

# Chemical Science

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

## COMMUNICATION

# Template-directed ligation on repetitive DNA sequences: a chemical method to probe the length of Huntington DNA

Cite this: DOI: 10.1039/x0xx00000x

Received 00th January 2012,  
Accepted 00th January 2012Anika Kern and Oliver Seitz<sup>a</sup>

DOI: 10.1039/x0xx00000x

www.rsc.org/

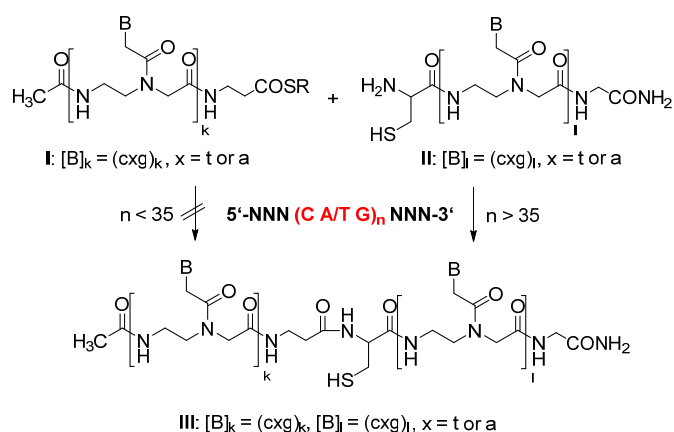
Several genomic disorders are caused by an excessive number of DNA triplet repeats. We developed a DNA-templated reaction in which product formation occurs only when the number of repeats exceeds a threshold indicative for the outbreak of Chorea Huntington. The combined use of native chemical PNA ligation and auxiliary DNA probes enabled reactions on templates obtained from human genomic DNA.

Nucleic acid controlled chemical reactions are emerging as useful tools for nucleic acid diagnosis and imaging.<sup>1-6</sup> Reactive probes offer advantages to the commonly used non-reactive hybridization probes.<sup>7</sup> For example, DNA/RNA-controlled reactions proceed with very high sequence specificity – at the single nucleotide level if desired – because adjacent and simultaneous annealing of two probe molecules is required to trigger the chemical reaction. Furthermore, chemical reactions can occur with turnover in template, which offers opportunities for the amplified detection of product signals.<sup>7, 8</sup> While significant efforts have been invested into the development of new chemistries applied in challenging environments such as in cell lysates, during polymerase chain reaction and within living cells, comparatively little attention has been paid to the nature of the targeted nucleic acid template itself.<sup>9-18</sup> A number of genes contain repeat sequences. Several disorders such as Huntington disease, myotonic dystrophy and spinocerebellar ataxia are caused by an excessive number of trinucleotide repeats.<sup>19-23</sup> For example, the expansion of triplet repeats in the human IT15 gene leads to Chorea Huntington Disease (HD).<sup>24, 25</sup> Alleles with less than 26 repeats are considered normal and are stable, but alleles that contain 27 – 36 triplet repeats may expand during meiosis, hence leading to increased repeats in the next generation. Eventually, alleles with more than 36 triplet repeats lead to HD pathology.<sup>26, 27</sup>

We envisioned that a nucleic acid directed reaction could be designed to provide product required that the number of repeat sequences exceeds a critical threshold value. This idea was based on

the notion that the number of triplet repeats in the template should correlate with the effective concentration of template available for binding of the reactive probes.

Our design of a reaction system that discriminates between repeat sequence DNA of varied length involves the native chemical ligation of peptide nucleic acid (PNA)-based probes (Scheme 1).<sup>28-32</sup>



Scheme 1: Principle of a native chemical ligation of PNA probes (I and II), in which formation of product III is indicative for a critical length of a triplet repeat DNA template.

