Information.

relaxases (TraI from plasmid R100 and TrwC from R388) showed very good orthogonal binding performance. Relaxases can be easily engineered and fused to other proteins of choice. As many relaxases from different sources and with unique recognition sequences are available, they potentially can be developed into a whole new class of orthogonal, sequence-selective protein linkers for DNA nanotechnology. Of particular interest could be their application for the biological generation of DNA–protein hybrid nanostructures and for the modification of DNA nanostructures in vivo.

Acknowledgements

We gratefully acknowledge financial support through the Volkswagen Foundation (grant 86 395-1), the European Commission (FP7 grant no. 248919 (BACTOCOM)), the DFG (SFB 1032 TP A2 and Nanosystems Initiative Munich), and EMBOEMBO. We thank F. Praetorius for providing the M13 origami scaffold and A. Kuzyk for initial work on this project.

Keywords: bacterial conjugation · bioconjugation · DNA nanotechnology · DNA origami · relaxase

How to cite: *Angew. Chem. Int. Ed.* **2016**, *55*, 4348–4352
Angew. Chem. **2016**, *128*, 4421–4425

[1] a) P. W. K. Rothemund, *Nature* **2006**, *440*, 297–302; b) S. M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf, W. M. Shih, *Nature* **2009**, *459*, 414–418; c) B. Kick, F. Praetorius, H. Dietz, D. Weuster-Botz, *Nano Lett.* **2015**, *15*, 4672–4676.
[2] N. M. Green, *Methods Enzymol.* **1990**, *184*, 51–67.
[3] N. V. Voigt, T. Tørring, A. Rotaru, M. F. Jacobsen, J. B. Ravnsbæk, R. Subramani, W. Mamdouh, J. Kjems, A. Mokhir, F. Besenbacher, K. V. Gothelf, *Nat. Nanotechnol.* **2010**, *5*, 200–203.
[4] E. Nakata, F. F. Liew, C. Uwatoko, S. Kiyonaka, Y. Mori, Y. Katsuda, M. Endo, H. Sugiyama, T. Morii, *Angew. Chem. Int. Ed.* **2012**, *51*, 2421–2424; *Angew. Chem.* **2012**, *124*, 2471–2474.
[5] a) T. Yamazaki, J. G. Heddle, A. Kuzuya, M. Komiyama, *Nanoscale* **2014**, *6*, 9122–9126; b) A. Shaw, E. Benson, B. Högberg, *ACS Nano* **2015**, *9*, 4968–4975.
[6] a) R. P. Goodman, C. M. Erben, J. Malo, W. M. Ho, M. L. McKee, A. N. Kapanidis, A. J. Turberfield, *ChemBioChem* **2009**, *10*, 1551–1557; b) W. Shen, H. Zhong, D. Neff, M. L. Norton, *J. Am. Chem. Soc.* **2009**, *131*, 6660–6661.
[7] B. Saccà, R. Meyer, M. Erkelenz, K. Kiko, A. Arndt, H. Schroeder, K. S. Rabe, C. M. Niemeyer, *Angew. Chem. Int. Ed.* **2010**, *49*, 9378–9383; *Angew. Chem.* **2010**, *122*, 9568–9573.
[8] J. Tominaga, Y. Kemori, Y. Tanaka, T. Maruyama, N. Kamiya, M. Goto, *Chem. Commun.* **2007**, 401–403.
[9] V. G. Metelev, E. A. Kubareva, O. V. Vorob'eva, A. S. Romanenkov, T. S. Oretskaya, *FEBS Lett.* **2003**, *538*, 48–52.
[10] a) S. Takeda, S. Tsukiji, T. Nagamune, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2407–2410; b) R. S. Sørensen, A. H. Okholm, D. Schaffert, A. L. B. Kodal, K. V. Gothelf, J. Kjems, *ACS Nano* **2013**, *7*, 8098–8104.
[11] a) Y. R. Yang, Y. Liu, H. Yan, *Bioconjugate Chem.* **2015**, *26*, 1381–1395; b) B. Saccà, C. M. Niemeyer, *Chem. Soc. Rev.* **2011**, *40*, 5910–5921.
[12] F. de la Cruz, L. S. Frost, R. J. Meyer, E. L. Zechner, *FEMS Microbiol. Rev.* **2010**, *34*, 18–40.
[13] M. Chandler, F. de la Cruz, F. Dyda, A. B. Hickman, G. Moncalián, B. Ton-Hoang, *Nat. Rev. Microbiol.* **2013**, *11*, 525–538.
[14] a) A. Guasch, M. Lucas, G. Moncalián, M. Cabezas, R. Pérez-Luque, F. X. Gomis-Rüth, F. de la Cruz, M. Coll, *Nat. Struct. Biol.* **2003**, *10*, 1002–1010; b) S. Datta, C. Larkin, J. F. Schildbach, *Structure* **2003**, *11*, 1369–1379; c) A. F. Monzingo, A. Ozburn, S. Xia, R. J. Meyer, J. D. Robertus, *J. Mol. Biol.* **2007**, *366*, 165–178; d) J. S. Edwards, L. Betts, M. L. Frazier, R. M. Pollet, S. M. Kwong, W. G. Walton, W. K. Ballentine, J. J. Huang, S. Habibi, M. Del Campo, J. L. Meier, P. B. Dervan, N. Firth, M. R. Redinbo, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 2804–2809; e) R. P. Nash, S. Habibi, Y. Cheng, S. A. Lujan, M. R. Redinbo, *Nucleic Acids Res.* **2010**, *38*, 5929–5943.

[15] a) M. K. Bhattacharjee, R. J. Meyer, *Nucleic Acids Res.* **1993**, *21*, 4563 – 4568; b) M. Llosa, G. Grandoso, F. de la Cruz, *J. Mol. Biol.* **1995**, *246*, 54 – 62.
[16] W. Pansegrau, E. Lanka, *J. Biol. Chem.* **1996**, *271*, 13068 – 13076.
[17] a) M. J. Harley, J. F. Schildbach, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 11243 – 11248; b) M. Lucas, B. Gonzalez-Perez, M. Cabezas, G. Moncalian, G. Rivas, F. de la Cruz, *J. Biol. Chem.* **2010**, *285*, 8918 – 8926.
[18] B. González-Pérez, M. Lucas, L. A. Cooke, J. S. Vyle, F. de la Cruz, G. Moncalián, *EMBO J.* **2007**, *26*, 3847 – 3857.

Received: November 5, 2015
Revised: December 17, 2015
Published online: February 24, 2016

