

# A Structure-Guided Switch in the Regioselectivity of a Tryptophan Halogenase

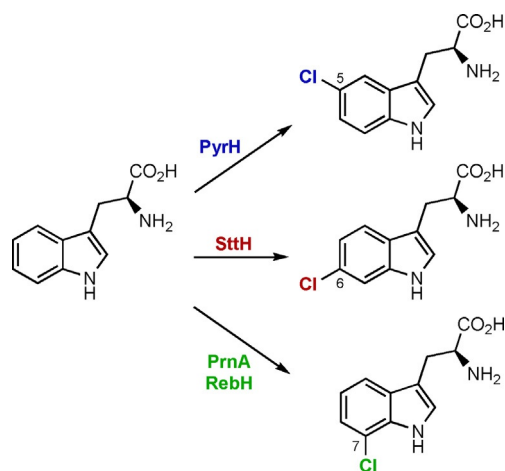
Sarah A. Shepherd, Binuraj R. K. Menon, Heidi Fisk, Anna-Winona Struck, Colin Levy, David Leys, and Jason Micklefield<sup>\*[a]</sup>

Flavin-dependent halogenases are potentially useful biocatalysts for the regioselective halogenation of aromatic compounds. Haloaromatic compounds can be utilised in the synthesis and biosynthesis of pharmaceuticals and other valuable products. Here we report the first X-ray crystal structure of a tryptophan 6-halogenase (SttH), which enabled key residues that contribute to the regioselectivity in tryptophan halogenases to be identified. Structure-guided mutagenesis resulted in a triple mutant (L460F/P461E/P462T) that exhibited a complete switch in regioselectivity; with the substrate 3-indolepropionate 75% 5-chlorination was observed with the mutant in comparison to 90% 6-chlorination for the wild-type SttH. This is the first clear example of how regiocomplementary halogenases can be created from a single parent enzyme. The biocatalytic repertoire of SttH was also expanded to include a range of indolic and non-indolic substrates.

Enzymes that can catalyse the regioselective halogenation of aromatic substrates could provide an attractive alternative to the traditional halogenation methods that are commonly used in synthesis. Halogenated aromatic compounds find extensive synthetic applications, particularly in transition-metal-catalysed crosscoupling reactions,<sup>[1]</sup> and are important constituents of pharmaceuticals,<sup>[2]</sup> agrochemicals<sup>[3]</sup> and other valuable materials.<sup>[4]</sup> Despite this, the traditional methods of producing haloaromatic compounds utilise harsh reaction conditions and often require harmful reagents, catalysts and solvents. These non-enzymatic methods also lack regiocontrol resulting in unwanted by-products that can be difficult to separate and are problematic to dispose of owing to their toxicity or persistence in the environment.<sup>[5]</sup> Consequently, there has been major interest in harnessing nature's halogenases, which employ benign inorganic halides in aqueous media, to effect cleaner and more regioselective halogenation reactions.

The first halogenating enzyme to be identified was the chloroperoxidase from the fungus, *Caldariomyces fumago*, which lacked regiocontrol owing to the free hypochlorous acid (HOCl) produced as the halogenating agent. Further examples of both haem- and vanadium-dependent haloperoxidases were later identified, which also generally lacked substrate specificity and regioselectivity.<sup>[6]</sup> More recently, Fe<sup>2+</sup>/α-ketoglutarate (αKG)-dependent and flavin-dependent halogenases, which effect the regioselective halogenation of precursors in the biosynthesis of a wide range of halogenated natural products, have been identified.<sup>[7]</sup> Many of the αKG- and flavin-dependent halogenases utilise substrates that are tethered to the carrier proteins of biosynthetic assembly-line enzymes thus making their use as biocatalysts limited. However, there are a number of flavin-dependent tryptophan halogenases that can regioselectively halogenate free tryptophan, and therefore have more potential for synthetic purposes.<sup>[8]</sup>

