Published online 29 December 2015

Nucleic Acids Research, 2016, Vol. 44, No. 4 ***1937–1943***
doi: 10.1093/nar/gkv1513

The first crystal structures of RNA–PNA duplexes and a PNA-PNA duplex containing mismatches—toward anti-sense therapy against TREDs

Agnieszka Kiliszek[1], Katarzyna Banaszak[1], Zbigniew Dauter[2] and Wojciech Rypniewski[1,*]

[1]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland and [2]Synchrotron Radiation Research Section, MCL, National Cancer Institute, Argonne National Laboratory, Argonne, IL 60439, USA

Received November 10, 2015; Revised December 15, 2015; Accepted December 16, 2015

ABSTRACT

PNA is a promising molecule for antisense therapy of trinucleotide repeat disorders. We present the first crystal structures of RNA–PNA duplexes. They contain CUG repeats, relevant to myotonic dystrophy type I, and CAG repeats associated with polyglutamine diseases. We also report the first PNA–PNA duplex containing mismatches. A comparison of the PNA homoduplex and the PNA–RNA heteroduplexes reveals PNA's intrinsic structural properties, shedding light on its reported sequence selectivity or intolerance of mismatches when it interacts with nucleic acids. PNA has a much lower helical twist than RNA and the resulting duplex has an intermediate conformation. PNA retains its overall conformation while locally there is much disorder, especially peptide bond flipping. In addition to the Watson–Crick pairing, the structures contain interesting interactions between the RNA's phosphate groups and the Π electrons of the peptide bonds in PNA.

INTRODUCTION

RNA-mediated diseases are genetic disorders in which a mutation results in toxic mRNA rather than gene silencing or malfunctioning protein. In a large class of the RNA-mediated diseases, known as Tri-nucleotide Repeat Expansion Disorders (TREDs), the mutated gene contains an abnormally expanded microsatellite sequence consisting of trinucleotide repeats (1). Tetra-, penta- or even hexanucleotide repeats also are known to be pathogenic (2). In a normal gene, the number of repeated units is usually less than 30 but when it exceeds 40–50, pathology begins (3).

Myotonic dystrophy type 1 was the first neurological disorder attributed to pathogenic RNA. It is linked to CTG repeats located in the 3′ untranslated region (3′UTR) of *DMPK* gene (4). The transcript contains extended CUG repeats and shows abnormal affinity for several cellular proteins, including MBNL—the regulator of alternative splicing which acts antagonistically to CUG-binding protein (CUG-BP). The sequestration of MBNL lowers the level of available protein while the level of CUG-BP increases. The combined effect is a change in the alternative splicing pattern (5). Although splicing proceeds, the mRNA is incorrect for the type or age of the cell.

Another type of pathogenic microsatellites consist of CAG repeats. They are located mainly in coding regions and their translation leads to toxic proteins containing polyglutamine (polyQ) runs. However, there are indications that RNA is also involved in the pathogenesis (6–9). Transcripts containing CAG repeats co-localize with MBNL and other proteins (i.e. SRSF6, nucleolin, U2AF65) involved in splicing or other processes affecting the function of the cell (10–12).

Myotonic dystrophy and polyQ diseases usually affect adults. They are progressive with no effective treatment but in recent years therapeutic approaches are under intensive development (13,14). One approach relies on antisense oligonucleotides (ASO) which serve as blockers binding directly to complementary target to prevent interactions of the repeat RNA with nuclear proteins (15). They can also inhibit the expression of toxic polyQ proteins. To be effective they should exhibit high affinity and target selectivity, good cellular uptake and stability in the cell. Natural oligomers are not suitable for this purpose because they are sensitive to nucleases. Several modified ASOs have been proposed instead. A promising class of compounds are peptide nucleic acids (PNA) which mimic nucleic acids while containing charge-neutral N-(2-aminoethyl)glycine instead of electronegative sugar phosphodiester backbone (Figure 1) (16). Despite significantly different chemical nature, PNA hybridizes with high affinity to complementary RNA or DNA, obeying the canonical Watson–Crick rules (17). On the other hand, PNA exhibits outstanding base-pairing selectivity, which is manifest by low tolerance to even single mismatches. Mismatches destabilize PNA–DNA du-

