Accepted Article

Title: Stable hairpin structures formed by xylose based nucleic acids

Authors: Charles-Alexandre Mattelaer, Mohitosh Maiti, Laurent Smets,
Munmun Maiti, Guy Schepers, Henri-Philippe Mattelaer,
Helmut Rosemeyer, Piet Herdewijn, and Eveline Lescrinier

This manuscript has been accepted after peer review and appears as an Accepted Article online prior to editing, proofing, and formal publication of the final Version of Record (VoR). This work is currently citable by using the Digital Object Identifier (DOI) given below. The VoR will be published online in Early View as soon as possible and may be different to this Accepted Article as a result of editing. Readers should obtain the VoR from the journal website shown below when it is published to ensure accuracy of information. The authors are responsible for the content of this Accepted Article.

To be cited as: *ChemBioChem* 10.1002/cbic.202000803

Link to VoR: <https://doi.org/10.1002/cbic.202000803>

FULL PAPER

Stable hairpin structures formed by xylose based nucleic acids

Charles-Alexandre Mattelaer,^{o[a]} Dr. Mohitosh Maiti,^{o[a]} Laurent Smets,^[b] Dr. Munmun Maiti,^[a] Guy Schepers,^[a] Dr. Henri-Philippe Mattelaer,^{§[a]} Prof. Dr. Helmut Rosemeyer,^[c] Prof. Dr. Piet Herdewijn,^{*[a]} and Prof. Dr. Eveline Lescrinier^{*[a]}

Dedicated to the 80th birthday of prof. Dr. W. Stec

- [a] Mr. Charles-Alexandre Mattelaer, Dr. Mohitosh Maiti, Mr. Laurent Smets, Dr. Munmun Maiti, Mr. Guy Schepers, Dr. Henri-Philippe Mattelaer, Prof. Dr. Piet Herdewijn, Prof. Dr. Eveline Lescrinier
Medicinal Chemistry
KU Leuven – Rega Institute for Medical Research
Herestraat 49 – box 1041, BE-3000 Leuven, Belgium
E-mail: Piet.Herdewijn@kuleuven.be; Eveline.Lescrinier@kuleuven.be
- [b] Mr. Laurent Smets
Research Centre for Operations Management, Faculty of Economics and Business, KU Leuven
Naamsestraat 69, BE-3000 Leuven, Belgium
- [c] Prof. Dr. Helmut Rosemeyer
Organische Materialchemie (OMC)
Universität Osnabrück – Institut für Chemie neuer Materialien
Barbarastraße 7, D-49076 Osnabrück, Germany

^o contributed equally to his manuscript

[§] current affiliation: Membrane Separations, Adsorption, Catalysis, and Spectroscopy for Sustainable Solutions (cMACS), KU Leuven (Arenberg), Celestijnenlaan 200f – box 2454, BE-3001 Leuven

^{*} co-corresponding authors for this manuscript

Supporting information for this article is given via a link at the end of the document.

Abstract: Xenobiology explores synthetic nucleic acid polymers as alternative carriers of genetic information to expand the central dogma. The xylo- and deoxyxylo-nucleic acids (XyNA and dXyNA), containing 3' epimers of riboses and deoxyriboses, are considered as potential candidates for an orthogonal system. In this study, thermal and spectroscopic analysis show that XyNA and dXyNA form stable hairpins. The dXyNA hairpin structure determined by NMR spectroscopy contains a flexible loop that locks the stem in a stable ladder-like duplex with marginal right-handed helicity. Reduced flexibility of the dXyNA duplex observed in the stem of the hairpin demonstrates that folding of dXyNA yields more stable structures as described so far.

Introduction

Xeno-nucleic acids (XNAs) are chemical analogues of natural nucleic acids, modified in at least one of their three main chemical moieties: nucleobase, sugar and phosphate backbone. XNA oligonucleotides and their monomers have a long and successful history in diverse areas of biotechnology (e.g. fluorescent base-modified nucleosides^[1]) and therapeutics (e.g. nucleoside analogues are a major class of anti-viral and anti-cancer compounds^[2,3], antisense and siRNA based oligonucleotides^[4,5] or aptamers^[6]). More recently, they entered the field of xenobiology as chemical information storage systems of genetic information that are capable of self-replication *in vitro* and *in vivo* for artificial life and biological computation^[7]. In order to develop

orthogonal life, it is important to have an available XNA species that is not able to 'cross talk' with natural nucleic acids but can support life (i.e. perform catalysis like RNA) and store genetic information like DNA^[8]. Examples of such orthogonal XNA's are p-RNA^[9], (deoxy)xylose nucleic acids^[10,11] and homo-DNA^[12–14]. They came into the spotlight as potential orthogonal genetic system since their homoduplexes exhibit similar or higher thermal stability to their DNA equivalents, while incorporating them into a regular DNA duplex leads to reduced thermodynamic stability. So far, only structures of homoduplexes – of which some are only published as 2D representations – are available on these XNA and these adopt ladder-like structures that differ significantly from helical structures that are typically formed by double stranded DNA and RNA^[9,11,15,16].

The 3D folded structure of nucleic acids is important for their biological function. Therefore, it is important to explore if orthogonal XNA can adopt other structural motifs than a double stranded helix. In RNA, hairpins are key functional and structural elements^[17], allowing enzyme-like activity: adding a metabolic role (e.g. ribozymes^[18]) to what was previously believed to be an information carrier. A typical example is the single stranded hammerhead ribozyme of merely 30 nucleotides in length folded in three hairpin motives to form its catalytic core.^[19] On the other hand, the presence of hairpin structures in DNA has been attributed to various functions in transcription and replication^[17] and are sought after recognition patterns for antigene therapies^[20] and more recently aptamers^[21,22]. These highly selective binders are usually single stranded oligonucleotide sequences where molecular interactions to the target is directed by the 3D folded

FULL PAPER

structure. To increase bioavailability of aptamers, specific synthetic nucleic acids have been developed which can be used *in vivo*. HNA aptamers containing a hexitol sugar^[22] or Spiegelmers based on L-sugars (in contrary to DNA/RNA having D-sugars)^[23] are both examples of biostable aptamers.

