

Synthetic Methods

Structure-Selective Catalytic Alkylation of DNA and RNA**

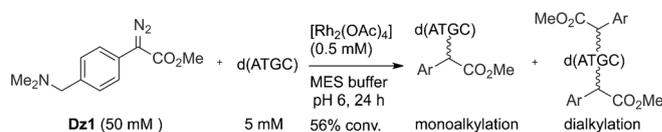
Kiril Tishinov, Kristina Schmidt, Daniel Häussinger, and Dennis G. Gillingham*

Deciphering the complex puzzle of nucleic acid (NA) chemistry and biology has proven incredibly difficult^[1] and continues unabated after more than a century.^[2] Chemists have begun to recognize that the extraordinary informational and structural properties that have made NAs so fundamental in biology could be exploited in other contexts.^[3] Simple, selective chemical strategies to modify NAs would therefore have a significant impact on studying their varied functions as well as repurposing their special abilities, yet most extant tools to target them are unselective or labor intensive.^[4] Although a number of reports exploring the potential of organometallic catalysis in biological environments have recently appeared,^[5] such strategies have not yet been applied towards selective NA alkylation. But untapped potential is suggested by the ability to tune the reactivity of metal complexes through judicious ligand selection and a number of studies demonstrating the feasibility of metal catalysis with DNA as a ligand.^[6] We establish herein that rhodium-catalyzed carbene transfer is a viable means of achieving the selective alkylation of a variety of NAs.

Although there are a range of chemical methods to modify NAs,^[7] researchers typically resort to the effort of building a new phosphoramidite for solid-phase synthesis whenever an unnatural NA is required. In addition, access to unnatural NAs beyond the size limit of solid-phase synthesis is especially difficult.^[7c] A handful of methods to target native NAs have been described and employ a reactive molecule linked to a guiding motif to direct specificity.^[8] While important in certain contexts, the complexity of these systems precludes their use as general synthetic strategies. A promising enzymatic method has recently been developed for NA modification which co-opts the natural S-adenosylmethionine (SAM)/methyltransferase system by employing SAM analogues bearing functional groups which can undergo subsequent chemoselective transformations.^[9] In the same spirit as enzymatic catalysis, organometallic approaches could bring unrivalled levels of efficiency and selectivity to the chemistry of NA modification. Indeed, as we show herein, despite the myriad of possible reaction pathways of a rhodium carbenoid with a NA, N-H insertion into exocyclic amines is preferred

(see equation for Table 1). We also find that the secondary structure of the NA can be exploited to guide selectivity in the alkylation. Taken together the recently developed enzymatic method and the organometallic approach we describe should offer researchers a considerable increase in flexibility for tailored NA synthesis.

Recent reports on employing rhodium carbenoids,^[10] and rhodium/peptide conjugates^[11] in alkylating various proteins, even in cell lysates,^[12] as well as structural studies of a number of Rh^{II}/DNA complexes,^[13] suggested dimeric rhodium complexes as a promising candidate catalyst system for NA alkylation. Indeed, in an initial proof-of-concept experiment the simple tetradexynucleotide d(ATGC) was treated with 10 mol% [Rh₂(OAc)₄] and the diazo substrate **Dz1** in aqueous buffer (Scheme 1). Under these conditions the **Dz1** substrate was completely consumed after 24 hours and consumption of the tetradexynucleotide (56% conversion) corresponded to the appearance of a number of new products whose masses indicate singly and doubly modified d(ATGC). Tandem MS analysis of the monoalkylation products allowed unambiguous assignment of the purine bases as the sites of modification (see Figure S1 in the Supporting Information).



Scheme 1. Proof-of-concept for rhodium-catalyzed NA alkylation.

