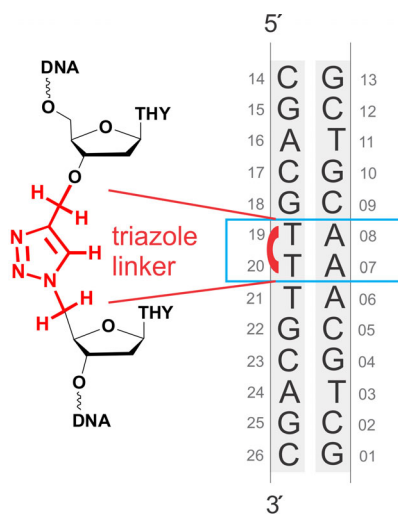


# Structure and dynamics of triazole-linked DNA - biocompatibility explained

André Dallmann,<sup>\*a,b</sup> Afaf H. El-Sagheer<sup>c,d</sup>, Lars Dehmel,<sup>a</sup> Clemens Mügge,<sup>a</sup> Christian Griesinger,<sup>e</sup> Nikolaus P. Ernsting<sup>a</sup> and Tom Brown<sup>\*c</sup>

The surge in molecular biology, genetic engineering and nanotechnology produces a strong demand for longer oligonucleotides, larger DNA and RNA constructs, and synthetic genes. However the synthesis of oligonucleotides with more than 100 bases by the phosphoramidite approach becomes increasingly inefficient in terms of yield and purity. This size barrier has recently been overcome by chemical ligation of several oligonucleotides using click chemistry,<sup>[1]</sup> whereby a triazole link is introduced in place of the natural phosphate backbone (Fig. 1). This is done using the CuAAC reaction,<sup>[2,3]</sup> an extremely efficient new version of Huisgen's 1,3-dipolar cycloaddition reaction.<sup>[4]</sup> Click chemistry has been used extensively in the nucleic acids field in recent years in a diverse range of applications.<sup>[5-10]</sup> The artificial linkage that is produced by click ligation is read through correctly and efficiently by DNA polymerases and is functional in vivo.<sup>[10]</sup> Moreover it supports catalytic activity at the ribozyme active site.<sup>[11]</sup>



**Fig. 1:** Duplex sequence of the Triazole-modified DNA (13merTAL). The Triazole linkage is marked by a red arc, its chemical structure and ligation to the adjacent nucleotides are shown enlarged. The sequence was chosen to enable comparison to the unmodified structure, which has already been solved.<sup>[12]</sup>

Therefore it is puzzling that thermodynamic destabilisation of the modified duplex is found by ultraviolet melting studies. Duplex instability normally occurs after incorporation of an incorrect nucleotide during replication. It triggers proofreading by the polymerase which excises several nucleotides from the daughter strand, after which replication resumes. At first sight one would not expect biocompatibility of the triazole linker. Here we resolve the apparent contradiction, and in doing so present the first NMR solution structure and base pair dynamics of a DNA duplex (13merTAL) where one strand contains a triazole linkage between two central T residues (cf. Fig. 1). Comparison with the unmodified duplex (13merRef, Dallmann et al.<sup>[12]</sup>) shows how structural and dynamic perturbations are distributed, and thus explains the biochemical results.

All NMR resonances could be assigned as with 13merRef, with the exception of the severely overlapped H5'/H5'' resonances. Integration and conversion of unambiguous, nontrivial NOESY crosspeaks yielded 351 distance restraints, which were used as primary experimental restraints to generate a locally well-defined B-DNA structure. Following this, a total of 39  $^1J_{CH}$  residual dipolar coupling (RDC) restraints (measured at natural abundance with a standard deviation of 0.6 Hz) were included into the refinement utilizing a single floating alignment tensor. The 10 best-energy, violation-free structures out of 100 calculated were used for further analysis. Their root-mean-squares-deviations (RMSD) are a measure of the precision of the calculation. In the following we discuss the corresponding average structures for 13merTAL and 13merRef<sup>[12]</sup> which are overlaid in Fig. 2. While the RMSD within either family of structures is very low (0.37 Å for 13merTAL and 0.30 Å for 13merRef), the RMSD between the two average structures is significantly higher (1.24 Å). This indicates structural changes upon the introduction of the triazole linkage, which are visible in Fig. 2.