In previous publications, we have demonstrated the self-pairing, the absence of cross-pairing with DNA and the altered duplex geometry of xylose and deoxyxylose nucleic acids (XyNA and dXyNA, respectively) (Figure 1).^[10,11,15] MD simulations of XyNA and dXyNA duplexes revealed flexibility of their ladder-like structure.^[15] Recently, the free energy landscapes of both XyNA and dXyNA duplexes was explored to investigate their structural and dynamical properties. Left-handed helical and ladder-type structures were the two lowest energy structures of this landscape, with the left-handed helix as global minimum. A dynamic equilibrium between both forms is proposed with more flexible XyNA duplexes than their dXyNA analogues due to a lower energy barrier between both low energy conformations.^[24]

Figure 1. Structure of the nucleotide residues from which DNA, RNA, dXyNA and XyNA oligomers are built. Nucleobases are attached at position indicated by the B.

In this work, we investigated the potential of dXyNA and XyNA to adopt a secondary structure such as a hairpin in a sequence with a suspected C4 tetraloop and determined effect of the loop on the structure and dynamic nature of dXyNA. The formation of hairpins increases the potential of (d)XyNA to form additional secondary and tertiary structures and may broaden their potential functions in xenobiology. Additionally, for other applications such as aptamer development, extensive insight in the structural implications of hybrid or fully modified XNA sequences is needed.

Results and Discussion

Thermal (UV) and circular dichroism (CD) study

First, we have evaluated the possibility of dXyNA and XyNA to form hairpins in comparison with the same DNA and RNA

Figure 2. T_m^{-1} vs $\ln C$ for all oligonucleotides except 13 for which no T_m could be determined (see Figure S1). The profiles were divided in two plots to reduce overlap. T_m values were obtained via thermal (UV) analysis.

sequences, respectively (Table 1). One 12-mer and two 14-mer sequences were analyzed in the deoxy series and one 12-mer and one 14-mer sequence was analyzed on the ribose/xylose series. The T_m of all deoxyxylose and xylose sequences was considerably higher than those of the corresponding DNA and RNA sequences. Previously, higher T_m values have been reported for xylose based nucleic acids^[25] as well as a concentration-dependent behavior of selected oligonucleotides^[26]. Concentration-dependent T_m studies of our library show a predominantly straight curve (Figure 2) for all entries, i.e. little or no change in T_m in function of oligonucleotide concentration, indicating hairpin formation. Some variation is observable for oligonucleotides 1, 2, 7 and 8 but are not indicative of a possible equilibrium with duplex formation.

Table 1. Thermal analysis of DNA, RNA, dXyNA and XyNA sequences. T_m is reported in presence of 0.1 M NaCl at 260 nm with 2 μ M oligonucleotide concentration.

Entry	Sequence	Length	T_m (°C)
1	d(ATCCCCCGGAT)	12	43.8
2	dx(ATCCCCCGGAT)	12	64.6
3	d(CATCCCCCGGATG)	14	57.3
4	dx(CATCCCCCGGATG)	14	74.9
5	d(TATACTTCGGTATA)	14	47.2
6	dx(TATACTTCGGTATA)	14	64.8
7	r(AUCCCCCGGAU)	12	48.6
8	x(AUCCCCCGGAU)	12	80.2
9	r(UAUACUUCGGUAUA)	14	58.0
10	x(UAUACUUCGGUAUA)	14	69.6
11	d(AGCAATCC CCCC GGATTGCT)	20	73.8
12	d(AGCAATCC)-dx(CCCC)-d(GGATTGCT)	20	70.8*
13	dx(AGCAATCC CCCC GGATTGCT)	20	> 85

d = DNA, dx = dXyNA, r = RNA, x = XyNA,

* recorded at 1 μ M oligonucleotide concentration

In a second series of experiments, a 20-mer with sequence d(AGCAATCCCCCGGATTGCT) was selected. To dissect the effect of deoxyxyloses in the loop and stem, not only the fully modified oligonucleotide (13) was prepared but also a chimeric construct with DNA residues in the stem and deoxyxylose residues in the loop (12). Temperature-dependent CD spectra (Figure 3) were recorded for all three constructs. The study was performed in a buffer solution (sodium cacodylate, $MgCl_2$, NaCl at pH 7.0) and in distilled water (pH 7.0).

FULL PAPER

The CD spectra of the DNA (**11**) and mixed DNA-dXyNA (**12**) constructs are very similar and significantly differ from the typical profile observed for dXyNA duplexes. The full dXyNA hairpin (**13**) shows a comparable but not identical CD profile as previously reported for dXyNA duplexes in H₂O at pH 7.0.^[10] Interestingly in oligonucleotide **13**, there is almost no temperature dependence in the range of 10 °C - 80 °C nor is there a significant difference between the spectra in water and the buffer solution which is in contrast to the observed spectra for the dXyNA duplex.^[10] The T_m of the DNA-dXyNA chimeric construct (70.8 °C) is somewhat lower than the T_m of the DNA hairpin (73.8 °C) indicating a destabilization of the DNA hairpin when introducing four dXyC residues in the loop structure.