PNA is a DNA analogue that has high affinity for complementary DNA.<sup>33-35</sup> This facilitates hybridization and reactions on folded nucleic acid targets (such as long DNA repeat sequences). The method requires two reactive PNA probes, one of which is equipped with a C-terminal thioester moiety (I) while the other bears an N-terminal cysteine residue (II). Hybridization with the complementary DNA sequence brings the reactive groups into proximity and triggers the native chemical ligation, whereas no reaction should occur in the absence of template.<sup>30</sup>

In the initial phase of probe development, we compared a probe set directed against the CAG repeats in the Huntington coding strand

with a probe set targeted against the CTG repeats of the corresponding non-coding strand. The latter was favoured by us owing to difficulties in the synthesis of (ctg)<sub>n</sub> containing PNA probes. Next we investigated whether templated reactions occur on repeat sequence templates. The cag-containing PNA thioester **1** and cysteinyl-PNA **2** were allowed to react in the absence and presence of a size-matched synthetic CTG-repeat DNA template (Figure 1a).

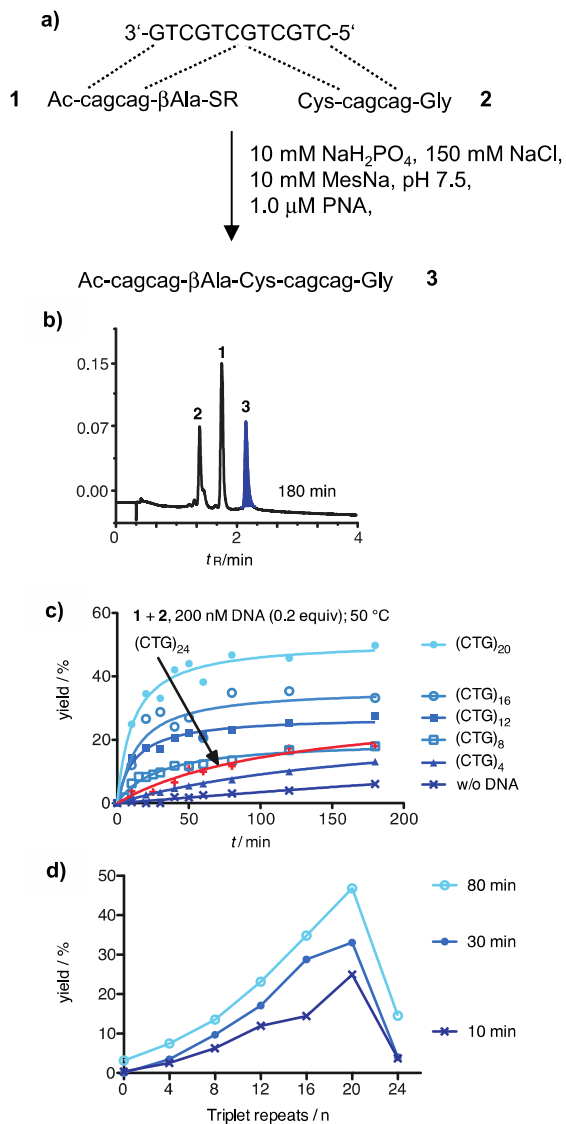


Figure 1: a) DNA templated native chemical PNA ligation between **1** and **2** at equimolar probe/template ratio; b) UPLC analysis of aliquots withdrawn from the reaction mixture after the specified reaction times; c) Time course of reactions in presence of synthetic ssDNA varying in triplet repeat length; d) Fractional increase in product yield after 10 min, 30 min or 80 min. Conditions: 1 μM reactive PNA probes in buffer (10 mM NaH<sub>2</sub>PO<sub>4</sub>, 150 mM NaCl, 10 mM MesNa, pH 7.5), 200 nM template when added, 50 °C. R = (CH<sub>2</sub>)<sub>2</sub>SO<sub>3</sub>H.