Previously, X-ray structures have been elucidated for the tryptophan 7-halogenases, PrnA<sup>[9]</sup> and RebH,<sup>[10]</sup> as well as a tryptophan 5-halogenase PyrH (Scheme 1).<sup>[11]</sup> These structures provided insights into the mechanism and regiocontrol of flavin-dependent halogenases. It is suggested, that the halogenases utilise O<sub>2</sub> to oxidise FADH<sub>2</sub> giving C4a-hydroperoxyflavin, which then reacts with chloride to produce HOCl. It is then proposed that HOCl reacts with an active site lysine to generate a chloramine, which chlorinates the substrate.<sup>[9,12]</sup> Although the position of the active-site lysine relative to the substrate is likely to be important in determining the regiochemical out-



**Scheme 1.** Reactions of flavin-dependent halogenases with tryptophan, and their respective products.

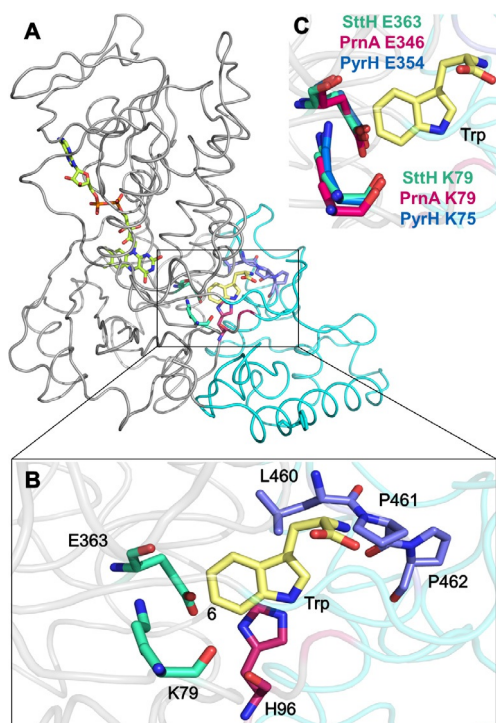
[a] Dr. S. A. Shepherd, Dr. B. R. K. Menon, H. Fisk, Dr. A.-W. Struck, Dr. C. Levy, Prof. D. Leys, Prof. J. Micklefield  
School of Chemistry and Manchester Institute of Biotechnology  
The University of Manchester  
131 Princess Street, Manchester, M1 7DN (UK)  
E-mail: Jason.micklefield@manchester.ac.uk

Supporting information for this article can be found under <http://dx.doi.org/10.1002/cbic.201600051>.

© 2016 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

come of the reactions, the factors that effect the regiocontrol of these enzymes are still not fully understood. For example, using the PrnA and PyrH X-ray crystal structures,<sup>[9,11]</sup> Lang et al. attempted to switch the regioselectivity of PrnA to that of PyrH by targeted mutagenesis.<sup>[13]</sup> However, of all the mutants tested only one mutation, F103A, had any effect on the regioselectivity of PrnA. The F103A mutant showed a modest change in regioselectivity with bromide giving a 2:1 mixture of 7- and 5-bromotryptophan, whereas the wild-type PrnA gives exclusively 7-bromotryptophan. This shift to produce 33% 5-bromotryptophan falls some way short of the change that would be required to create a new regiocomplementary enzyme. Here we describe the first structure of a tryptophan 6-halogenase, SttH, which provides further insights into the factors affecting the regioselectivity of these flavin-dependent halogenases. Moreover, these structural insights were used to expand the biocatalytic repertoire of this enzyme, and to guide mutagenesis leading to a complete switch in the regioselectivity from 90% chlorination at the 6-position to 75% in favour of chlorination at the 5-position of 3-indolepropionic acid.