To quantify the structural changes, helical parameters were calculated for each of the 10 best-energy structures with the program 3DNA.<sup>[13]</sup> Overall they

<sup>a</sup> Dr. A. Dallmann, Dipl.-Chem. L. Dehmel, Dr. C. Mügge, Prof. Dr. N. P. Ernsting  
Department of Chemistry, Humboldt-Universität zu Berlin  
12489 Berlin, Germany

<sup>b</sup> present address: Department of Chemistry, TU München  
85748 Garching, Germany  
E-mail: andre.dallmann@tum.de

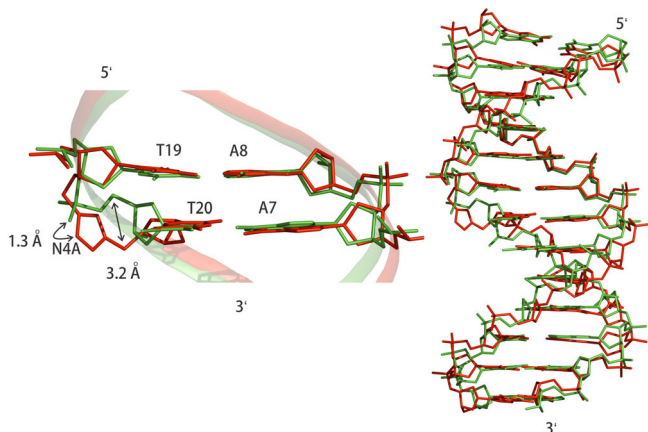
<sup>c</sup> Prof. Dr. Tom Brown and Dr. Afaf H. El-Sagheer  
School of Chemistry, University of Southampton

Highfield, Southampton. SO17 1BJ, UK  
E-mail: tb2@soton.ac.uk

<sup>d</sup> Dr. Afaf H. El-Sagheer  
Chemistry Branch, Dept. of Science and Mathematics, Faculty of Petroleum and Mining Engineering, Suez Canal University  
Suez, 43721, Egypt

<sup>e</sup> Prof. Dr. C. Griesinger  
Max-Planck Institute for Biophysical Chemistry  
37077 Göttingen, Germany

fall well within the ranges for standard B-DNA. This is backed up by  $^{31}\text{P}$  chemical shifts which do not change significantly (cf. Fig. S5). Perturbations are observed throughout the helix, but mainly for the triazole-linked and adjacent pairs (for details see section 6 of SI). The most pronounced changes are in base pair step parameters “rise” (A8-C9 by  $+0.86\text{ \AA}$ ), “shift” (A7-A8  $+1.41\text{ \AA}$ , A8-C9  $-0.95\text{ \AA}$ ), and “slide” (A6-A7  $-0.67\text{ \AA}$ , A7-A8  $-0.58\text{ \AA}$ , A8-C9  $-0.60\text{ \AA}$ ), and in base pair parameter “propeller twist” (A7:T20  $+20.5^\circ$ , A8:T19  $+22.5^\circ$  and C9:G18  $+15.0^\circ$ ). All of these indicate less optimal stacking interactions in the central region.



**Fig. 2:** Overlay of the average structures of 13merTAL (red) and 13merRef (green). The triazole linkage does not affect the overall conformation of the DNA duplex.

The T19-T20 base stacking step is compared in the close-up. The backbone of 13merTAL snakes down and up to accommodate its extra length. Thus base stacking between the two T:A base pairs and their neighbors is achieved. This requires the 5'-carbon of 13merTAL to point down, shifted by  $3.2\text{ \AA}$  relative to its position in 13merRef. The triazole ring nitrogen N4A of 13merTAL and a phosphate oxygen of 13merRef are only  $1.3\text{ \AA}$  apart, suggesting that the triazole is a plausible phosphodiester H-bonding surrogate.