*To whom correspondence should be addressed. Tel: +48 61 8528503; Fax: +48 61 8520532; Email: wojtekr@ibch.poznan.pl

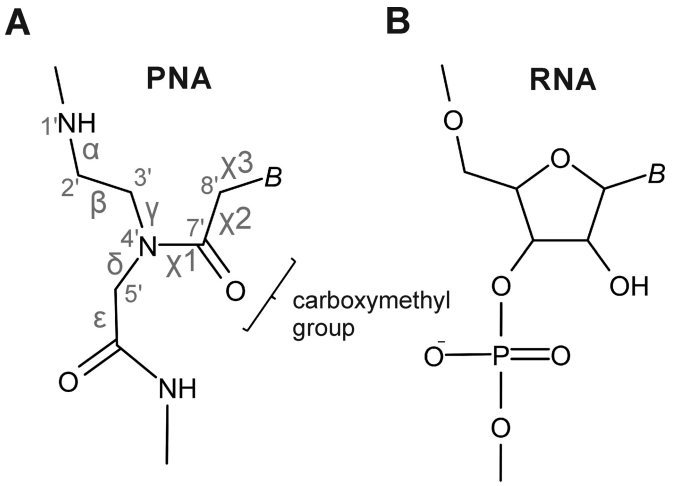

Figure 1. The chemical structure of PNA (**A**) and RNA (**B**). '*B*' stands for a nucleobase.

plexes by at least a factor of two, in terms of free energy, compared to the effect of mismatches on DNA–DNA duplexes (18–21). PNA is resistant to nucleases and proteases (22) and can be easily modified to adjust its permeability across the cell membrane (23). Promising results of using PNA against CAG repeats were obtained by Corey and colleagues (24,25). They targeted mutant transcripts of the huntingtin gene associated with Huntington disease (HD) and mutated mRNA of ATX3 related to spinocerebellar ataxia type 3 (SCA3). One of the antisense PNA, composed of six CTG repeats inhibited selectively the translation of the polyQ protein in HD cells, while not affecting the expression of genes containing normal CAG repeats and not causing cellular toxicity.

Structural studies aimed at understanding the molecular bases of the pathogenic mechanism of TREDs revealed that extended CUG and CAG repeats formed hairpin structures. The main part of the hairpin is the double-stranded stem which interacts with proteins. Crystallography revealed that duplexes of the CNG tracks, representing the stem of the hairpin, have the form of A-RNA consisting of G–C and C–G pairs interrupted by non-canonical N–N pairs which bestow specific properties on the repeats (26).

The main goal of the research presented here was to solve the crystal structures of RNA containing CUG and CAG repeats in complex with antisense PNA. We wanted to see how the PNA recognized the RNA and how it affected the structure of the repeats. We also wanted to understand PNA's outstanding sequence selectivity. Here we present crystal structures of two duplexes of RNA, containing CUG and CAG repeats, with complementary PNA; and one PNA–PNA duplex with T–T mismatches. To our knowledge these are the first crystallographic RNA–PNA structures. The PNA–PNA model is the first example of a PNA duplex containing non-Watson–Crick pairs.

MATERIALS AND METHODS

Synthesis, purification and crystallization of the RNA–PNA and PNA–PNA duplexes

The PNA oligomers were synthesized and purified commercially in the laboratory of prof. Piotr Rekowski from the University of Gdańsk. Crude RNA oligomers were purchased from Future Synthesis and were subsequently purified using the TLC method on silica gel plates with ammonia/1-propanol/water solvent. RNA was eluted with water and lyophilized under vacuum using Speed-Vac.

All oligomers were dissolved in water and then mixed to the final concentration of 1 mM each. $MgCl_2$ was added only to the solution of r(GCAGCAGC)-p(GCTGCTGC). Before the crystallization, all PNA and RNA mixtures were annealed for 3–5 min at 95°C, then cooled to ambient temperature within 2–3 h. Crystals were grown by the sitting drop method at 19°C. The crystals of r(GCAGCAGC)-p(GCTGCTGC) grew in 40 mM $MgCl_2$, 50 mM HEPES pH 7.0 and 1.6 M $(NH_4)_2SO_4$. The crystallization media of the second r(GCUGCUGC)-p(GCAGCAGC) duplex contained 200 mM KCl, 50 mM $MgCl_2$, 50 mM TRIS pH 7.5 and 10% PEG 4000. The $p(GCTGCTGC)_2$ duplex crystallized in 6 mM $MgCl_2$, 0.03 M HEPES-Na pH 7.0 and 2.4 M LiCl.