Given that there is no structure for a tryptophan 6-halogenase, we explored expression of a number of candidate enzymes for crystallography trials. From this, we found that SttH from *Streptomyces toxytricini* gave good expression in *Escherichia coli* and catalysed the halogenation of tryptophan to give exclusively 6-chlorotryptophan, as reported previously.<sup>[14]</sup> The X-ray structure of SttH was then determined at 2.7 Å (Figure 1)

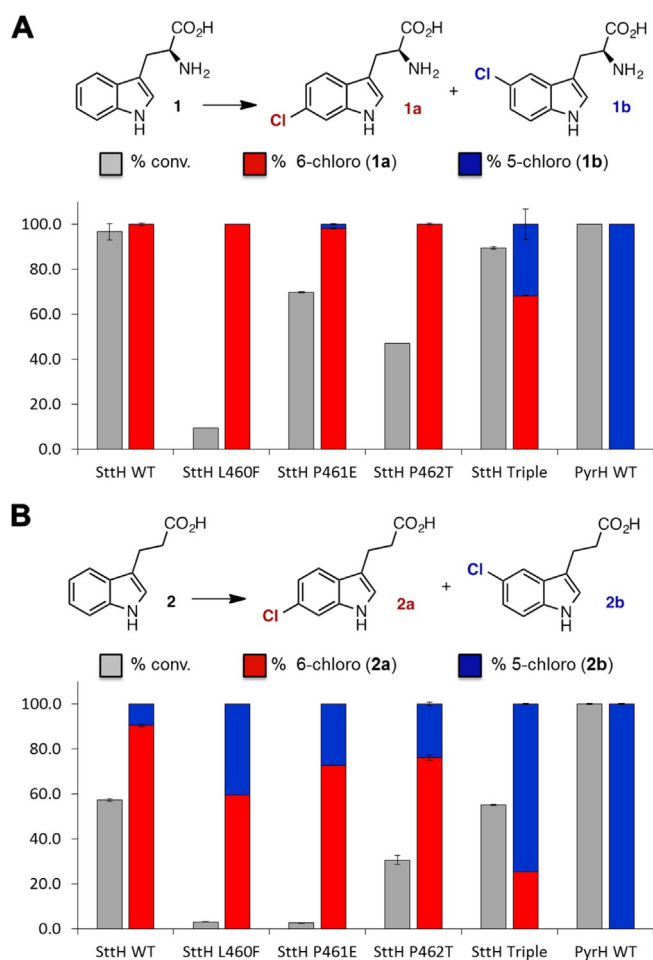


**Figure 1.** A) Crystal structure of SttH (PDB ID: 5HY5) showing typical box (grey) and triangular pyramid (cyan) of flavin-dependent tryptophan halogenases. B) Selected active-site residues of SttH with tryptophan modelled. C) Overlay of catalytic lysine and glutamate in PyrH, SttH and PrnA.

to reveal a dimer, with each monomer exhibiting a box and triangular pyramid structure, as previously observed with PrnA, RebH and PyrH.<sup>[9–11]</sup> Within the box structure are the conserved flavin-dependent tryptophan halogenase sequences GxGxxG and WxWxIP. At the interface with the triangular pyramid are the catalytic residues K79 and E363, which align with the active-site lysine and glutamate of PrnA and PyrH (Figure 1C).<sup>[14,15]</sup> In addition, SttH residues H96 and F98 are positioned for  $\pi$ -stacking with the indole moiety of the substrate (Figure S2), whilst the P97 carbonyl and Y463 hydroxy groups can potentially hydrogen bond with the indole NH (Figure S3).