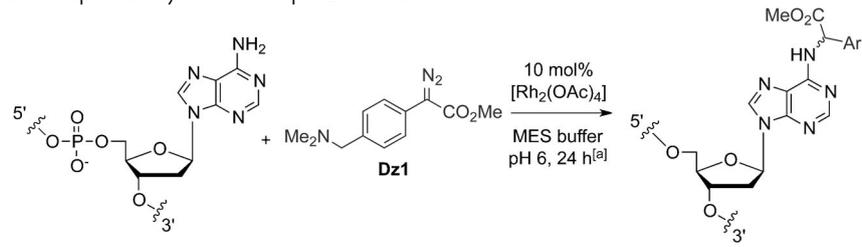
With the feasibility of targeting NAs through rhodium-carbenoid catalysis established, we undertook a more comprehensive investigation of the reaction with a series of hairpin sequences (Table 1). Hairpins were chosen because they contain a number of common NA structural elements in a single molecule. The first hairpin we tested contained thymidine (T) rings in its turn region and was otherwise double-stranded (entry 1). We knew that T was unreactive from the experiment with d(ATGC), so this molecule would allow an assessment of the susceptibility of double-stranded stretches to the alkylation. Interestingly this hairpin was completely unreactive, thus revealing the prospect of exploiting double strands as a type of shielding motif to enable the targeting of specific unpaired bases in a given NA. To test this possibility alkylation was attempted on hairpins containing adenine as an overhang base (entries 2, 3, and 5) or in the turn region (entry 4). As expected both of these motifs were viable substrates. The specificity of the process for unpaired NA sequences offers a new strategic tool for post-synthetic NA

[*] K. Tishinov, K. Schmidt, Priv.-Doz. D. Häussinger, Prof. Dr. D. G. Gillingham
Department of Chemistry, University of Basel
St. Johanns-Ring 19, CH-4056, Basel (Switzerland)
E-mail: dennis.gillingham@unibas.ch

[**] This work was supported by start-up funds from the University of Basel. Karl Gademann and Florian Seebeck are gratefully acknowledged for valuable discussions. Raphael Wyss is thanked for his help in interpreting some of the NMR spectra.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201205201>.

Table 1:
Structure-specific alkylation of hairpin DNA and RNA.



Entry	Sequence	Structure	Conv. [%]	Product	Significance
1	d(CGAACGTTT TTCGTTCCG)		0	no reaction	double-stranded sequences are unreactive
2	d(CGAACGTTT TTCGTTCGA)		21		3' overhang regions can be targeted
3 ^[b]	d(ACGGAAT TCCGTTTTT CGGAATCCG)		19		5' overhang regions can be targeted
4	d(CGAACGTTA TTCGTTCCG)		22 ^[c]		turn regions/bulges can be targeted
5	r(CUAGCAUUU UUUGCUGA)		33 ^[c]		short hairpin RNAs can be targeted

[a] Conditions: 5 mM oligonucleotide, 50 mM **Dz1**, 0.5 mM $[\text{Rh}_2(\text{OAc})_4]$, 100 mM MES, pH 6. [b] Site of alkylation confirmed by tandem MS analysis of the fragments resulting from a restriction digest (see the Supporting Information). [c] Oligonucleotide conversion after addition of a second bolus of **Dz1** and incubating for a further 24 h.

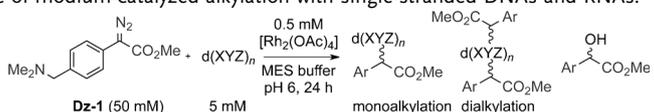
modification and may prove practical for chemical tailoring of DNA architectures in the burgeoning field of DNA nanotechnology.^[3]

Short hairpin RNAs and short dsRNAs are molecules of high current interest because of their central role in the RNA interference (RNAi) pathway.^[14] RNAs that participate in this pathway always contain 3' overhangs, which are essential to trigger the assembly of the active ribonucleoprotein complex. The successful 3'-alkylation of a hairpin RNA (Table 1, entry 5) is therefore especially significant as it establishes the method as a convergent approach to the synthesis of base-modified RNAs. Such molecules have been employed to study and modulate RNAi,^[15] but thus far the diversity of the modified RNA pool has been limited because of the laborious linearity of solid-phase synthesis methods. The convergent approach described here

could alleviate many of the current problems.

The preference for purine nucleobases observed for the alkylation of d(ATGC) (Scheme 1) warranted a systematic evaluation of which features in the oligonucleotide primary structure were important for the modification reaction. Homotetramers were analyzed to probe the relative propensity of each base towards alkylation (Table 2, entries 1–4). Consistent with the results from the hairpin series the homotetramers d(T)₄ and r(U)₄ were completely unreactive, thus indicating that neither the ribophosphate backbone nor the T and U nucleobases are the site of reaction.^[16] This data was confirmed by MS/MS experiments with three trinucleotides containing A, C, or G nested between T residues (entries 5–7); here alkylation was observed only at the A, C, or G nucleobases (see Figures S5–S10 in the Supporting Information). In contrast to the result with d-(ATGC), where C was untouched, alkylation of d(C)₄ proceeded efficiently (entry 4), thus suggesting that purine bases are modified pref-

Table 2: Scope of rhodium-catalyzed alkylation with single-stranded DNAs and RNAs.