The formation of product **3** was monitored by reversed phase-ultra high performance chromatography (RP-UPLC, Figure 1b). This enabled the analysis of picomol amounts within short time (4 minutes). During the optimization of the native chemical ligation between **1** and **2** we noticed a surprisingly high background reaction in absence of template (Figure S37). To avoid probe-probe

interactions the reaction temperature was increased to 50 °C. We next added 200 nM CTG-repeat DNA templates of varied length. This concentration is within the range of amplicons obtained after PCR. Of note, the rates of native chemical PNA ligation proved critically dependent on the number of repeats (Figure 1c). The positive correlation between product yield and repeat length extended until 20 CTG repeats (Figure 1d). Then, at 24 CTG repeats the efficiency of the PNA ligation suddenly dropped. Increases of probe concentration to 5 μM did not lead to an improvement. Melt experiments (Table S1) suggested that the template adopted a superstructure that had significantly higher stability ((CTG)<sub>20</sub>, T<sub>M</sub> = 43 °C; (CTG)<sub>32</sub>, T<sub>M</sub> = 66 °C) than the complexes formed between the reactive PNA probes and the template (**1**•(CTG)<sub>4</sub>, T<sub>M</sub> = 39 °C; **2**•(CTG)<sub>4</sub>, T<sub>M</sub> = 38 °C). The PNA probes were elongated by three nucleotides to increase the template affinity. However, the rather large extent of background ligation was problematic (Figure S38).

We assumed that the use of aminomodified PNA should offer a solution to the problem of PNA-PNA interaction and, at the same time, confer the required increase of the template affinity. According to this, charge repulsion would help reduce PNA-PNA interactions whereas charge attraction would foster the formation of PNA-DNA template complexes. For that purpose, the aminoethyllysine cytidine building block **c\*** was prepared (SI) and incorporated into the cag-containing PNA thioester probe **4** (Figure 2a). The cysteinyl-PNA **5** was equipped with a C-terminal lysine. The extent of background

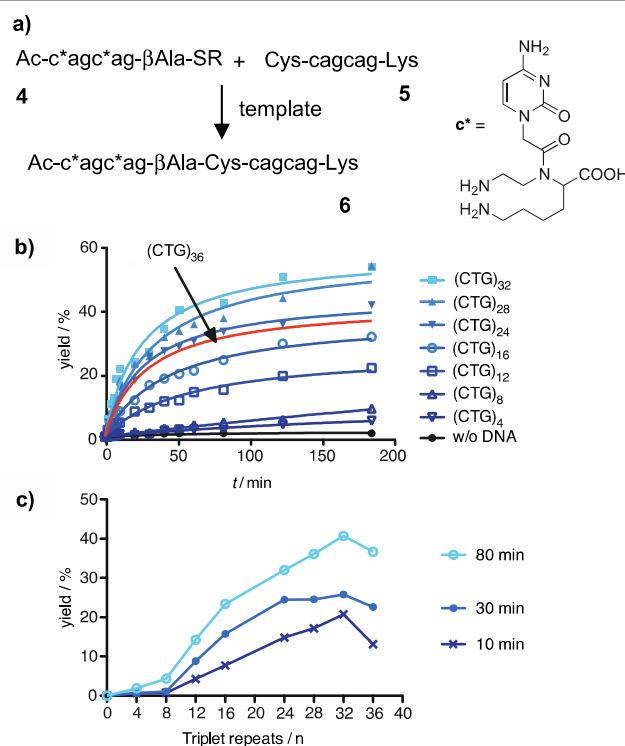


Figure 2: a) Native chemical ligation reaction between aminomodified PNA probes **4** and **5**; b) time course of product **6** formed in absence or presence of 100 nM DNA templates varying in the number of triplet repeats; c) template dependence of yields obtained after 10 min, 30 min or 80 min. (Conditions: see Figure 1, R = (CH<sub>2</sub>)<sub>2</sub>SO<sub>3</sub>H).

reaction between the aminomodified reactive probes **4** and **5** in absence of template was very small (Figure S37) and the scope of templated reactions was extended to sequences longer than 20 repeats. The yield of the templated ligation between **4** and **5** continued to increase until a length of 32 triplet repeats (Figure 2b, Figure S39). This corresponds to the length, which may lead to the Huntington phenotype in offspring. However, the ligation efficiency was severely affected, again, when the DNA length exceeded 32 repeats (Figure 2c). The addition of denaturing agents such as DMSO, glycerol, betaine or formamide allowed increases of product formation (Figure S41-S42). But still, the fractional increase in ligation rate did not surpass a length threshold of 32 repeats.