Figure 3. Temperature dependent CD spectra of oligonucleotides **11–13** in aqueous buffer solution containing sodium cacodylate (10 mM), NaCl (100 mM) and MgCl₂ (10 mM) at pH 7.0 (left) and H₂O at pH 7.0 (right). Legend and 2D overview of hairpin structure is given in the top right panel since no CD spectra were recorded for these conditions.

NMR diffusion experiments

Diffusion ordered spectroscopy (DOSY) experiments were used to determine the translational diffusion constant of **12** and **13** in a D₂O solution at 10 °C to determine the presence of a hairpin or a duplex structure in the NMR samples. Measured values (**12** $\approx 7.1 \cdot 10^{-11}$ m²/s, **13** $\approx 5.2 \cdot 10^{-11}$ m²/s) are in the same range as reported for dXyNA duplexes ($6.5\text{--}8.8 \cdot 10^{-11}$ m²/s) with 8 basepairs.^[10] These results confirm the hairpin fold of **12** and **13** in the NMR samples. Faster diffusion of **13** relative to the dXyNA duplex for which the NMR structure was determined ($8.8 \cdot 10^{-11}$

m²/s) could be indicative of a more compact structure for **13**. However, it is worth noting that the DOSY spectra for the dXyNA duplexes were recorded in high salt concentration conditions whereas no salt was added to the sample of **13** since this could promote duplex formation.

NMR resonance assignment of non-exchangeable protons

Standard methods were applied in the DNA stem of **12** to assign non-exchangeable protons starting from a classic anomeric-to-aromatic proton walk (Figure S2).^[27]

For **13**, a similar strategy was applied as reported for dXyNA duplexes.^[10] Only intrasidue H1' to nucleobase H8 (purine residues) or H6 (pyrimidine residues) interactions were observed in a NOESY spectrum with mixing time 150 ms at 25 °C in both the stem region as well as the loop region. All H1' signals in **13** had strong intrasidue NOE to H4'. From strong, well-resolved, intrasidue H1'–H4' interactions, sequential assignment could be achieved through weak interresidue H1'(n) H4'(n+1) (Figure 4). However, this walk was interrupted between residues 11 and 12 in the loop region. Additionally, the missing H1'(1) to H4'(2) interaction can be attributed to fraying of the duplex end at 25 °C using 150 ms mixing time.

For both **12** and **13**, the H2', H2'', H3', H4', H8, H6 and H7 (T-Me) protons of each individual residue were assigned through correlations observed in DQF-COSY, TOCSY and NOESY spectra. The H2 protons of adenine bases were assigned using ¹H-¹³C HMBC spectra correlating H2 and H8 protons via C4 in adenine.^[28]

The sequential assignment of exchangeable protons in **13** was finished by assigning H5' and H5'' through very strong intrasidue NOE correlation with the aromatic H8/H6 protons to H5''. Sequential H8/6 (n) to H5'(n+1) interactions confirmed the assignment obtained through H1'–H4' interactions.

NMR resonance assignment of exchangeable protons

Exchangeable protons were assigned using the NMR data in H₂O/D₂O (9/1) at 10 °C. For both **12** and **13**, six signals were clearly visible in the imino region of the 1D ¹H spectrum at 25 °C. Lowering the temperature to 10 °C broadened all signals and one additional signal appeared. The NOE patterns presented all the characteristics of a Watson-Crick type base pairing in the stem region. The imino H3 protons of thymine were identified by their strong correlation to H2 of their adenine base-pair partner. Correlations between imino protons provided assignment of H1 in guanine residues. All imino protons of the nucleobases were identified except those of the two terminal basepairs (G13:H1 and T20:H3) which could be attributed to their solvent exposure. All ³¹P signals are highly overlapping in a narrow ppm range (−0.69 to −1.63 ppm for **12** and −0.08 to −1.71 for **13**). The ¹H-³¹P HETCOR experiment allowed assignment of all ³¹P in **12**, but signals were too weak in the spectrum of **13** for assignment purposes.

FULL PAPER

Figure 4. Left: 2D overview of the hairpin structure. All residues are dXyNA. Right: Expansion of NOESY spectrum in D₂O (150 ms mixing time at 25 °C) for **13**. NOE walk between intrasidue H1'-H4' (labelled with residue number) and sequential H1'(n)-H4'(n+1) are indicated with a solid line for residues 2 to 11 and a dashed line for residues 12 to 20.

NMR experimental restraints

Distance restraints between non-exchangeable protons were derived from a NOESY spectrum with a mixing time of 150 ms. Based on isolated spin-pair approximation (ISPA)^[29], inter-proton distances were calculated and categorized 1.8-3.2 Å, 1.8-4.2 Å, 1.8-5.0 Å and 1.8-6.5 Å for strong, medium, weak and very weak cross-peak intensities, respectively. The calibration of NOE cross-peak intensities was performed against the H5-H6 cross-peak as internal standard.

Sugar puckers of the 2'-endo deoxyribose (residues 1-8 and 13-20 in **12**) and three deoxyxyloses in the loop of **12** (10-12) were inferred from the strong scalar couplings of H2' to H1' (and H3' in deoxyxylose) in combination with weak scalar couplings of H2'' to H1'. The S-type sugar conformation of these residues was further confirmed from strong NOE interaction between nucleobase protons H6 (pyrimidine) or H8 (purine) to H2'. Dihedral restraints were applied on DNA residues for H1'-C1'-C2'-H2' = 157.4° (± 20°), H2'-C2'-C3'-H3' = -34.6° (± 20°), H3'-C3'-C4'-H4' = -100.5° (± 20°).