Entry	Substrate	Total alkylation products	Conv. NA [%] ^[a]	Conv. Dz-1 [%] ^[a]	Selectivity factor (N-H/O-H) ^[b]
		mono-alkylation			
1	d(T) ₄	0	0	20	≈ 0
2	r(U) ₄	0	0	30	≈ 0
3	d(A) ₄	4	39	40	–
4 ^[c]	d(C) ₄	6	45	> 98	–
5 ^[d]	d(TAT)	2	20	> 98	282
6 ^[d,e]	d(TGT)	2	25	> 98	282
7 ^[c,d]	d(TCT)	2	16	> 98	210
8 ^[d]	d(ATGC)	4	56	> 98	–
9	r(ACUGCU)	3	68	> 98	9970
10	d(ACT GCT)	5	56	> 98	–
11	d(CTC TCT)	2	32	> 98	1550
12	d(CTG GCT)	6	42	> 98	1050
13	d(TTT ATT TGT TTC TTT)	4	37	> 98	–

[a] Conversion refers to consumption of starting material (either NA or diazo compound) as determined by HPLC. [b] Approximated from the ratio of N-H and O-H insertion products multiplied by the effective concentration of potential N-H versus O-H insertion sites. Only calculated for reactions that do not give dialkylation. [c] **Dz1** is completely consumed after 3 h. [d] Site of alkylation determined by MS/MS analysis (see the Supporting Information). [e] Equimolar mixture of d(TAT) and d(TGT) (5 mM each), 1 mM $[\text{Rh}_2(\text{OAc})_4]$, 50 mM **Dz1**, 24 h (see the Supporting Information).

entially but that C residues can also be targeted. The remainder of the RNAs and DNAs shown in Table 2 provide further evidence of the generality of the alkylation process.

Diazo substrate that did not lead to NA alkylation was instead predominately converted into the corresponding O-H insertion product (see Table 2). Despite this side reaction a significant percentage of catalyst turnovers were directed to the oligonucleotide. Considering the large excess of water in the system this indicates a substantial preference for N-H insertion over O-H insertion (see the selectivity factor; Table 2). This preference may simply be a consequence of the increased nucleophilicity of the amine towards the intermediate rhodium carbenoid, but a number of observations speak to a more intimate role of the NA: First, while the alkylation of d(TGT) (Table 2, entry 6) and d(G)₄ (see Figures S4 and S9 in the Supporting Information) was slow and often led to the formation of a precipitate, a mixture of d(TAT) with these oligonucleotides re-established a smooth alkylation of both NAs. Second, in the absence of a NA **Dz1** slowly delivered O-H insertion under the standard reaction conditions (< 10% conv. in 24 h); but when d(TAT) was present not only was the NA alkylated but the rate of the O-H insertion side reaction also increased. A full understanding of the modulation of catalytic activity that we observe with certain NAs will have to await detailed mechanistic studies. Nevertheless, these results clearly demonstrate that certain NA sequences can have a profound impact on the catalytic activity of the rhodium(II) species.

To characterize the exact site of alkylation on the nucleobases, NMR experiments with the three trinucleotides d(TAT), d(TGT), and d(TCT) modified with a ¹³C-labeled version of the diazo substrate **Dz1** (¹³C-**Dz1**), were performed. In all three cases, the ³¹P NMR spectra showed two sets of two phosphate signals, thus indicating the formation of two different alkylated species (Figure 1, and Figure S14 in the Supporting Information).

Further analysis of these species in H₂O showed two sets of resonances for the proton H8 on the ¹³C-labeled carbon atom (C8) of the tag (Figure 2a). For each proton a large coupling to ¹³C (¹J = 145.7 Hz) was observed as well as a smaller coupling to the NH proton of the exocyclic nitrogen atom (³J = 5.4 Hz). Unambiguous proof for attachment of C8 to the exocyclic nitrogen atom of the nucleobase was obtained by H/D exchange experiments in D₂O, which led to a collapse of the latter couplings (Figure 2b).