From sequence alignments it is evident that SttH is more like PyrH than PrnA, with insertions present in PyrH and SttH between residues SttH 155 and 167 and a deletion between SttH 457 and 464 compared with PrnA (Figure S1). These subtle differences around the active site of the enzymes lead to alterations in the binding mode of tryptophan and effect the regiocontrol observed with these enzymes. By comparing the apo structures of PyrH and SttH, it can also be noted that many of the other active-site residues are very closely aligned, including sequences of residues such as QFPYAYHF (SttH residues 171–178) and PYYHGxxxYS (SttH residues 455–464). Other than residues between SttH G148 and G167, which lack electron density in the SttH structure, the only differences evident in the active-site region between the structures of PyrH and SttH are those of PyrH residues F451, E452 and T453 and SttH L460, P461 and P462. These residues are of particular interest because they are in close proximity to the active site in PyrH and SttH, and are positioned directly above the  $\alpha$ -amino acid moiety of the substrate, tryptophan (**1**). Moreover, there is a loop insertion in PrnA in this region that is suggested to contribute to its regioselectivity.<sup>[11]</sup> Each of these residues was mutated in SttH to the corresponding residue in PyrH, that is, L460F, P461E and P462T. Individually, each mutation reduced the relative activity of the enzyme with **1**, but did not have a significant effect on the observed regioselectivity, with 6-chlorotryptophan (**1a**) remaining the major product (Figure 2). Interestingly however, the triple mutant SttH L460F/P461E/P462T exhibited similar activity to the wild-type SttH, with tryptophan as a substrate, but produced 32% 5-chlorotryptophan (**1b**) and 68% 6-chlorotryptophan, whereas the wild-type SttH only produces the 6-chlorinated product.

A second substrate 3-indolepropionic acid (**2**), which lacks the amino group of tryptophan and therefore has more flexibility in the active site owing to the absence of a potential interaction with the backbone carbonyl of SttH G459 (Figure S4), causes a more significant shift in regioselectivity than that of the tryptophan (Figure 2). With wild-type SttH 10% 5-chloro-3-indolepropionic acid (**2b**) is produced and 90% 6-chloro-3-indolepropionic acid (**2a**), whereas the SttH triple mutant produces 75% **2b** and 25% **2a**. Notably, the relative activity of the SttH triple mutant was similar to that of the wild-type enzyme, with 3-indolepropionic acid (**2**) as substrate, thus indicating that the complete switch in regioselectivity can be achieved without impacting on catalytic efficiency. The wild-type PyrH enzyme, with either **1** or **2** as a substrate, produced exclusively 5-chlorinated products. In an effort to switch the regioselectivity



**Figure 2.** Percentage conversion of A) tryptophan or B) 3-indolepropionic acid with SttH wild type, single and triple mutants, as well as PyrH wild type. Purified halogenase (10  $\mu\text{M}$ ) was incubated with agitation at 30 °C for 1 h with Fre (1  $\mu\text{M}$ ), GDH2 (6  $\mu\text{M}$ ), FAD (7.5  $\mu\text{M}$ ), NADH (200  $\mu\text{M}$ ),  $\text{MgCl}_2$  (50 mM), glucose (20 mM) and substrate (0.5 mM) in a total volume of 100  $\mu\text{L}$  in potassium phosphate buffer (10 mM, pH 7.0).

ty of PyrH from 5- to 6-halogenation, the corresponding PyrH triple mutant (F451L/E452P/T453P) was generated with the corresponding residues observed in SttH. However, this triple mutant was found to be inactive with both substrates **1** and **2**.

Previous studies have indicated that PrnA<sup>[16]</sup> and RebH<sup>[17]</sup> can halogenate indolic substrates, as well as tryptophan. Here we investigate the substrate specificity of SttH with *N*-methyltryptophan (**3**) in addition to non-indolic aromatic substrates such as kynurenine (**4**), anthranilamide (**5**) and other anilines (Table 1). As with the natural substrate **1**, halogenation occurs solely at the 6-position of **3** resulting in 6-chloro product (**3a**). However, upon moving to the non-indolic substrate **4**, chlorination did not occur at the 4-position, *meta* to the amino group, as might be expected. Instead kynurenine is chlorinated at the intrinsically more reactive 5-position by SttH. This suggests the greater flexibility of the kynurenine side chain, and perhaps also reduced  $\pi$ -stacking interactions with H96 and F98 (Figure S2) compared with tryptophan, enables the active-site lysine residue, K79, to deliver the electrophilic chloroamine to the most reactive 5-position, *para* to the amino group. Presum-