In previous ligation chemistries, auxiliary strands have been added to prevent reannealing of a DNA template strand produced by PCR.<sup>11, 30, 36</sup> However, intramolecular hybridization is a very efficient process and the length requirements for oligonucleotides that block the formation of superstructures in the Huntington template were unknown. To avoid slippage along the template, we designed a blocker strand that is equipped with a 5'-terminal hexanucleotide sequence which anchored the CAG repeat sequence to the 3'-flanking region of the CTG repeat tract. The so-called partial blocker DNA strands **7** and **8** offered 12 or 28 CAG repeats for binding interactions with the CTG repeat part of the Huntington non-coding strand (Figure 3a). The partial blockers **7** or **8** were

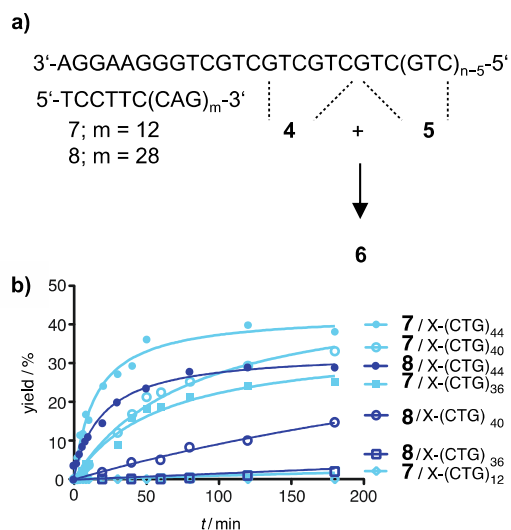


Figure 3: a) Principle of native chemical PNA ligation aided by partial DNA blockers **7** and **8** and b) time course of the product formation triggered by DNA templates varying in triplet repeats length for reaction systems involving **7** or **8**. Conditions: 1  $\mu$ M reactive PNA probes in buffer (10 mM NaH<sub>2</sub>PO<sub>4</sub>, 150 mM NaCl, 10 mM MesNa, pH 7.5), 100 nM DNA template, 100 nM partial blocker, R = (CH<sub>2</sub>)<sub>2</sub>SO<sub>3</sub>H.

added to the **4** + **5** ligation. Of note, the ligations proceeded remarkably well on DNA templates containing 44 triplet repeats (Figure 3b). No product was formed when the template strand contained less triplet repeats than the partial blocker. Reactions performed in presence of the short blocker **7** provided higher ligation yields than reactions in presence of the long blocker **8**. This can be attributed to the increased number of template bases available for

annealing of reactive probes. These results showed that a DNA-directed chemical reaction can distinguish between healthy and pathological DNA length.

To test the reaction system under realistic conditions we analyzed human genomic DNA templates. Samples from a Chorea Huntington patient (purchased from the Coriell Institute) and human wild-type DNA were subjected to PCR, which was used to amplify the trinucleotide repeat segment of the IT15 gene. This was achieved by using a biotin-modified forward primer and an Alexa700-labeled reverse primer and adapting a known PCR method.<sup>37, 38</sup> The amplified DNA was captured by streptavidin-magnetic beads and NaOH treatment was used to release single stranded DNA templates (SI). The yield of PCR and the concentration of template strands were determined via fluorescence measurements. The reactive probes and partial blocker **7** were added to the PCR-DNA single strand. Aliquots were withdrawn after 0-180 min reaction time and analyzed via UPLC (Figure 4). The DNA template obtained through