In accordance with these findings, ROE contacts between the cytidine H5 and the *ortho* protons of the aromatic ring in the case of d(TCT) are clearly visible in the ROESY spectra (see Figure S15 in the Supporting Information). A further corroboration was obtained by HMBC experiments, wherein no ³J correlations to C2 and C4 were observed; these correlations should be very intense for an alkylation product in the endocyclic 3-position. It is therefore clear that the alkylation takes place on the exocyclic nitrogen atom in all three cases and that a roughly equimolar mixture of the two diastereoisomers are formed as a result of the new stereogenic center on C8. In the case of d(TCT) a complete assignment of both diastereoisomers was accomplished (see Tables S2–S4 in the Supporting Information) and the only major differences in

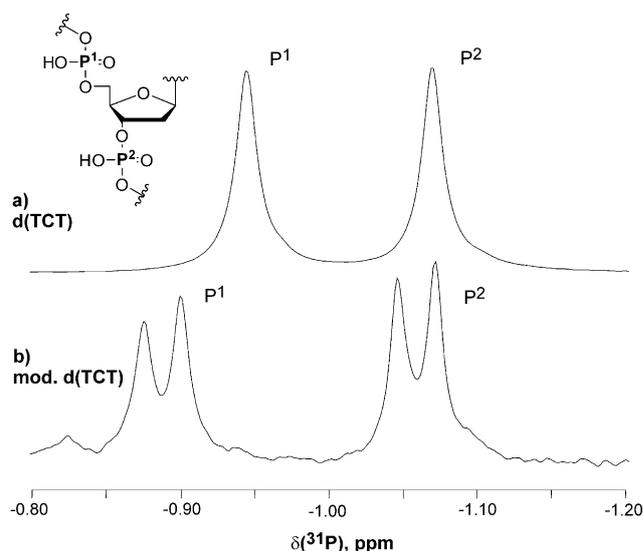


Figure 1. ³¹P NMR spectra of d(TCT) before (a) and after (b) reaction with ¹³C-**Dz1**.

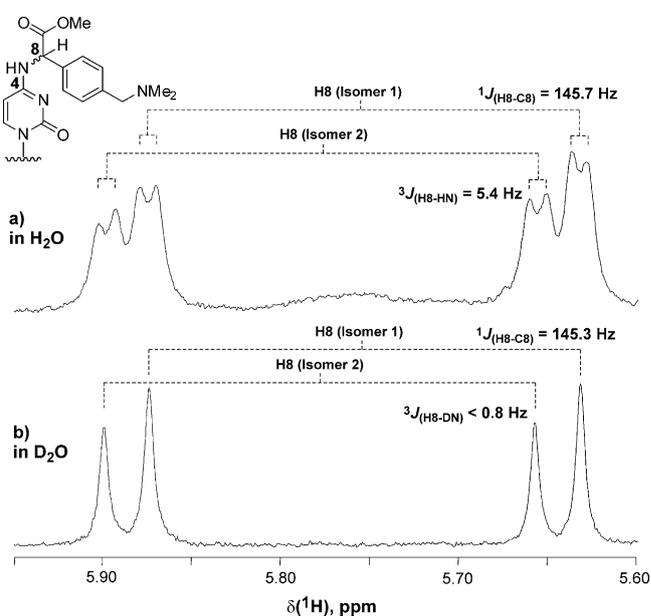


Figure 2. H-8/NH proton spin–spin coupling for ¹³C-**Dz1**-modified d(TCT) in H₂O (a) and in D₂O (b).

chemical shift with respect to the unmodified d(TCT) are observed for C4 and H5—again consistent with exocyclic alkylation for both isomers. Tandem MS/MS (MS²) and MS³ experiments with the alkylation products of the trinucleotides d(TAT), d(TGT), and d(TCT) further substantiated modification at the exocyclic nitrogen atom of the nucleobase as the ions observed after the fragmentation of these alkylated trinucleotides are in full agreement with those expected for exocyclic N-H insertion products (see Figures S11–S13 in the Supporting Information).^[17]

As the results collected in Table 3 indicate the process showed little dependence on the structure of the diazo

Table 3: Scope of diazo substrates and reaction conditions in NA alkylation.