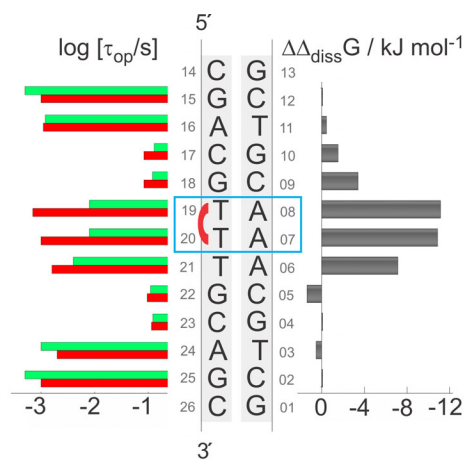
Thermodynamic destabilization of the duplex, when the triazole link is introduced, is found by ultraviolet melting studies. Both 13merRef and 13merTAL duplexes give sharp cooperative melting transitions at  $62.9^\circ\text{C}$  and  $55.3^\circ\text{C}$ , respectively, for  $3\mu\text{M}$  duplex concentration (see section 2 of SI). But the decrease of melting point is only a global measure of how the modification affects base pairing and base stacking. A more detailed view was already provided, above, in terms of structural changes for every base pair. Similarly the energetic changes can be assigned to specific base pairs, as will be discussed next.

Rates of water exchange reflect the structural rearrangements in the vicinity of the triazole. A first indicator of increased exchange is the fast decay with increasing mixing time of the diagonal T19 and T20 imino proton signals, as observed in the NOESY-spectrum of 13merTAL (cf. also<sup>[12]</sup>). We determined base pair lifetimes

( $\tau_{op}$ ) and inverse apparent dissociation constants  $1/K_D$ , using the relationship

$$\tau_{ex} = \tau_{op} + \frac{1}{K_D} \cdot \frac{1}{[B]} \quad (1)$$

The imino proton exchange time  $\tau_{ex}$  is measured as a function of the inverse catalyst concentration ( $1/[B]$ ). From a linear fit (see Fig. S14 and S15 in section 7 of SI)<sup>[14]</sup> we obtain extrapolated lifetimes and inverse apparent dissociation constants. These are compared in Fig. 3 for 13merRef and 13merTAL. Terminal base pair lifetimes could not be measured due to “base pair fraying” which is commonly observed at the helical termini.<sup>[15,16]</sup>



**Fig. 3:** Left part: Basepair lifetimes  $\tau_{op}$  compared for 13merRef (green bars) and 13merTAL (red bars). They were determined by measuring the  $T_1$  relaxation rates of imino proton signals, as a function of the concentration of TRIS base which catalyzes exchange with water.<sup>[12]</sup> The right part shows how the triazole linker affects the Gibbs energy  $\Delta_{diss}G$  for dissociating a base pair. The distribution is obtained from the apparent dissociation constants  $K_D$ .

Lifetimes for the inner base pairs in 13merRef become strongly reduced in 13merTAL, contrary to the outer pairs. The introduction of the triazole backbone reduces  $\tau_{op}$  for the directly connected pair A08:T19 (A07:T20) to 8.8% (12.5%) of the value for the unperturbed duplex. The activation energy for opening is therefore decreased, by  $-6.0$  ( $-5.5$ )  $\text{kJ/mol}$  in this case. The decrease is  $-2\text{ kJ/mol}$  for the adjacent pair A06:T21 and  $-1\text{ kJ/mol}$  for G18:C09, G10:C17 (cf. Fig. S15 in section 7 of SI). The entire center is affected dynamically by a single triazole backbone substitution (cf. Fig. 3).

Apparent dissociation constants are proportional to the Gibbs energy  $\Delta_{diss}G$  for dissociating the base pair. Therefore, the ratios  $K_D^{TAL}/K_D^{Ref}$  describe how the thermodynamic destabilisation by the triazole is distributed among the base pairs. In Fig. 3 we show the corresponding changes  $\Delta\Delta_{diss}G$  as gray bars. The destabilisation is seen to be spread over the three central A:T pairs and G18:C09.