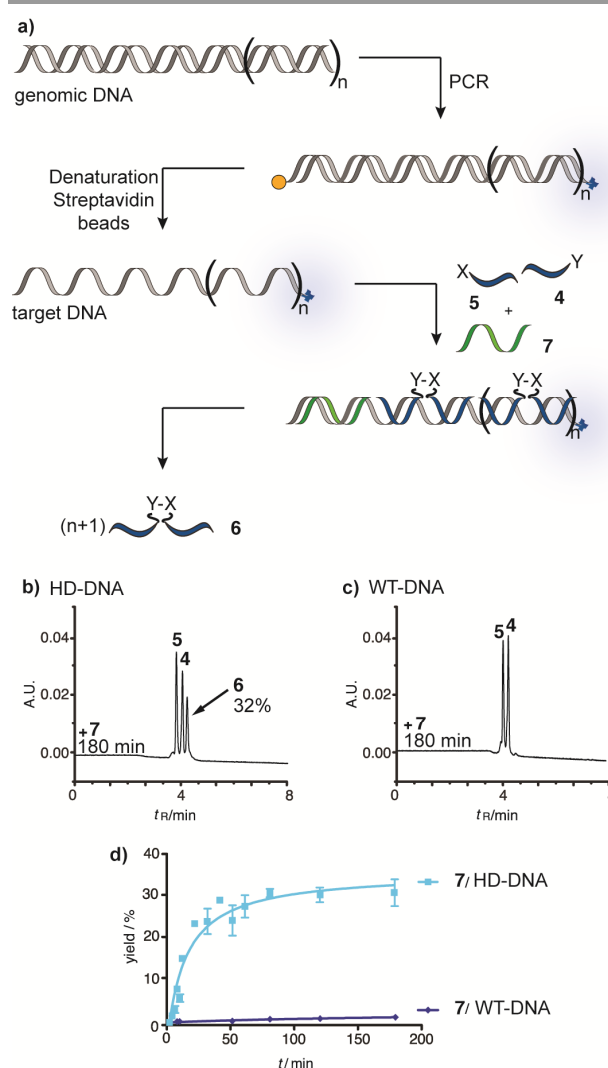


Figure 4: a) Principle of native chemical PNA ligation on templates obtained via PCR from human genomic DNA. UPLC traces (b, c) and time course (d) of reactions performed on DNA from b) a Chorea Huntington patient (HD-DNA) or c) a healthy individual (WT-DNA). Conditions: see Figure 3.

amplification of Huntington Disease genomic DNA (HD-DNA) triggered the formation of ligation product **6** (Figure 4a). The ligation product was detectable after 10 min reaction time (10 % yield) and reached 32% in 180 minutes. By contrast, no product was formed when the reaction system involved the wild-type DNA template (Figure 4b). The analysis was repeated with partial blocker **8** (Figure S44-S45). Again, no product was formed when the template was produced from wild-type genomic DNA, whereas UPLC analysis revealed the time-dependent synthesis of ligation product **6** in reactions including the HD-DNA template. We wish to note that the signal detected via RP-UPLC is robust and can be obtained nearly as rapidly as by a fluorometric read-out. Based on peak height the signal-to-noise on a genomic DNA template is 46 (after 10 min reaction time). The data shows that the PNA-based reaction system responds to the length of repeat sequences in a way that enables ligation when a certain threshold is surpassed. The combined use of reactive probes and auxiliary probes provided a clear yes/no response to the diagnosis problem. The comparison with data obtained with synthetic DNA templates suggested that the HD-DNA has more than 28 repeats. The rate profile resembles reactions with > 40 repeat sequences (Figure S44-S45).