Entry	Diazo substrate	[Diazo] [mM] ^[a]	Buffer	Conv. NA [%] ^[a]
1	Dz1	50	pH 6, 100 mM MES	56
2	Dz1	50	pH 6, 75 mM <i>t</i> BuNHOH ^[c]	52
3	Dz1	50	pH 6, 100 mM KP _i	52
4	Dz2	50	pH 6, 100 mM MES	24
5	Dz3	50	pH 6, 100 mM MES	48
6	Dz3	50	pH 6.5, 100 mM KP _i	51
7	Dz3	50	pH 7.5, 100 mM KP _i	50
8	Dz4	25	pH 6, 100 mM MES	30
9 ^[d]	Dz5	50	pH 6, 100 mM MES	51
10	Dz6	25	pH 6, 100 mM MES	40
11	Dz7	12.5	pH 6, 100 mM MES	< 5

[a] Concentrations below 50 mM indicate substrates of low solubility. [b] Conversion refers to consumption of the starting NA as determined by HPLC analysis. [c] *t*BuNHOH has previously been shown to impact aqueous rhodium-catalyzed reactions.^[10] [d] d(TTT ATT TGT TTC TTT) was used instead of d(ATGC). Boc = *tert*-butoxycarbonyl, KP_i = potassium phosphate buffer, MES = 2-(*N*-morpholino)ethanesulfonic acid.

substrate or the reaction conditions. The most important limitation in terms of substrate scope we encountered was maintaining water solubility at the concentrations typically needed to achieve efficient alkylation (25–50 mM). Although the substrate could be customized for any specific application, the propargyl-containing molecules **Dz4** and **Dz5** (entries 8 and 9) should prove particularly versatile because they allow the copper(I)-catalyzed azide–alkyne cycloaddition to be employed to generate diverse structures from a single diazo precursor—a feature demonstrated by the click reaction of both a rhodamine fluorophore and a biotin derivative with an acetylene-tagged oligonucleotide (Figure 3).

The practical potential of the alkylation process was demonstrated through the simple generation of fluorescently labeled PCR amplicons. A modified T7 promoter primer (T7PP) was first alkylated with **Dz5** and then used in a PCR reaction. We expected a PCR to be successful despite the unspecific alkylation profile with single-stranded NAs since

double-helix formation and polymerase extension should only occur with those primers that get alkylated near their 5'-end. Indeed this seems to be the case as robust PCR amplification is seen in all samples that employ alkylated T7PPs (Figure 4, Lanes 3–5). The experiment in Lane 5 is particularly important as it highlights the potential of the technique for generating labeled genes; here the propargyl group is linked through a click reaction to a rhodamine derivative prior to the PCR. The amplification product of this plasmid PCR is clearly visible with a standard UV lamp before the addition of ethidium bromide, thus indicating that labeled amplicons are accessible through this simple technique.

The process outlined herein establishes the feasibility of employing rhodium catalysis to achieve the direct alkylation of native NA structures. Moreover its predictable selectivity profile allows the strategic targeting of unpaired nucleobases such as those present in single strands, bulge regions, and overhangs. Key features of the present process are its simplicity and directness: the catalyst is commercially available, the diazo compounds are readily prepared, and native NAs are viable substrates. Challenges for future development include further refining the base specificity and achieving efficient alkylation at high dilution. Given the enormous excess of water present,

the preference for N-H insertion observed here, as well as by others in the protein-labeling field,^[10] is intriguing. We are currently investigating the nature of this selectivity. The new catalytic alkylation process offers a powerful strategy in the labeling and modification of NA derivatives, a strategy which we foresee being applied to the tailoring of DNA architectures and in the labeling of NAs for biological study or therapeutic applications.

Received: July 2, 2012

Published online: October 22, 2012

Keywords: insertion · nucleic acids · rhodium · water chemistry · synthetic methods