**Table 1.** Conversion of substrates after 1 hour with SttH wild type.

Substrate	Conversion [%]	Product
tryptophan ( <b>1</b> )	97 $\pm$ 4	<b>1a</b>
3-indolepropionic acid ( <b>2</b> )	57 $\pm$ 1.5	<b>2a</b> : <b>2b</b> (9:1)
<i>N</i> -methyltryptophan ( <b>3</b> )	68 $\pm$ 3	<b>3a</b>
kynurenine ( <b>4</b> )	79 $\pm$ 4	<b>4a</b>
anthranilamide ( <b>5</b> )	43 $\pm$ 2	<b>5a</b>
anthranilic acid ( <b>6</b> )	1.1 $\pm$ 0.1	<b>6a</b>
<i>N</i> -phenylanthranilic acid ( <b>7</b> )	20 $\pm$ 2	<b>7a</b>

Red indicates conversion to expected product. Blue indicates conversion to chemically favoured products. Assay conditions shown in Figure 2.

ably, the subtle differences in the active-site architecture of SttH compared with PyrH are not sufficient to prevent the movement of kynurenine aryl group, so the more electronically favoured *para*-chlorination reaction predominates. The same regioselectivity is also evident with the smaller aromatic substrates **5** and anthranilic acid (**6**), which are both solely chlorinated at the 5-position. Finally, the biaryl compound *N*-phenylanthranilic acid (**7**) is also halogenated by SttH, thus showing the potential for the halogenation of larger aromatic compounds with this enzyme.

When comparing the activity of SttH with various substrates (Table 1), it is unsurprising that tryptophan was the best substrate. The addition of a methyl group to the nitrogen of the tryptophan indole ring (**3**) reduced activity. The activity was further decreased by the loss of the  $\alpha$ -amino group (**2**) and this also led to reduced regiocontrol, about 90% 6-chlorination observed (Figure 2). With non-indolic substrates, the most electronically favoured products were produced with **4** (with its similar side chain to tryptophan) displaying close to 80% conversion. Anilines **5** and **7** also displayed good conversion; however, upon switching the amide of **5** to the acid of **6**, activity is severely reduced.

From the kinetics of the selected substrates **1**, **4** and **5** (Table 2), it is evident that substrate binding has a significant

**Table 2.** Kinetics of selected substrates with SttH wild type.

Substrate	$K_m$ [ $\mu\text{M}$ ]	$k_{cat}$ [ $\text{min}^{-1}$ ]	$k_{cat}/K_m \times 10^{-3}$ [ $\text{min}^{-1} \mu\text{M}^{-1}$ ]
tryptophan ( <b>1</b> )	0.8 $\pm$ 0.1	0.65 $\pm$ 0.02	825 $\pm$ 95
kynurenine ( <b>4</b> )	241 $\pm$ 32	0.51 $\pm$ 0.02	2.1 $\pm$ 0.30
anthranilamide ( <b>5</b> )	1075 $\pm$ 154	1.21 $\pm$ 0.04	1.1 $\pm$ 0.17

effect on the overall catalytic efficiency of the enzyme. Generally, the turnover varies between 0.6 and 1.2 min<sup>-1</sup>; however, the  $K_m$  varied between 0.8  $\mu\text{M}$  for tryptophan and 1 mM for anthranilamide. Presumably kynurenine has lower affinity for the active site, compared with tryptophan, owing to greater side chain flexibility and reduced  $\pi$ -stacking interactions; this is consistent with the observed regioselectivity. In addition to this, anthranilamide also loses contacts with residues S54 and Q171, which are likely to bind the  $\alpha$ -amino acid moiety of **1** and **4**; this leads to even lower binding affinity (Figure S4).

Previously Frese et al. demonstrated that crosslinked enzyme aggregates (CLEAs) incorporating the 7-chlorotryptophan halogenase RebH can be used to chlorinate the natural substrate tryptophan on a gram scale.<sup>[18]</sup> By applying this method, CLEAs of SttH were produced and used to halogenate the unnatural substrate anthranilamide on a 100 mg scale (isolated yield 25%). This could be improved by optimisation and catalyst recycling.