X-ray data collection, structure solution and refinement

X-ray diffraction data were collected on the BL 14.2 beam line at the BESSY synchrotron in Berlin and on EMBL P13, PETRA III, Hamburg. Only the PNA–PNA crystal was additionally cryoprotected by 3M sodium formate (v/v) in the mother liquor. All three crystals diffracted to atomic resolution (Table 1). The data were integrated and scaled using the program suite XDS (27) or DENZO and SCALEPACK (28). The space group and the cell parameters of the two RNA–PNA crystals were virtually the same (Table 1). The structure of r(GCAGCAGC)-p(GCTGCTGC) was solved by molecular replacement (MR) using PHASER (29). The search model was one RNA strand of GGCAGCAGCC oligomer (PDB code: 3NJ6). The phases from MR were used in a free-atom refinement using ARP/wARP (30). The R-factor/R-free statistics were improving continuously from the first cycle of adding/removing atoms and after 40 cycles they reached values of 34/37%. The free-atom electron density map showed the RNA and PNA atoms clearly resolved. The manual rebuilding and map inspection were done using Coot (31). The second r(GCAGCAGC)-p(GCTGCTGC) structure was solved by taking the phases from the first r(GCAGCAGC)-p(GCTGCTGC) model. The final refinement of all structures was performed in Refmac5 (32).

In the case of PNA–PNA, the space group was assigned as $P2_12_12_1$. POINTLESS (33) indicated that the space group could be Pbca but closer inspection revealed reflections which would be systematically absent in the centrosymmetric space group. Solution of the $p(GCTGCTGC)_2$ model was obtained by direct methods using SHELXD (34). A run of 1000 phase trials produced clear solutions with 62 phase sets characterized by the correlation coefficient (CC) all/weak larger than 59/32%,

Table 1. Data collection and refinement statistics

	rCAG-pCTG	rCUG-pCAG	p(GCTGCTGC)$_2$
Data collection			
Space group	$P3_121$	$P3_121$	$P2_12_12_1$
Cell dimensions			
a, *b*, *c* (Å)	35.2, 35.2, 68.9	35.1, 35.1, 69.1	28.8, 42.9, 64.3
Resolution (Å)	30.0–1.15 (1.22–1.15)[a]	30.0–1.14 (1.21–1.14)	30.0–1.06 (1.08–1.06)
R_{sym}	0.064 (0.75)	0.051 (0.90)	0.068 (0.60)
$I / \sigma I$	18.7 (3.0)	23.1 (2.0)	23.9 (2.2)
Completeness (%)	98.2 (96.4)	98.4 (90.3)	99.8 (99.7)
Redundancy	9.0 (8.7)	9.8 (7.4)	4.5 (4.1)
Refinement			
Resolution (Å)	30.0–1.15 (1.22–1.15)	30.0–1.14 (1.21–1.14)	30.0–1.06 (1.08–1.06)
No. reflections	16914 (891)	17374 (1109)	35227 (1464)
R_{work} / R_{free}	13.9 / 17.7	12.1 / 15.3	16.3 / 19.3
No. atoms			
RNA or PNA	555	416	1681
Ligand/ion	2	3	21
Water	124	121	229
B-factors (Å^2)			
RNA or PNA	12.7	12.5	11.4
Ligand/ion	46.9	24.5	20.6
Water	26.1	26.4	24.3
R.m.s. deviations			
Bond lengths (Å)	0.024	0.025	0.02
Bond angles (°)	2.8	2.6	2.8
PDB code	5EME	5EMF	5EMG

[a]Values in parentheses are for highest-resolution shell.

whereas all remaining trials had these values smaller than 43/21%. The CC all/weak values are the correlation coefficients between the observed (E_{obs}) and calculated (E_{calc}) normalized structure factors for 6757 reflections used for phase evaluation and a group of weaker reflections, not taking part in phase estimation, respectively. The best SHELXD phase trial (with CC all/weak of 61.63/36.02%) produced the atomic model comprising 340 atoms. The electron density map computed from this model was clearly interpretable and allowed building of the complete and correct sequence of residues in the PNA molecule, that was submitted to further refinement. Despite the high resolution, the PNA–PNA structure had a high level of disorder, especially in the backbone. Most of it could be modeled as two alternative conformation. Restrains for the PNA backbone were generated based on the small-molecule database CSD (35). The helical parameters were calculated with 3DNA (36). All pictures were drawn using UCSF Chimera (37) and PyMOL v0.99rc6 (38). Atomic coordinates of the crystallographic models have been deposited with the Protein Data Bank (accession codes 5EME, 5EMF and 5EMG).