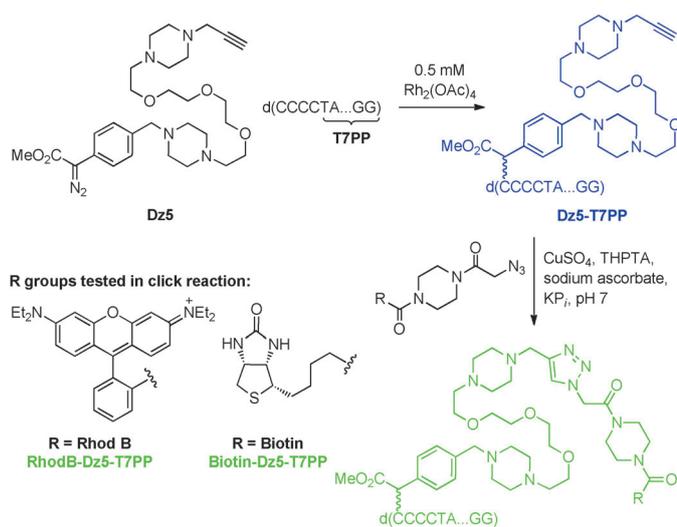


Figure 3. Rhodium-catalyzed NA alkylation and subsequent click reaction. T7PP, which contains a short 5'-extension, is alkylated under standard conditions with the propargyl-containing diazo substrate **Dz5**. The product of the alkylation, **Dz5-T7PP**, can then be further modified with azide-containing substrates under standard click conditions.^[18] Reactions with biotin and rhodamine B derivatives are shown. THPTA = tris(hydroxypropyltriazolyl)methylamine.

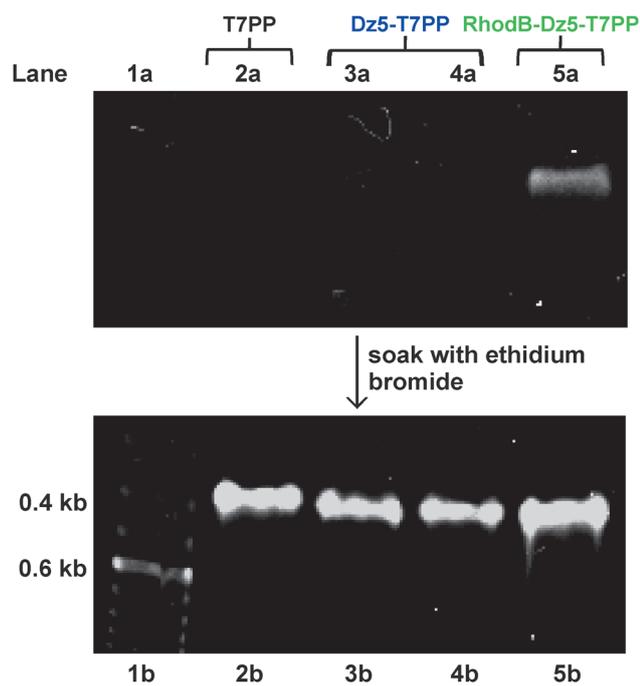


Figure 4. Application of the catalytic alkylation in PCR labeling. Gel: The gel is run without ethidium bromide to allow visualization of fluorescently labeled products (Lanes 1a–5a) and then soaked with ethidium bromide to visualize all NA species (Lanes 1b–5b). Lane 1: 100 bp DNA ladder. Lane 2: Control reaction with unmodified T7PP and plasmid DNA. Lane 3: T7PP is alkylated with **Dz5** to give **Dz5-T7PP** and this entire crude reaction mixture is then employed in a PCR reaction with plasmid DNA. Lane 4: Same as Lane 3 except the **Dz5-T7PP** is purified before carrying out the PCR. Lane 5: RhodB-**Dz5-T7PP** from the click reaction shown in Figure 3 is employed in the PCR.