In summary, the first crystal structure of a tryptophan 6-halogenase (SttH) has been determined. By comparing the structure of SttH with those of other halogenases, including PyrH, it is clear how subtle differences in the active site,  $\pi$ -stacking interactions and contacts to the  $\alpha$ -amino acid moiety can alter the position of the aromatic moiety relative to the catalytic Lys residue thereby affecting the orientation of the subsequent electrophilic substitution reaction. The observed structural differences between the halogenases were exploited to create a SttH triple mutant, L460F/P461E/P462T, which showed the first complete switch in regioselectivity of this class of enzymes: with 3-indolepropionate as substrate, wild-type SttH gives 6-chloro-3-indolepropionate, whereas 5-chloro-3-indolepropionate was predominately produced by the triple mutant. The new regiocomplementary SttH variant displayed similar activity to the wild-type enzyme. Further assays revealed an additional five substrates that can be regioselectively halogenated by SttH, and with CLEAs, the halogenase can be stabilised for use on a preparative scale. Taken together, these results provide guidance for future efforts to engineer regiocomplementary halogenases for a wider range of aromatic substrates of synthetic utility.<sup>[17–19]</sup>

## Acknowledgements

We acknowledge the BBSRC (grants BB/K00199X/1 and BB/I008055/1), GlaxoSmithKline and CoEBio3 for support. We thank Diamond Light Source for access to beamline i24 (MX-7146), which contributed to the results presented here.