RESULTS

rCUG-PNA and rCAG-PNA

X-ray structures of two RNA–PNA duplexes were analyzed: r(GCUGCUGC)-p(GCAGCAGC) and r(GCAGCAGC)-p(GCTGCTGC), abbreviated here as rCUG-PNA and rCAG-PNA. The statistics of the data and the final model are shown in Table 1. Both duplexes are fully complementary with Watson–Crick base pairs. They have the form of A-RNA with an unusually low twist: 26° for rCUG-PNA and rCAG-PNA, compared to 33° for a canonical A-RNA (Figure 2B and C). The rise is low (2.4 Å) compared to values found for A-helices 2.6–3.3 Å. The base pair inclination angle is 14°, somewhat lower than for a typical A-RNA (16°). Average X-displacement, ≈−6 Å, exceeds typical values of A-RNA (−5 Å) or B-DNA (−4 Å) (Table 2).

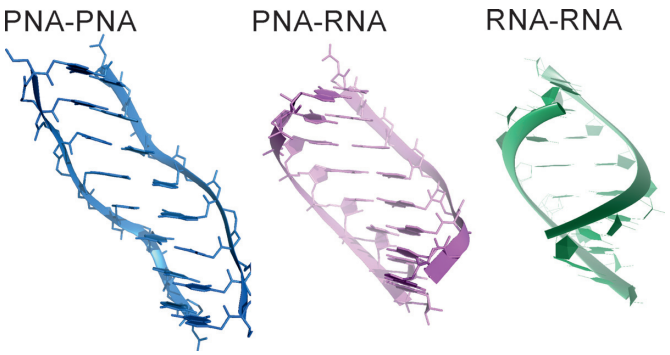

Figure 2. Comparison of the overall structure of mixed RNA–PNA duplex against PNA and RNA (pdb code 3glp) duplexes. PNA has a markedly lower twist than RNA and the mixed duplex has an intermediate value of the twist.

Crystal packing is similar in both duplexes. The 3′-end of each duplex is stacked end-to-end against a molecule related to it by the crystallographic 2-fold axis. Therefore, the 3′-end of the RNA strand is next to the N-terminus of the PNA strand of the other duplex, and *vice versa*. The other end of the duplex makes no base-stacking interactions but adjoins the PNA strand of a neighboring molecule (Supplementary Figure S1). The phosphate group of the 3′-terminal RNA residue interacts with the last peptide link on the PNA's C-terminal side. The distance between the phosphate O atom and the carbonyl C atom is 3.0 Å—closer than the sum of

Table 2. Helical parameters

	rCUG-PNA rCAG-PNA	$p(GCTGCTGC)_2$	A-RNA[a]	P-PNA[b]
Form	A	P	A	P
Twist(°)	26	20 (18 for G–C; 24 for T–T)	33	19
Rise (Å)	2.4	3.9	2.6-3.3	3.5
Inclination (°)	14	12 (9 for G–C; 19 for T–T)	16	3
x-displacement (Å)	−6	−9 (−6 for G–C; -15 for T–T)	−5	−7

[a]Parameters for A-RNA were taken from Saenger (1984) (40).
[b]The parameters were calculated using seven available 3D structures of PNA-PNA duplexes (PDB codes: 1PUP, 1RRU, 1QPY, 1HZS, 2K4G, 3MBS, 3MBU) (35,44–48).