The reported work represents the first attempt to assess the length of triplet repeat DNA by means of a DNA-templated reaction. The PNA native chemical ligation provided detectable product after 10 min reaction time and the UPLC analysis was obtained within 4 minutes. Therefore, the time required for analysis by PNA ligation is shorter than the time required for the commonly used analysis of PCR products by gel electrophoresis. However, the need for single stranded DNA template adds capture and denaturation steps to the PCR protocol and, thereby, increases the total analysis time. Sequencing methods could, in principle, provide a more direct means of counting the number of triplet repeats. Yet, long read lengths are required in order to avoid ambiguities in alignment. The resulting need for the separation of the differently sized DNA amplicons calls for electrophoretic methods, which also are time consuming. A reduction of the analysis time required in our method should be feasible. We propose the use of PNA-based invaders such as the G-clamp modified  $\gamma$ PNA<sup>39</sup> introduced by Ly as this would make the capture/denaturation step redundant.

## Conclusions

We identified a reaction system, which allowed the discrimination between Chorea Huntington DNA and wild-type DNA. The reaction system relied on reactive PNA probes, which contained thioester and cysteinyl units required to drive a native chemical ligation reaction. Of note, aggregation of repeat sequence PNA and the formation of template superstructures presented a formidable challenge. A solution to the problem of false positive ligation and folding of large repeat templates was provided by the combined use of amino-modified PNA probes and auxiliary, non-reactive DNA hybridization probes. Under these conditions, the native chemical PNA ligation proceeded at high rate and afforded detectable signals after 10 min reaction time provided that the template was longer than the auxiliary probe (partial blocker). The length-reactivity

relationship is within a range that allows the characterization of various polyglutamine diseases.

## Notes and references

<sup>a</sup> A. Kern, Prof. Dr. O. Seitz

Institut für Chemie, Humboldt-Universität zu Berlin

Brook-Taylor-Straße 2. 12489 Berlin (Germany)

Fax: (+49) 30-2093-7266

E-mail: oliver.seitz@chemie.hu-berlin.de

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/c000000x/

This work was supported by Deutsche Forschungsgesellschaft (DFG).

1. K. Gorska and N. Winssinger, *Angew. Chem. Int. Ed.*, 2013, **52**, 6820-6843.
2. X. Li and D. R. Liu, *Angew. Chem. Int. Ed.*, 2004, **43**, 4848-4870.
3. S. M. Gryaznov, R. Schultz, S. K. Chaturvedi and R. L. Letsinger, *Nucleic Acids Res.*, 1994, **22**, 2366-2369.
4. A. P. Silverman and E. T. Kool, *Chem. Rev.*, 2006, **106**, 3775-3789.
5. A. A. Martí, S. Jockusch, N. Stevens, J. Ju and N. J. Turro, *Acc. Chem. Res.*, 2007, **40**, 402-409.
6. A. Shibata, H. Abe and Y. Ito, *Molecules*, 2010, **17**, 2446-2463.
7. T. N. Grossmann, A. Strohbach and O. Seitz, *ChemBioChem*, 2008, **9**, 2185-2192.
8. J. Michaelis, A. Roloff and O. Seitz, *Org. Biomol. Chem.*, 2014, **12**, 2821.
9. E. M. Harcourt and E. T. Kool, *Nucleic Acids Res.*, 2012, **40**, e65.
10. A. Roloff and O. Seitz, *Bioorg. Med. Chem.*, 2013, **21**, 3458-3464.
11. Y. Xu, N. B. Karalkar and E. T. Kool, *Nat Biotechnol.*, 2001, **19**, 148-152.
12. S. Sando, H. Abe and E. T. Kool, *J. Am. Chem. Soc.*, 2004, **126**, 1081-1087.
13. A. P. Silverman and E. T. Kool, *Trends Biotechnol.*, 2005, **23**, 225-230.
14. H. Abe and E. T. Kool, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 263-268.
15. R. M. Franzini and E. T. Kool, *J. Am. Chem. Soc.*, 2009, **131**, 16021-16023.
16. K. Gorska, I. Keklikoglou, U. Tschulena and N. Winssinger, *Chem. Sci.*, 2011, **2**, 1969-1975.
17. K. K. Sadhu and N. Winssinger, *Chem. Eur. J.*, 2013, **19**, 8182-8189.
18. H. Saneyoshi, Y. Ito and H. Abe, *J. Am. Chem. Soc.*, 2013, **135**, 13632-13635.
19. A. R. LaSpada, E. M. Wilson, D. B. Lubahn, A. E. Harding and K. H. Fischbeck, *Nature*, 1991, **352**, 77-79.
20. M. Mahadevan, C. Tsilfidis, L. Sabourin, G. Shutler, C. Amemiya, G. Jansen, C. Neville, M. Narang, J. Barceló, K. O'Hoy, S. Leblond, J. Earle-Macdonald, P. J. deJong, B. Wieringa and R. G. Korneluk, *Science*, 1992, **255**, 1253-1255.
21. J. D. Brook, M. E. McCurrach, G. G. Harley, A. J. Buckler, D. Church, H. Aburatani, K. Hunter, V. P. Stanton, J.-P. Thirion, T. Hudson, R. Sohn, B. Zemelmann, R. G. Snell, S. A. Rundle, S. Crow, J. Davies, P. Shelbourne, J. Buxton, C. Jones, V. Juvonen, K. Johnson, P. S. Harper, D. J. Shaw and D. E. Housman, *Cell*, 1992, **68**, 799-808.
22. C. Tsilfidis, A. E. McKenzie, G. Mettler, J. Barcelo and R. G. Korneluk, *Nat. Genet.*, 1992, **1**, 192-195.
23. Y. H. Fu, D. P. Pizzuti, M. Pieretti, J. S. Sutcliffe, S. Richards, A. J. M. H. Verkerk, J. J. A. Holden, R. G. F. Jr., S. T. Warren, B. A.