Residues in **13** have nicely resolved signals at 25 °C. Deoxyxylose anomeric protons show one strong COSY cross-peak to H2'/H2'' while the second peak is very weak or even absent. The second proton covalently bound to C2' could be assigned to strong intrasidue NOE interactions with the anomeric proton and aromatic H8 or H6 in purine and pyrimidine nucleotides, respectively. Due to its proximity to the nucleobase, the latter was assigned as H2' while the proton with strong *J*-coupling was assigned as H2''. Together with NOE interactions observed between H2'' and H4', the observed intrasidue NOE and *J*-coupling interactions agree with a 3'-endo pucker of the deoxyxylose ring systems in **13**. Therefore, dihedral restraints could be applied for residues with resolved H2' and H2'' signals

(1-7, 9-12 and 14-20) that fit the above mentioned COSY and NOESY interactions. Restraints were applied to H1'-C1'-C2'-H2' = 99.2° (± 20°), H2'-C2'-C3'-O3' = 39.4° (± 10°), H3'-C3'-C4'-H4' = -162.0° (± 20°). Error margins for loop residues were adjusted to ± 30°.

The normal ³¹P chemical shifts of DNA residues were used to restrain α and ζ torsion angles (0° ± 120°).^[30] Additional backbone dihedral angle restraints in the stem of **12** are based on ³J_{H3'P} (ε = 180° ± 60°), ³J_{H4'H5'} and ³J_{H4'H5''} (γ = 36.3° ± 20°).

From COSY and ¹H-³¹P HETCOR experiments of **13**, *J*-coupling could be determined to derive restraints on backbone dihedral angles in **13**. Since ³J_{H5'P} and ³J_{H5''P} are both small (below 10 Hz), β was restricted to the trans region. Also, ³J_{H3'P} was below 10 Hz and as such ε was restrained to 250° ± 100° to exclude the forbidden region characterized with ³J_{H3'P} above 10 Hz. Resolved long range cross-peaks between H4'(n) and P(n+1) in residues 2, 8 and 13 are characteristic for a 'W' conformation of the involved atoms. This observation allowed to narrow ε restraints in these residues to 180° ± 60°. Both ³J_{H4'H5'} and ³J_{H4'H5''} are small, resulting in weak COSY and TOCSY signals between involved protons and allowing to restrain O5'-C5'-C5'-H4' in a *trans* conformation. Hydrogen bond restraints were implemented as NOE distance restraints on base pairs for which imino signals are observable in H₂O/D₂O (9/1) at 10 °C. An overview of all restraints used during NMR structure calculation and violations is summarized in Table 2.

Table 2. Summary of experimental restraints that were applied during torsion angle dynamics for generating an NMR structure ensemble and restrained MD.

	12	13
Intranucleotide distance	220	220
Internucleotide distance	169	162
Dihedral angle	123	106

FULL PAPER

Structure refinement of chimeric DNA-dXyNA hairpin (12)

Starting from an extended strand, 44 % of the structures calculated by torsion angle dynamics in vacuum fulfilled all experimental distance and dihedral restraints. Ten structures closest to the average of this ensemble were submitted to the next stage of restrained molecular dynamics (MD) in explicit water with ions.

MD simulation results indicate that overall mobility of dXyNA residues 9-12 in the loop region is increased, partly due to the lack of any stacking interactions, as illustrated by the RMSF (Figure 5). The stem, in contrast, shows to be a quite rigid region and its overall structure is hardly affected by the increased mobility of the loop residues.

Figure 5. RMSF (Å) averages for each residue of the chimeric DNA-dXyNA hairpin (**12**) after 20 ns of restrained MD simulation in explicit solvent. Averages were calculated after MD simulation for each of the ten selected models.

Backbone torsion angle variations for stem residues in the structural ensemble of **12** structures obtained after restrained MD simulations (Table S1) correspond to typical right-handed B-type helical values with divergent dihedral profiles for loop residues.^[27] In all residues, nucleobases were found to be in the anti-conformation based on glycosidic torsion angle χ values. Overall, residues in the rigid stem adopt an S-type (C2'-endo) conformation, while the increased flexibility in loop residues results in an overall greater standard deviation for both backbone torsions as well as puckering phase of the deoxyxylose rings (Table S1). The sugar moiety of residues xC9 and xC10 adopted more an N-type (C3'-endo) conformation in all ten models whereas residues xC11 and xC12 show fluctuation between the S-type and N-type conformation. A 'crankshaft' conformation is observable in two structures at α and γ dihedrals in residues xC12 and dG13. Unfortunately, it was not possible to determine $^3J_{H4^1/H5^1/5'}$ couplings in these residues due to overlap in the COSY spectrum. This would have allowed to retrieve experimental data on the conformation of γ torsion angle in these residues.

In all obtained structures, stacking of the base pairs in the stem is extended with residue xC9 of the loop. In contrast, canonical tetraloop structures usually extend stacking interactions of the stem with three residues in the loop where only one residue forms a sharp turn.^[17] This reduced amount of stacking interactions is probably the reason for the reduction in T_m (70.8 °C) compared with **11** consisting of only DNA nucleotides (T_m = 73.8 °C). Since the introduction of four deoxyxylose residues in the loop structure destabilizes the DNA hairpin, the significant increase in T_m observed in dXyNA of different length (12 and 14 nucleotides) cannot be attributed to stabilizing effects in the loop region. The results hint towards interactions in the stem region of fully

modified dXyNA hairpins which should be responsible for their increased stability.

Structure refinement of dXyNA hairpin (13)

Torsion angle dynamics in vacuum starting from an extended strand yielded 96 structures out of 100 that fulfilled all experimental distance and dihedral restraints. Ten structures with the least deviations from the experimental restraints were submitted to the next stage of restrained MD in explicit water and ions.