the Van der Waals radii, and the O-C line is perpendicular to the peptide plane. The O atom also forms a salt bridge with the terminal amine group of another PNA molecule (distance = 2.9 Å) and a water molecule (2.8 Å) (Figure 3A). In the rCUG-PNA structure, the two P-O bonds appear to be unequal: the length of the interacting O atom is 1.47 Å (e.s.d. = 0.02 Å), the other bond is 1.42 Å long (e.s.d. = 0.03 Å). The difference, if true, would indicate an increased polarity of the longer bond. In the rCAG-PNA duplex, the distances cannot be measured reliably because of a double conformation of the RNA backbone. In the other phosphate groups, the P-O bonds appear equivalent. Another interaction (at 3.1 Å) between carbonyl C and O from a phosphate group is observed between guanine 4 and cytosine 5 residue from a neighboring duplex. At the other end of the duplex, the N-terminal carbonyl C from the carboxymethyl group is approached (3.1–3.2 Å) in a similar way by the O2' atom at the 5′ RNA end of another duplex (Figure 3B). The carboxyl groups of PNA are exposed to a large solvent-filled cavity and interact only with ordered water molecules.

PNA–PNA

Crystals of $p(GCTGCTGC)_2$ were obtained from a mixture of p(GCTGCTGC) and r(GCUGCUGC). Instead of RNA–PNA duplexes containing U-T mismatches, we obtained crystals of PNA–PNA with T–T non-canonical base pairs (Table 1).

PNA forms semi-infinite columns in which the right-handed and left-handed duplexes $p(GCTGCTGC)_2$ are stacked alternately (Supplementary Figure S2). Each duplex contains two T–T mismatches lodged between standard C–G and G–C pairs. The T–T pairs can be classified according to Leontis and Westhof as WC–WC pairs (39) or an asymmetric homo-pyrimidine XVI (40) (Figure 4). There are two hydrogen bonds between the thymine rings: N3H···O4 (2.7–2.9 Å donor–acceptor distance) and O2···HN3 (3.1–3.3 Å). One of the rings is inclined toward the minor groove more than the other. The distance between the PNA strands, measured between the C8' atoms, is 10.7 Å for the C–G pairs and 9.0 Å for the T–T pairs. The helical twist is low: 18° for the canonical base pairs and 24° for the T–T pairs. Rise is very high: 3.9 Å/b.p.. Angle is 5–11° for the canonical base pairs and 17–20° for T–T. The T–T pairs are associated with a jump in X-displacement between each T and the following G (Table 2; Figure 2A).

One of the T–T pairs forms an inner complex with a Na^+ ion in the 'major groove' (the term major groove is used by analogy with nucleic acids, but this side of the PNA duplex is in fact convex). The other T–T pair is associated with a Na^+ ion which interacts with it though ordered water molecules in the minor groove and forms an inner complex with the carbonyl O atom of the less inclined T. There are 12 ordered Cl^- ions: most interacting in the minor groove with the *exo*-amino groups of guanine rings or on the other side of the duplex, interacting with the *exo*-amino groups of the cytosine rings. Some are wedged between the backbones of the neighboring molecules and interact with the carbonyl C atoms.

PNA backbone conformation

The carbonyl O atoms of the carboxymethyl moieties point in the direction of the C-terminus and most make intrastrand CH···O bonds with the C8 atoms of guanine/adenine or C6 atoms of cytosine/thymine rings. In some residues, they also participate in CH···O interactions with the C8' atoms (donor–acceptor distance 3.1–3.3 Å) (Figure 5). All the amide bonds are *trans* and can be observed in one of two conformations differing by a 180° flip of the peptide plane. In one, the carbonyl O atom points toward the minor groove, in the other it points outward (Figure 5). The conformation of the peptide bond has little effect on the overall structure. This is clear from a comparison of the two duplexes, in the two different crystal structures, which are closely superposable although they have different distributions of the two conformations. Some residues are statically disordered, showing both conformations. The backbone dihedral angles at the Cα atom of the glycine moieties correspond to the two most common conformations found on the Ramachandran (ψ–φ) diagram for glycine residues in proteins. A detailed comparison of the PNA backbone conformations in this and other solved structures is given in Supplementary Table S1.

Comparison of PNA–RNA and PNA–PNA duplexes

A comparison with nucleic acids is not straightforward because the duplexes containing PNA have only one groove, corresponding to the minor groove of nucleic acids. The other side of the duplex has the form of a crest formed by the nucleobases (Supplementary Figure S3). The groove is filled with ordered water molecules.