The biocompatibility of the triazole linkage<sup>[10]</sup> can finally be explained. The two 13mer DNA duplexes examined in this study have different chemical structure solely in the backbone region between residues T19 and T20. The substitution introduces local changes in helical parameters but leaves the overall B-form intact. In particular, the bases on either side of the triazole linkage are able to form normal Watson-Crick base pairs. Therefore the base orientation required for correct recognition by DNA polymerase is preserved. But for wider biological function the site-specific interactions should also be retained. The triazole linker must mimic the local as well as global conformation of nucleic acids. Our NMR study indicates that the N4A atom of the triazole ring of 13merTAL lies close to one of the branched oxygen atoms at the 5'-side of T20 in 13merRef (see Fig. 2). The triazole template strand can thus mimic the hydrogen bond acceptor positions of the phosphate group in complex with DNA polymerase.<sup>[17]</sup>

Why does the thermodynamic destabilisation by the triazole linkage not prevent biocompatibility? The likely reason is that it is not concentrated on the triazole region alone, but is spread over 4 or 5 normal Watson-Crick base pairs. For all of the affected base pairs the opening rates are increased. These thermodynamic effects render the bases more flexible, compared to the native phosphate backbone. The increased linker length combined with a decrease in rigidity causes the 5'-carbon of T20 of 13merTAL to be shifted by 3.2 Å relative to its position in 13merRef (Fig. 2). Increased flexibility may be beneficial due to a more favorable entropy term for complex formation. For example DNA polymerases, which can process up to 1000 bases per second,<sup>[18,19]</sup> are limited by the base pair opening rates. This is consistent with the observation that PCR amplification of GC-rich DNA can only be achieved by addition of NaOH or betaine, which help to disrupt the stable secondary and tertiary structures.<sup>[20,21]</sup> Note however, that increased flexibility could also reduce stabilizing site-specific interactions, for example the inter-strand H-bonding or minor-groove narrowing that is observed in A-tract DNA.<sup>[22,23]</sup> Here the changes in the local environment due to the triazole linkage will certainly have an effect on DNA bending and rigidity, and thus DNA recognition.

The thermodynamics of the triazole linkage was explored in additional UV melting experiments. The increase in absorbance when a duplex melts (hyperchromicity) is a direct measure of base stacking. The base opposite T19 is chosen to be either matched (T19:A08 as in 13merTAL) or mismatched (T.C, T.G and T.T); corresponding results are collected in Table S1. A mismatch lowers the melting point by -10.9°C on average in this case. To place this in context, a similar loss of stability (-11.6°C, Table S2) is observed for mismatched Ref du-

plexes. Particularly informative is the average decrease in hyperchromicity which is caused by a mismatch: -18.8% for 13merTAL but only -7.5% for 13merRef. Clearly the combination of triazole linkage and mismatched base causes a severe loss of base stacking, much more than in the native mismatch duplex. We can now build a picture of replication at the site of the triazole linkage. When the correct nucleotide is selected a Watson-Crick base pair will form within the DNA helix and this will be accepted by the DNA polymerase. In contrast, on the rare occasion that a mismatched nucleotide is incorporated, the resulting major duplex instability and local melting will quickly lead to removal of the rogue nucleotide by the proofreading activity of the enzyme.

In conclusion, we have shown that a DNA duplex containing an isolated triazole linkage is structurally similar to normal DNA, and its relative instability is distributed over 4 or 5 base pairs. In contrast, mismatches around the triazole linkage cause major structural and thermodynamic perturbations. Thus, the biophysical results in this paper explain why only correct nucleotides are incorporated around the modified backbone during replication.