Backbone torsion angle variations of stem residues in the ten structures of **13** after restrained MD simulation (Table S2) correspond greatly to observed ladder-like duplex in dXyNA.^[10,15] In all residues, nucleobases were found to be in the anti-conformation based on the glycosidic torsion angle χ values. In contrast to the chimeric variant (**12**), backbone torsions vary greatly from both A- and B-type duplexes. This results in a stem region which resembles more a ladder-like structure instead of a helix (Figure 6).^[10,15]

This fold resembles greatly the linear folds of p-RNA^[9] and homo-DNA^[16] characterized by a strong inclination of base pairs relative to the backbone. The backbone dihedrals of **13** adopt different values compared to p-RNA^[9] and homo-DNA^[16] (Figure S3), especially for 3' torsions ϵ and ζ . This is, however, not totally unexpected given the inversed chirality of C3' which opens up alternatives for steric optimization. In addition to the inversed chirality of C3', the conformation of pyranose rings in p-RNA and homo-DNA versus the furanose ring in dXyNA influence the difference in value of δ . Given the different chemical constitution of these XNA species, the helical stem of **13** is characterized by an almost identical slide (Table S3) between basepairs (on average 4.39 Å) as ladder-like duplexes of homo-DNA^[16] (4.4 Å). Subsequently, interstrand stacking can be observed for purine-purine or purine-pyrimidine motifs whereas pyrimidine-pyrimidine interstrand stacking is far less prominent. For **13**, this is observable for residues xT6:xT16 and xC3:xC19 whose orientation is also less ideal for intrastrand stacking interactions. Stacking of the stem is only extended via xC9 in 9 of the 10 structures. No extension of stacking interactions was observed in the remaining model. For the deoxyxylose puckering, the N-type conformation (puckering phase = 0° - 36°) dominates in all residues for this model.

Figure 6. Overview of **11** (left), **12** (center) and **13** (right) structures after MD simulations. dXyNA residues are indicated in gray.

In contrast to canonical hairpin structures (cfr. infra), extension of stacking interactions between stem and loop region is only

FULL PAPER

observed with maximum one residue (xC9) in the structural ensembles of **12** and **13**. This has been reported earlier for a FANA-RNA hybrid hairpin containing DNA residues in the loop.^[31] Our results agree with the fact that the chemical composition of the loop is the dominant factor in the structure of the loop since introduction of dXyNA residues in an otherwise unmodified hairpin (**12**) resulted in a similar overall structure for the loop compared to the fully modified hairpin (**13**).

Early thermodynamic studies indicated concentration-dependent variation in T_m for dXyNA oligonucleotides which would indicate an equilibrium between hairpin or duplex formation based on oligonucleotide concentration.^[26] The structural flexibility of xylose based nucleic acids was further investigated using NMR spectroscopy^[10,11] and MD simulations^[15]. These experiments indicated that (deoxy)xylose based nucleic acids show a structural versatility in the adaptation of different helical forms, i.e. left-handed helix or ladder-like helix. NMR spectra of **13** do not show multiple sets of signals originating from different duplex forms in solution and its CD spectra do not show temperature dependence, as reported for dXyNA previously.^[10] Apparently, the closing loop locks the stem of **13** in a single conformation and prevents it to adopt the alternative duplex that exists for self-complementary dXyNA octamers. The loop itself is the most flexible region, with only xC9 tending to stack onto the closing basepair of the stem in both **12** and **13**. The increased stability of the dXyNA duplex thus allows the absence of loop stabilizing interactions.

In silico model of DNA hairpin (**11**)

To examine the influence of the dXyNA residues on the overall hairpin structure, an *in silico* model of **11** – containing only DNA residues – was generated based on the observed base pair profile in both **12** and **13**. The model was refined using a combined approach of restrained simulated annealing (100 ns) in implicit solvent followed by an unrestrained explicit solvent MD simulation (20 ns). In the obtained model, stacking of the base pairs in the stem is extended with residue dC9 and dC10 at the 5' side and with residue dC12 at the 3' side. As previously described, tetra- or pentaloops are preferred for DNA hairpins since these minimize steric hindrance while optimizing stacking interactions in the hairpin.^[17] The obtained model corresponds to this description since only a single residue forms the sharp turn and the other three residues extend stacking interactions of the stem, in this case both at the 5' start and 3' end of the loop. This increased amount of stacking interactions with dC9, dC10 and dC12 provides increased stabilization of the full DNA hairpin in contrast to the chimeric construct (**12**). In all residues of this DNA oligomer, sugars adopt an S-type conformation and nucleobases were found to be in the anti-conformation. Additionally, backbone dihedral angles in the stem correspond to those of B-type helices. Taken together, the *in silico* model of **11** resulted in a canonical hairpin fold. As such, the overall influence of the dXyNA residues in the loop of **12** only have a minor influence on the overall structure of the hairpin. The major difference is limited to the loop where less stacking interactions are observed.

Conclusion

Experimental data demonstrate that the inversion of chirality of the C3' stereogenic center of all five-membered residues in the

sugar phosphate backbone of natural nucleic acids does not hamper the formation of a hairpin fold for deoxyxylose and xylose nucleic acids.

Combining thermal analysis with CD and NMR spectroscopy, we studied the structural characteristics of several modified hairpins. Thermal melting profiles for different sequences and different lengths of dXyNA and XyNA oligonucleotides were characteristic for hairpin formation. The structure elucidation of a 20-mer dXyNA hairpin showed the structural preference for a ladder-like stem in the stem of the studied hairpin. Apparently, the presence of a closing loop limits the flexibility of the dXyNA duplex that otherwise exists in a dynamic equilibrium of different helical forms as observed in NMR and molecular dynamics studies on dXyNA duplexes^[10,15,24].

Presence of dXyNA residues in the loop, both in a chimeric construct as in a fully modified sequence, results solvent exposed nucleobases in the loop that have less stacking interactions with stem residues.