PNA strands are hydrated in a characteristic way: in the minor groove each residue has an ordered water molecule wedged between the base and the peptide backbone. The

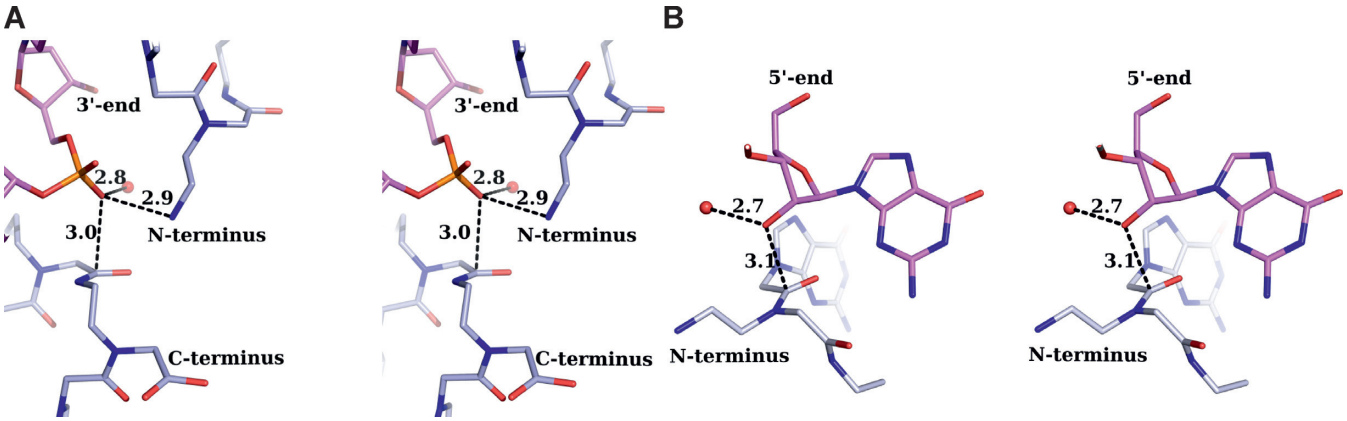

Figure 3. Stereo pairs showing Π…lone-pair interactions implied by the close contacts between the peptide plane of PNA and: the phosphate group of RNA (**A**), the hydroxyl O2′ atom of the ribose moiety (**B**). Both O atoms make a close approach to the electron-depleted C atom of the peptide bond. Distances are marked in Å. Water molecules are shown as red spheres.

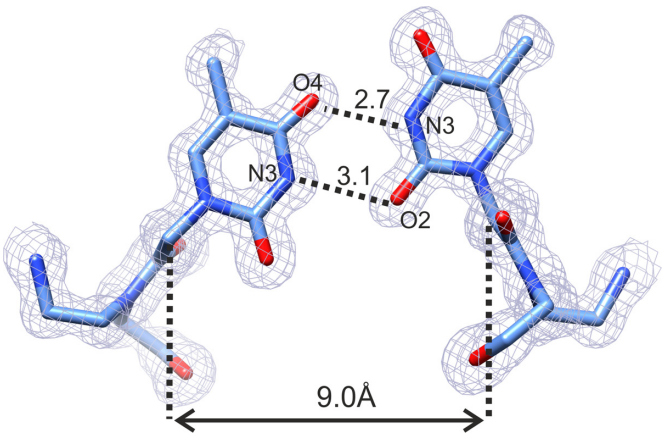

Figure 4. T–T base pairing interactions in the PNA–PNA duplex. Hydrogen bonds and donor–acceptor distances are indicated. The inter-strand distance is also shown, measured between the C8' atoms. Contours represent the $2F_o$-F_c electron density map at the 1 σ level.