## Experimental Section

All NMR measurements were carried out at 298 K on a BRUKER AVANCE 600 at 3 mM duplex concentration. Structure calculations were performed with XPLOR-NIH v2.20. Duplex melting was monitored by optical absorption at 260 nm at total duplex concentration of 3.0 μM. Coordinate and input files are deposited at the protein databank with PDB ID codes 2kuz (13merRef) and 2l8i (13merTAL). Full experimental details are presented in the Supporting Information.

## Acknowledgement

We thank the Deutsche Forschungsgemeinschaft for a grant (ER 154/9) to NPE. AHE-S and TB have received funding from the European Community 7th Framework Programme [FP7/2007-2013], grant agreement [HEALTH-F4-2008-201418] entitled READNA.

## References

- [1] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angewandte Chemie Int. Ed. Engl.* **2001**, *40*, 2004–2021.
- [2] V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angewandte Chemie Int. Ed. Engl.* **2002**, *41*, 2596–9.
- [3] C. W. Tornøe, C. Christensen, M. Meldal, *The Journal of Organic Chemistry* **2002**, *67*, 3057–3064.
- [4] R. Huisgen, *Angewandte Chemie Int. Ed. Engl.* **1963**, *2*, 633–645.

- [5] J Gierlich, G. A. Burley, P. M. E. Gramlich, D. M. Hammond, T Carell, *Organic Letters* **2009**, *8*, 3639–3642.
- [6] P. M. E. Gramlich, S. Warncke, J. Gierlich, T. Carell, *Angewandte Chemie Int. Ed. Engl.* **2008**, *47*, 3442–3444.
- [7] D. M. Hammond, A. Manetto, J. Gierlich, V. a. Azov, P. M. E. Gramlich, G. a. Burley, M. Maul, T. Carell, *Angewandte Chemie Int. Ed. Engl.* **2007**, *46*, 4184–7.
- [8] F. Seela, V. R. Sirivolu, *Chemistry & biodiversity* **2006**, *3*, 509–14.
- [9] F. Seela, V. R. Sirivolu, *Helvetica Chimica Acta* **2007**, *90*, 535–552.
- [10] A. H. El-Sagheer, A. P. Sanzone, R. Gao, A. Tavassoli, T. Brown, *Proceedings of the National Academy of Sciences* **2011**, *108*, 11338–11343.
- [11] A. H. El-Sagheer, T. Brown, *Proceedings of the National Academy of Sciences* **2010**, *107*, 2–7.
- [12] A. Dallmann, L. Dehmel, T. Peters, C. Mügge, C. Griesinger, J. Tuma, N. P. Ernsting, *Angewandte Chemie Int. Ed. Engl.* **2010**, *49*, 5989–5992.
- [13] X.-J. Lu, W. K. Olson, *Nucleic Acids Research* **2003**, *31*, 5108–5121.
- [14] J.-L. Leroy, D. Broseta, M. Gueron, *Journal of Molecular Biology* **1985**, *184*, 165–178.
- [15] M Leijon, A Gräslund, *Nucleic Acids Research* **1992**, *20*, 5339–43.
- [16] S. Nonin, J.-L. Leroy, M. Gueron, *Biochemistry* **1995**, *34*, 10652–10659.
- [17] Y Li, S Korolev, G Waksman, *The EMBO journal* **1998**, *17*, 7514–25.
- [18] J. J. Schwartz, S. R. Quake, *Proceedings of the National Academy of Sciences* **2009**, *106*, 20294–9.
- [19] A. M. van Oijen, J. J. Loparo, *Annual review of biophysics* **2010**, *39*, 429–48.
- [20] R. K. Agarwal, a Perl, *Nucleic Acids Research* **1993**, *21*, 5283–4.
- [21] W Henke, K Herdel, K Jung, D Schnorr, S. a. Loening, *Nucleic Acids Research* **1997**, *25*, 3957–8.
- [22] N. V. Hud, V. Sklenar, J. Feigon, *Journal of Molecular Biology* **1999**, *286*, 651–660.
- [23] R. Stefl, H. Wu, S. Ravindranathan, V. Sklenar, J. Feigon, *Proceedings of the National Academy of Sciences* **2004**, *101*, 1177–1182.