The ability of dXyNA to fold in a highly stable hairpin with increased flexibility in its loop suggest that more sequential variety in this loop could be possible for target interactions compared to natural nucleic acids without compromising fold stability. These interesting findings makes dXyNA an excellent candidate for aptamer development, next to its potential as an orthogonal carrier of genetic information.

Finally, the refined structures of both the chimeric DNA-dXyNA as well as the full dXyNA hairpin have been deposited in the PDB database (entry 7BFS and 7BFX, respectively).

Experimental Section

Oligonucleotide synthesis and design

The 2'-deoxyxylose nucleosides were synthesized following classical nucleoside chemistry starting from 2'-deoxyribose nucleosides. In the pyrimidine series, the stereochemistry at the 3'-position is inverted via the intermediate formation of an 2,3'-anhydro bond. In the purine series, this epimerization reaction is carried out by a Mitsunobu-type reaction on the 3'-O-mesylate. The 2'-deoxy-xylo nucleosides were converted in the phosphoramidate building blocks for oligonucleotide synthesis. The nucleobases were protected with a dimethylamino formamidino group (G, A) or with a benzoyl group (C, U). Details of oligonucleotide synthesis have been previously described.^[10]

The easier synthesis of cytosine analogues combined with the general trend for optimal stability in tri- or tetraloops determined our choice for loop in the selected sequences. Additionally, identical bases in the loop would reduce the chance of possible interactions between loop residues as observed in other types of loops, e.g. GNRA loops.^[32] For NMR analysis, the choice fell on a longer, but similar sequence with C4-loop present in the ColE1 Cruciform wild-type.^[33]

UV melting experiments

Samples were prepared either in a buffer solution containing 100 mM NaCl, 0.1 mM EDTA and 20 mM potassium phosphate (pH 7.5). The concentration of the oligonucleotide was determined by measuring the absorbance at 260 nm at 80 °C and by assuming that the (deoxy)xylose nucleosides have similar extinction coefficients in the denatured state as natural nucleosides: $\epsilon = 15060 \text{ Lmol}^{-1}\text{cm}^{-1}$ (A), $\epsilon = 8560 \text{ Lmol}^{-1}\text{cm}^{-1}$ (T), $\epsilon = 12180 \text{ Lmol}^{-1}\text{cm}^{-1}$ (G), $\epsilon = 7100 \text{ Lmol}^{-1}\text{cm}^{-1}$ (C) due to the limited change in contribution of the isomeric sugar. Melting curves were measured with a Varian Cary 100 BIO spectrometer. The temperature of the solution was measured with a thermometer that was directly immersed in the cuvette. Temperature control and data acquisition were carried out automatically with an IBM compatible computer using Cary WinUV (Agilent, Santa Clara, USA) thermal application software. The samples were heated from 15 to

FULL PAPER

85 °C at a rate of 0.2 °C/min and were cooled at the same speed. Melting temperatures were determined by plotting the first derivative of the absorbance (at 260 nm) as a function of temperature.

CD experiments

Samples were measured in a temperature range of 10 °C – 80 °C with an interval of 10 °C employing a Jasco 600 (Jasco, Tokyo, Japan) spectropolarimeter in a thermostatically controlled 1 cm cuvette connected with a Lauda Res 6 bath. The oligomers were dissolved in water and in buffer containing sodium cacodylate (10 mM), MgCl₂ (10 mM), NaCl (100 mM) at pH 7.0.

NMR experiments

Spectra were recorded on a Bruker Avance II 600 MHz (1H) UltrashieldTM Plus with CP TCI 600 S3 H-C/N-D-05 Z probe (Bruker, Fällanden, Switzerland). Unless stated otherwise, NMR experiments were performed in D₂O and recorded at 10 °C. Since signals of **13** are broadened at 10 °C while nice linewidths are observed at 25 °C, spectra were recorded at 25 °C for this sample. All data were processed using Topspin v3.2 provided by Bruker and the resulting spectra were analyzed using the computer aided resonance assignment (CARA) software.^[34] Lyophilized oligonucleotides were dissolved in D₂O or in H₂O/D₂O (9/1). Concentration of oligomers was 2.9 mM and 1.9 mM for **12** and **13**, respectively. Samples were annealed prior to NMR experiments by briefly heating at 70 °C and cooling in an ice bath. The pD was adjusted to 6.6 and 6.9 for **12** and **13** samples, respectively. The DQF-COSY (³¹P decoupled), TOCSY and NOESY^[35] spectra with pre-saturation were recorded using 64 scans, 2048 complex data points in the *t*₂ dimension and 1024 increments in the *t*₁ dimension. Recordings of NOESY spectra employed mixing times of 150 ms and 300 ms. For the TOCSY experiment, a clean MLEV17 pulse program was used with a mixing time of 60 ms. For all homonuclear [¹H,¹H] spectra in D₂O, a spectral width of 6009.62 Hz (10.0138 ppm) in both dimensions was used. The NOESY spectra in H₂O/D₂O 9/1 also acquired recordings at 10 °C to reduce the exchange with water and were performed using 64 scans for each FID, 2048 data points in the *t*₂ dimension and 1024 increments in the *t*₁ dimension. The water suppression was achieved by excitation sculpting with gradients.^[36] The spectral width was 13227.51 Hz (22.0410 ppm) in both *t*₁ and *t*₂ dimensions.