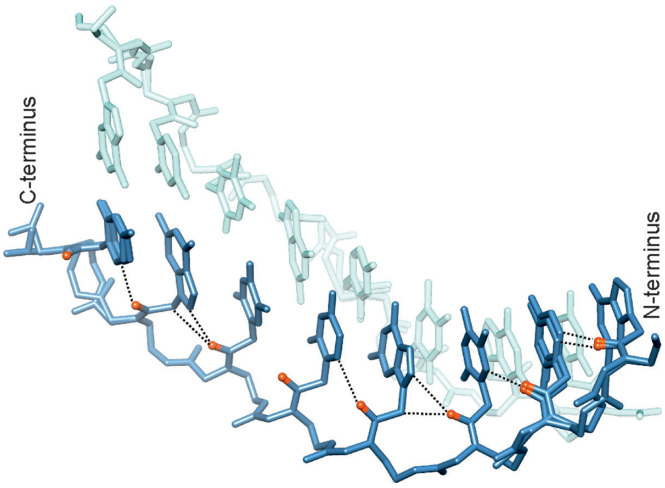

Figure 5. Intra-strand interactions in PNA. The carbonyl O atoms (red) of the carboxymethyl moiety usually make a H-bond with the neighboring nucleobase. The peptide planes are found in two possible conformations, differing by 180°. Some residues show static disorder with both conformations superposed, having only a local effect on the structure.

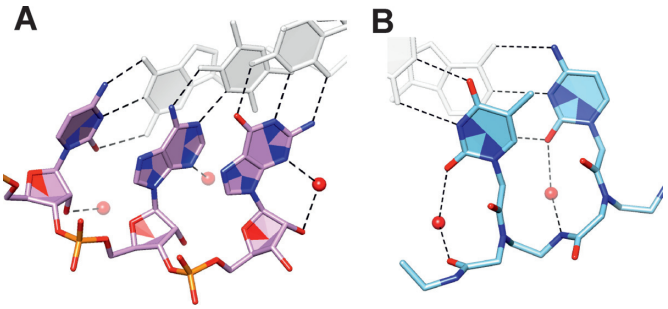

Figure 6. Hydration patterns in RNA (**A**) and PNA chains (**B**). Water molecules are shown as red spheres, hydrogen bonds are indicated with dashed lines. In PNA, water molecules link the nucleobases with the peptide backbone. Flipping of the peptide bond does not affect the position of the water molecule.

base always acts as a hydrogen bond acceptor (N3 in purines, O2 in pyrimidines) while the backbone supplies either an NH group or a carbonyl O, depending on the orientation of the peptide plane (Figure 6). Hydration of RNA strands in the minor groove consists of water molecules interacting with the 2′ hydroxyl groups and with the functional groups of the base rings (same as above). The ribose ring and the bases can share one water molecule or they can interact with separate waters.

Base-stacking is observed primarily within steps CG/GC, in both PNA–RNA (9–11 Å^2) and PNA–PNA (8 Å^2) duplexes. On the other hand, pairs T–T in PNA–PNA interact the least, and the effect is mainly cross-strand with guanine rings in step TG/CT (4 Å^2) (Supplementary Figure S4).

DISCUSSION

PNA is a fully synthetic molecule that has raised hopes and expectations as a tool in biotechnology and medicine (16). It can be delivered to the living cell and is resistant to proteases and nucleases, which makes it a promising therapeutic agent. PNA oligomers hybridize in a sequence-specific manner to DNA and RNA. Their strong sequence specificity is apparent also in this study. Only fully com-

plementary RNA–PNA duplexes could be obtained, while co-crystallization trials of mismatching RNA with PNA resulted only in crystals of mismatched PNA–PNA. Such sequence selectivity is a desirable property for antisense therapeutic strategies against TREDs, which target anomalous RNA structures containing mismatches while minimizing the risk of interfering with the genetic material.

PNA has been described as a flexible analogue of nucleic acids but it is clear that it has a well-defined intrinsic architecture, which has to be considered when explaining PNA's properties or designing strategies for applying it.

A comparison of the PNA–RNA and PNA–PNA structures reveals PNA's own conformational preferences and the effect it exerts within a heteroduplex. A homoduplex of PNA is strongly unwound (20°/bp, 18 bp/turn) compared to RNA or DNA, but when it accommodates complementary RNA, the resulting double helix shows an intermediate conformation (twist = 26°) between pure PNA and typical A-RNA (≈33°). The two studied RNA–PNA duplexes are isomorphic, having identical helical parameters despite different sequences. The comparisons indicate that RNA and PNA have a similar ease of adapting to one another and the resulting duplex is halfway between PNA–PNA and RNA–RNA structures.