- Oostra, D. L. Nelson and C. T. Caskey, *Cell*, 1991, **67**, 1047-1058.
24. L. Margolis and C. A. Ross, *Clin. Chem.*, 2003, **49**, 1726-1732.
25. The Huntington's Disease Collaborative Research Group, *Cell*, 1993, **72**, 971-983.
26. C. R. L. Teo, W. Wang, H. Y. Law, C. G. Lee and S. S. Chong, *Clin. Chem.*, 2008, **54**, 964-972.
27. F. O. Walker, *The Lancet*, 2007, **369**, 218-228.
28. A. Mattes and O. Seitz, *Chem. Commun.*, 2001, 2050-2051.
29. C. Dose and O. Seitz, *Org. Biomol. Chem.*, 2004, **2**, 59.
30. S. Ficht, A. Mattes and O. Seitz, *J. Am. Chem. Soc.*, 2004, **126**, 9970-9981.
31. C. Dose, S. Ficht and O. Seitz, *Angew. Chem. Int. Ed.*, 2006, **45**, 5369-5373.
32. C. Dose and O. Seitz, *Bioorg. Med. Chem.*, 2008, **16**, 65-77.
33. P. E. Nielsen, M. Egholm, R. H. Berg and O. Buchardt, *Science*, 1991, **254**, 1497-1500.
34. M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S. M. Freier, D. A. Driver, R. H. Berg, S. K. Kim, B. Norden and P. E. Nielsen, *Nature*, 1993, **365**, 566-568.
35. P. Wittung, P. E. Nielsen, O. Buchardt, M. Egholm and B. Norden, *Nature*, 1994, **371**, 561-563.
36. A. Roloff and O. Seitz, *Chem. Sci.*, 2012, **4**, 432.
37. Y. P. Goldberg, S. E. Andrew, L. A. Clarke and M. R. Hayden, *Hum Mol Genet*, 1993, **2**, 635-636.
38. O. Riess, A. Noerremoelle, S. A. Soerensen and J. T. Epplen, *Hum Mol Genet*, 1993, **2**, 637-638.
39. V. Chenna, S. Rapireddy, B. Sahu, C. Austin, E. Pedrosso and D. H. Ly, *ChemBioChem*, 2008, **9**, 2388-2391.