¹H-³¹P HETCOR^[37] spectra were acquired with 256 scans, 4096 data points in the proton dimension (*t*₂) and 256 increments in the phosphorus dimension (*t*₁) over sweep widths of 4000 Hz and 808 Hz, respectively. A natural abundance ¹H-¹³C HSQC spectrum for sugar nuclei was recorded with coherence selection optimized for CH-groups (¹J_{CH} = 150 Hz) with 32 scans for the **12** sample and 64 scans for the **13** sample, 2048 complex data points in the *t*₂ dimension and 1024 increments in the *t*₁ dimension. The spectral width was 8.50 ppm and 120 ppm for the *t*₂ and *t*₁ dimension, respectively. A ¹H-¹³C HSQC^[38] spectrum to correlate nucleobase protons (H2, H6 and H8) was also recorded with coherence selection optimized for CH-groups (¹J_{CH} = 180 Hz) using 64 scans for the **12** sample and 128 scans for the **13** sample, 1024 complex data points in the *t*₂ dimension and 256 increments in the *t*₁ dimension. The spectral width was 10.01 ppm and 39.99 ppm for the *t*₂ and *t*₁ dimension, respectively. The ¹H-¹³C HMBC spectrum^[28] was optimized for nucleobase protons (H2 and H8 of adenine residues) coupling to C4 (³J_{CH} = 200 Hz) with 64 scans for the **12** sample and 128 scans for the **13** sample. The HMBC spectra were recorded using 256 complex data points in the *t*₁ dimension for both samples. The *t*₂ dimension of **12** covered 4096 complex data points, while 2048 complex data points were used for the *t*₂ dimension of **13**. The spectral width for both NMR spectra was 10.01 ppm and 70.00 ppm for the *t*₂ and *t*₁ dimension, respectively.

A diffusion ordered spectroscopy (DOSY) experiment^[39] with stimulated echo and bipolar gradient pulses was used to determine the translational diffusion constant of **12** and **13** in D₂O at 10 °C. All 2D data were apodized in each dimension with a shifted sine-bell square function.

NMR solution structure

Restrained torsion angle dynamics starting from an extended strand and using NMR derived restraints. was performed with X-PLOR-NIH V3.851.^[40] In the topology file a patch was included to define the altered configuration

of C3' in deoxyxylose nucleic acid residues. After the torsion angle molecular dynamics round on **12** using 123 dihedral angle restraints, 169 interresidue and 220 intraresidue restraints, 44 structures out of 100 had converged to very similar structures with similar total energies and no violations of the NOE and dihedral restraints.

MD Simulations

Restrained molecular dynamics refinement was performed using the AMBER (v18) software package.^[41] The structures were solvated with water (TIP3P) in a truncated octahedral box with a 10 Å radius. Additionally, the solvated system was neutralized by adding 19 Na⁺ ions. The OL15 parameter set^[42] for DNA was employed with a modified topology to account for the inversed chirality of C3' for dXyNA residues^[11]. Solvated MD experiments followed the general procedure of minimization, heating, equilibration of density and production MD. The minimization consisted of 10 000 steps of steepest descent followed by 10 000 steps of conjugated gradient. After the minimization, the SHAKE algorithm^[43] was applied for bonds involving hydrogen atoms and the system was heated using the Langevin thermostat with a collision frequency of 1 ps⁻¹ to 300 K at constant volume in 50 ps (2 fs per timestep). Subsequently, the density was equilibrated at constant pressure for 500 ps (2 fs per timestep). Finally, production MD was performed for 20 ns (2 fs per timestep) at constant pressure. A cutoff value of 8 Å was employed for non-bonded interactions and all NMR derived restraints were applied with a force constant of 20 kcal/(mol*Å) during all simulations. The refined structures were analyzed with the Curves+ webserver (v1.31).^[44]

Acknowledgements

The authors thank the KULeuven (C14/29/102) for financial help. Charles-Alexandre Mattelaer is an SB PhD fellow at the Research Foundation – Flanders (FWO) (1S23217N). We thank Chantal Biernaux and Inge Aerts for editorial help.

Keywords: XNA • NMR Spectroscopy • Molecular Dynamics • Structural Elucidation • Aptamers

- [1] R. W. Sinkeldam, N. J. Greco, Y. Tor, *Chem. Rev.* **2010**, *110*, 2579–2619.
- [2] L. P. Jordheim, D. Durantel, F. Zoulim, C. Dumontet, *Nat. Rev. Drug Discov.* **2013**, *12*, 447–464.
- [3] M. K. Yates, K. L. Seley-Radtke, *Antiviral Res.* **2019**, *162*, 5–21.
- [4] J. Xia, A. Noronha, I. Toudjarska, F. Li, A. Akinc, R. Braich, M. Frank-Kamenetsky, K. G. Rajeev, M. Egli, M. Manoharan, *ACS Chem. Biol.* **2006**, *1*, 176–183.
- [5] A. Khvorova, J. K. Watts, *Nat. Biotechnol.* **2017**, *35*, 238–248.
- [6] A. D. Keefe, S. Pai, A. Ellington, *Nat. Rev. Drug Discov.* **2010**, *9*, 537–550.
- [7] V. B. Pinheiro, A. I. Taylor, C. Cozens, M. Abramov, M. Renders, S. Zhang, J. C. Chaput, J. Wengel, S.-Y. Peak-Chew, S. H. McLaughlin, P. Herdewijn, P. Holliger, *Science (80-.)*. **2012**, *336*, 341–344.
- [8] P. Herdewijn, P. Marliere, *Chem. Synth. Biol.* **2011**, *6*, 201–226.
- [9] I. Schlönvogt, S. Pitsch, C. Lesueur, A. Eschenmoser, B. Jaun, R. M. Wolf, *Helv. Chim. Acta* **1996**, *79*, 2316–2345.
- [10] M. Maiti, V. Siegmund, M. Abramov, E. Lescrier, H. Rosemeyer, M. Froeyen, A. Ramaswamy, A. Ceulemans, A. Marx, P. Herdewijn, *Chem. - A Eur. J.* **2012**, *18*, 869–879.
- [11] M. Maiti, M. Maiti, C. Knies, S. Dumbre, E. Lescrier, H.