- [1] R. Dahm, *Hum. Genet.* **2008**, *122*, 565–581.
- [2] L. Salmena, L. Polisenio, Y. Tay, L. Kats, P. P. Pandolfi, *Cell* **2011**, *146*, 353–358.
- [3] F. A. Aldaye, A. L. Palmer, H. F. Sleiman, *Science* **2008**, *321*, 1795–1799.
- [4] G. T. Hermanson, *Bioconjugate techniques*, Elsevier/Academic Press, Amsterdam, **2008**.
- [5] a) C. Streu, E. Meggers, *Angew. Chem.* **2006**, *118*, 5773–5776; *Angew. Chem. Int. Ed.* **2006**, *45*, 5645–5648; b) R. M. Yusop, A. Unciti-Broceta, E. M. V. Johansson, R. M. Sánchez-Martín, M. Bradley, *Nat. Chem.* **2011**, *3*, 239–243; c) V. Hong, N. F. Steinmetz, M. Manchester, M. G. Finn, *Bioconjugate Chem.* **2010**, *21*, 1912–1916.
- [6] a) A. J. Boersma, D. Coquière, D. Geerdink, F. Rosati, B. L. Feringa, G. Roelfes, *Nat. Chem.* **2010**, *2*, 991–995; b) G. Roelfes, B. L. Feringa, *Angew. Chem.* **2005**, *117*, 3294–3296; *Angew. Chem. Int. Ed.* **2005**, *44*, 3230–3232; c) P. J. Deuss, R. den Heeten, W. Laan, P. C. J. Kamer, *Chem. Eur. J.* **2011**, *17*, 4680–4698.
- [7] a) K. A. Browne, *J. Am. Chem. Soc.* **2002**, *124*, 7950–7962; b) R. J. Heetebrij, E. G. Talman, M. A. v. Velzen, R. P. M. v. Gijlswijk, S. S. Snoeijers, M. Schalk, J. Wiegant, F. v. d. Rijk, R. M. Kerkhoven, A. K. Raap, H. J. Tanke, J. Reedijk, H. J. Houthoff, *ChemBioChem* **2003**, *4*, 573–583; c) S. H. Weisbrod, A. Marx, *Chem. Commun.* **2008**, 5675–5685.
- [8] a) Q. Zhou, S. E. Rokita, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 15452–15457; b) K. Onizuka, Y. Taniguchi, S. Sasaki, *Bioconjugate Chem.* **2009**, *20*, 799–803; c) C.-X. Song, C. He, *Acc. Chem. Res.* **2011**, *44*, 709–717; d) M. Di Antonio, F. Doria, S. N. Richter, C. Bertipaglia, M. Mella, C. Sissi, M. Palumbo, M. Freccero, *J. Am. Chem. Soc.* **2009**, *131*, 13132–13141.
- [9] a) C. Dalhoff, G. Lukinavicius, S. Klimasauskas, E. Weinhold, *Nat. Chem. Biol.* **2006**, *2*, 31–32; b) Y. Motorin, J. Burhenne, R. Teimer, K. Koynov, S. Willnow, E. Weinhold, M. Helm, *Nucleic Acids Res.* **2011**, *39*, 1943–1952.
- [10] J. M. Antos, J. M. McFarland, A. T. Iavarone, M. B. Francis, *J. Am. Chem. Soc.* **2009**, *131*, 6301–6308.
- [11] B. V. Popp, Z. T. Ball, *J. Am. Chem. Soc.* **2010**, *132*, 6660–6662.
- [12] a) Z. Chen, B. V. Popp, C. L. Bovet, Z. T. Ball, *ACS Chem. Biol.* **2011**, *6*, 920–925; b) Z. Chen, F. Vohidov, J. M. Coughlin, L. J. Stagg, S. T. Arold, J. E. Ladbury, Z. T. Ball, *J. Am. Chem. Soc.* **2012**, *134*, 10138–10145.
- [13] a) H. T. Chifotides, K. R. Dunbar, *Acc. Chem. Res.* **2005**, *38*, 146–156; b) H. T. Chifotides, K. R. Dunbar, *J. Am. Chem. Soc.* **2007**, *129*, 12480–12490.
- [14] J. Kurreck, *Angew. Chem.* **2009**, *121*, 1404–1426; *Angew. Chem. Int. Ed.* **2009**, *48*, 1378–1398.
- [15] a) H. Peacock, A. Kannan, P. A. Beal, C. J. Burrows, *J. Org. Chem.* **2011**, *76*, 7295–7300; b) A. R. Hernández, L. W. Peterson, E. T. Kool, *ACS Chem. Biol.* **2012**, *7*, 1454–1461.
- [16] The possibility that d(T)₄ is simply an inhibitor of the catalyst was excluded by a control experiment where d(T)₄ was added to the reaction of d(A)₄ and no change was observed in its alkylation profile (see Figure S3 in the Supporting Information).
- [17] S. S. Jensen, X. Ariza, P. Nielsen, J. Vilarrasa, F. Kirpekar, *J. Mass Spectrom.* **2007**, *42*, 49–57.
- [18] V. Hong, S. I. Presolski, C. Ma, M. G. Finn, *Angew. Chem.* **2009**, *121*, 10063–10067; *Angew. Chem. Int. Ed.* **2009**, *48*, 9879–9883.