**Keywords:** aryl halides • biocatalysis • halogenation • mutagenesis • regioselectivity

- [1] a) S. Z. Tasker, E. A. Standley, T. F. Jamison, *Nature* **2014**, *509*, 299–309; b) J. F. Hartwig, *Nature* **2008**, *455*, 314–322; c) F.-S. Han, *Chem. Soc. Rev.* **2013**, *42*, 5270–5298; d) E. J. Cho, T. D. Senecal, T. Kinzel, Y. Zhang, D. A. Watson, S. L. Buchwald, *Science* **2010**, *328*, 1679–1681; e) K. C. Nicolaou, P. G. Bulger, D. Sarlah, *Angew. Chem. Int. Ed.* **2005**, *44*, 4442–4489; *Angew. Chem.* **2005**, *117*, 4516–4563.
- [2] Y. Lu, Y. Liu, Z. Xu, H. Li, H. Liu, W. Zhu, *Expert Opin. Drug Discovery* **2012**, *7*, 375–383.
- [3] P. Jeschke, *Pest Manage. Sci.* **2010**, *66*, 10–27.
- [4] O. Bolton, K. Lee, H.-J. Kim, K. Y. Lin, J. Kim, *Nat. Chem.* **2011**, *3*, 205–210.
- [5] a) F. Alonso, I. P. Beletskaya, M. Yus, *Chem. Rev.* **2002**, *102*, 4009–4091; b) S. A. Rowlands, A. K. Hall, P. G. McCormick, R. Street, R. J. Hart, G. F. Ebell, P. Donecker, *Nature* **1994**, *367*, 223; c) B. C. Kelly, M. G. Ikononou, J. D. Blair, A. E. Morin, F. A. P. C. Gobas, *Science* **2007**, *317*, 236–239; d) B. Xu, H.-Z. Zhu, Y.-L. Lin, K.-Y. Shen, W.-H. Chu, C.-Y. Hu, K.-N. Tian, S. A. Mwakaganda, X.-Y. Bi, *Water Air Soil Pollut.* **2012**, *223*, 4429–4436; e) J. M. Sufliata, J. A. Robinson, J. M. Tiedje, *Appl. Environ. Microbiol.* **1983**, *45*, 1466–1473.
- [6] a) K.-H. van Pée, C. Dong, S. Flecks, J. Naismith, E. P. Patallo, T. Wage, *Adv. Appl. Microbiol.* **2006**, *59*, 127–157; b) P. Bernhardt, T. Okino, J. M. Winter, A. Miyanaga, B. S. Moore, *J. Am. Chem. Soc.* **2011**, *133*, 4268–4270.
- [7] a) A. Butler, M. Sandy, *Nature* **2009**, *460*, 848–854; b) C. S. Neumann, D. G. Fujimori, C. T. Walsh, *Chem. Biol.* **2008**, *15*, 99–109; c) P. Gao, Y. Huang, *Appl. Environ. Microbiol.* **2009**, *75*, 4813–4820; d) S. Cadel-Six, C. Duga, A. M. Castets, R. Rippka, C. Bouchier, N. T. de Marsac, M. Welker, *Mol. Biol. Evol.* **2008**, *25*, 2031–2041.
- [8] a) D. Khare, B. Wang, L. Gu, J. Razelun, D. H. Sherman, W. H. Gerwick, K. Håkansson, J. L. Smith, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 14099–14104; b) P. C. Dorrestein, E. Yeh, S. Garneau-Tsodikova, N. L. Kelleher, C. T. Walsh, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13843–13848.
- [9] C. Dong, S. Flecks, S. Unversucht, C. Haupt, K.-H. van Pée, J. H. Naismith, *Science* **2005**, *309*, 2216–2219.
- [10] E. Bitto, Y. Huang, C. Bingman, *Proteins Struct. Funct. Bioinf.* **2008**, 289–293.
- [11] X. Zhu, W. De Laurentis, K. Leang, J. Herrmann, K. Ihlefeld, K.-H. van Pée, J. H. Naismith, *J. Mol. Biol.* **2009**, *391*, 74–85.
- [12] a) E. Yeh, L. Cole, E. Barr, J. Bollinger, *Biochemistry* **2006**, *45*, 7904–7912; b) E. Yeh, L. C. Blasiak, A. Koglin, C. L. Drennan, C. T. Walsh, *Biochemistry* **2007**, *46*, 1284–1292.
- [13] A. Lang, S. Polnick, T. Nicke, P. William, E. P. Patallo, J. H. Naismith, K.-H. van Pée, *Angew. Chem. Int. Ed.* **2011**, *50*, 2951–2953; *Angew. Chem.* **2011**, *123*, 3007–3010.
- [14] J. Zeng, J. Zhan, *Biotechnol. Lett.* **2011**, *33*, 1607–1613.
- [15] a) X. Robert, P. Gouet, *Nucleic Acids Res.* **2014**, *42*, W320–W324; b) S. Zehner, A. Kotzsch, B. Bister, R. D. Süßmuth, C. Méndez, J. A. Salas, K.-H. van Pée, *Chem. Biol.* **2005**, *12*, 445–452.
- [16] M. Hölzer, W. Burd, H.-U. Reibig, K.-H. van Pée, *Adv. Synth. Catal.* **2001**, *343*, 591–595.
- [17] a) J. T. Payne, M. C. Andorfer, J. C. Lewis, *Angew. Chem. Int. Ed.* **2013**, *52*, 5271–5274; *Angew. Chem.* **2013**, *125*, 5379–5382; b) J. T. Payne, C. B. Poor, J. C. Lewis, *Angew. Chem. Int. Ed.* **2015**, *54*, 4226–4230; *Angew. Chem.* **2015**, *127*, 4300–4304.
- [18] M. Frese, N. Sewald, *Angew. Chem. Int. Ed.* **2015**, *54*, 298–301; *Angew. Chem.* **2015**, *127*, 302–305.
- [19] S. A. Shepherd, C. Karthikeyan, J. Latham, A.-W. Struck, M. L. Thompson, B. R. K. Menon, M. Q. Styles, C. Levy, D. Leys, J. Micklefield, *Chem. Sci.* **2015**, *6*, 3454–3460.

Manuscript received: January 27, 2016

Accepted article published: February 3, 2016

Final article published: March 30, 2016