FULL PAPER

- Rosemeyer, A. Ceulemans, P. Herdewijn, *Nucleic Acids Res.* **2015**, *43*, 7189–7200.
- [12] K. Augustyns, A. van Aerschot, C. Urbanke, P. Herdewijn, *Bull. des Sociétés Chim. Belges* **1992**, *101*, 119–130.
- [13] J. Hunziker, H.-J. Roth, M. Böhringer, A. Giger, U. Diederichsen, M. Göbel, R. Krishnan, B. Jaun, C. Leumann, A. Eschenmoser, *Helv. Chim. Acta* **1993**, *76*, 259–352.
- [14] M. Egli, P. Lubini, P. S. Pallan, *Chem. Soc. Rev.* **2007**, *36*, 31–45.
- [15] A. Ramaswamy, D. Smyrnova, M. Froeyen, M. Maiti, P. Herdewijn, A. Ceulemans, *J. Chem. Theory Comput.* **2017**, *13*, 5028–5038.
- [16] J. H. Sheehan, J. A. Smith, P. S. Pallan, T. P. Lybrand, M. Egli, *Crystals* **2019**, *9*, 1–14.
- [17] G. Varani, *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 379–404.
- [18] K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, T. R. Cech, *Cell* **1982**, *31*, 147–157.
- [19] K. R. Birikh, P. A. Heaton, F. Eckstein, *Eur. J. Biochem.* **1997**, *245*, 1–16.
- [20] M. Zacharias, *Biophys. J.* **2001**, *80*, 2350–2363.
- [21] I. Jeddi, L. Saiz, *Sci. Rep.* **2017**, *7*, 1–13.
- [22] E. Ereemeeva, A. Fikatas, L. Margamuljana, M. Abramov, D. Schols, E. Groaz, P. Herdewijn, *Nucleic Acids Res.* **2019**, *47*, 4927–4939.
- [23] D. Eulberg, S. Klussmann, *ChemBioChem* **2003**, *4*, 979–983.
- [24] D. J. Sharpe, K. Röder, D. J. Wales, *J. Phys. Chem. B* **2020**, *124*, 4062–4068.
- [25] H. Rosemeyer, F. Seela, *Nucleosides, Nucleotides and Nucleic Acids* **1995**, *14*, 1041–1045.
- [26] H. Rosemeyer, M. Krecmerova, F. Seela, *Helv. Chim. Acta* **1991**, *74*, 2054–2067.
- [27] S. S. Wijmenga, M. M. . W. Mooren, C. W. Hilbers, *NMR of Macromolecules. A Practical Approach*, Oxford University Press, New York, NY, **1993**.
- [28] M. J. P. van Dongen, S. S. Wijmenga, R. Eritja, F. Azorin, C. W. Hilbers, *J. Biomol. NMR* **1996**, *8*, 207–212.
- [29] B. A. Borgias, T. L. James, **1989**, pp. 169–183.
- [30] S. S. Wijmenga, B. N. M. van Buuren, *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *32*, 287–387.
- [31] J. F. Trempe, C. J. Wilds, A. Y. Denisov, R. T. Pon, M. J. Damha, K. Gehring, *J. Am. Chem. Soc.* **2001**, *123*, 4896–4903.
- [32] K. E. Richardson, M. S. Adams, C. C. Kirkpatrick, D. W. Gohara, B. M. Znosko, *Biochemistry* **2019**, *58*, 4809–4820.
- [33] N. B. Blatt, S. E. Osborne, R. J. Cain, G. D. Glick, *Biochimie* **1993**, *75*, 433–441.
- [34] R. L. J. Keller, *Swiss Fed. Inst. Technol. Zurich* **2004**, *1*–159.
- [35] B. Borah, J. S. Cohen, A. Bax, *Biopolymers* **1985**, *24*, 747–765.
- [36] T. L. Hwang, A. J. Shaka, *J. Magn. Reson. Ser. A* **1995**, *112*, 275–279.
- [37] V. Sklenář, H. Miyashiro, G. Zon, H. Todd Miles, A. Bax, *FEBS Lett.* **1986**, *208*, 94–98.
- [38] J. Schleucher, M. Schwendinger, M. Sattler, P. Schmidt, O. Schedletsky, S. J. Glaser, O. W. Sorensen, C. Griesinger, *A General Enhancement Scheme in Heteronuclear Multidimensional NMR Employing Pulsed Field Gradients*, **1994**.
- [39] C. S. Johnson, *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 203–256.
- [40] C. D. Schwieters, J. J. Kuszewski, N. Tjandra, G. M. Clore, *J. Magn. Reson.* **2003**, *160*, 65–73.
- [41] D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York, P. A. Kollman, *Univ. Calif.* **2018**.
- [42] M. Zgarbová, J. Šponer, M. Otyepka, T. E. Cheatham, R. Galindo-Murillo, P. Jurečka, *J. Chem. Theory Comput.* **2015**, *11*, 5723–5736.
- [43] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327–341.
- [44] C. Blanchet, M. Pasi, K. Zakrzewska, R. Lavery, *Nucleic Acids Res.* **2011**, *39*, 68–73.

FULL PAPER

Entry for the Table of Contents

Thermal and spectroscopic analysis of xylose (XyNA) and deoxyxylose nucleic acids (dXyNA) indicate the formation of stable hairpin structures. The solution structure of a fully modified dXyNA hairpin indicates the structural preference for a ladder-like stem with a non-canonical loop.

Institute and/or researcher Twitter usernames: not applicable