The PNA p(GCTGCTGC) duplex can also be compared to an RNA duplex of similar sequence: $(GCUGCUGC)_2$ (41) (Figure 2). In addition to having a lower helical twist, PNA shows a greater backbone flexibility, allowing the thymine rings to come closer (inter-strand distance = 9.0 Å) and form two hydrogen bonds, whereas in RNA the predominant form is the 'stretched' U–U pair with only one H-bond and 10.4 Å between the strands. Stacking interactions in the two comparable oligomers are slightly more extensive in the RNA duplex.

The rules of sequence complementarity are strongly enforced in interactions between PNA and nucleic acids. None of the available PNA–DNA structures contain mismatches and our attempts to match RNA with PNA also resulted in fully complementary duplexes. On the other hand, we demonstrate that mismatches are tolerated within a PNA–PNA duplex. Perhaps mismatches in PNA homoduplexes could form as easily as in RNA structures. The relative instability of heteroduplexes containing mismatches can be explained by the additional need of the two strands to overcome their different conformational preferences (Figure 2).

Flipping of the peptide bond in PNA occurs readily and has no effect on the overall structure. Consequently, water molecules interacting with the carbonyl oxygen atom or the amine group have no effect on the structure either. The other carbonyl group, in the carboxymethyl moiety, has a fixed orientation, despite the rotational freedom about the $\chi 1$, $\chi 2$ and $\chi 3$ torsion angles (Figure 1 and Supplementary Table S1). The orientation of the CO groups toward the C-terminus confers polarity on the whole PNA chain. In addition, the carbonyl oxygen atoms are hydrogen bonded with aromatic CH groups of neighboring nucleobases (C8H8 for guanine and adenine rings, C6H6 for cytosine and thymine). The donor–acceptor distances of 3.2 Å (hydrogen–acceptor 2.3 Å) attest to relatively strong interactions for an H-bond involving a C atom. Polarization of the CH bonds could be facilitated by the aromatic environment of the bases, making carbon a better donor.

In addition to Watson–Crick base pairing, crystal structures reveal interesting PNA–RNA interactions between the backbones of neighboring duplexes. Electronegative groups (O^- from phosphate or hydroxyl O from ribose) approach the carbonyl C atoms of the peptide bond. There is no evidence that this interaction destabilizes the peptide bond but the geometry is similar to that found during the nucleophilic attack of a serine protease on a peptide bond. Such an arrangement indicates at least that the C-atom is electron-depleted.

Finally, PNA's role has been proposed as a primordial carrier of genetic information (42). Two conditions were set to support this speculation: that components of PNA be found in primitive microorganisms and that sequence-specific PNA–PNA and PNA–RNA interactions be demonstrated. The first condition has been met when N-(2-aminoethyl)glycine was found in cyanobacteria (43), while examples of meaningful interactions between the oligomers are presented in this paper. The sensitive element of PNA is the carbonyl carbon of the peptide bond and the catalytically active group of RNA is the 2′ hydroxyl, and we see them interacting in the crystal structure.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Science Centre (Poland) [UMO-2011/01/B/NZ1/04429]; Ministry of Science and Higher Education (Poland) [0450/IP1/2013/72, 01/KNOW2/2014]; European Commission (the Seventh Framework Programme); BioStruct-X project [Contract No. 283570]. Funding for open access charge: Institute of Bioorganic Chemistry, Polish Academy of Sciences.
Conflict of interest statement. None declared.

REFERENCES

1. Orr,H.T. and Zoghbi,H.Y. (2007) Trinucleotide repeat disorders. *Annu. Rev. Neurosci.*, **30**, 575–621.
2. Nelson,D.L., Orr,H.T. and Warren,S.T. (2013) The unstable repeats–three evolving faces of neurological disease. *Neuron*, **77**, 825–843.
3. Mirkin,S.M. (2007) Expandable DNA repeats and human disease. *Nature*, **447**, 932–940.
4. Brook,J.D., McCurrach,M.E., Harley,H.G., Buckler,A.J., Church,D., Aburatani,H., Hunter,K., Stanton,V.P., Thirion,J.P., Hudson,T. *et al.* (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell*, **68**, 799–808.
5. Ranum,L.P. and Cooper,T.A. (2006) RNA-mediated neuromuscular disorders. *Annu. Rev. Neurosci.*, **29**, 259–277.
6. Hsu,R.J., Hsiao,K.M., Lin,M.J., Li,C.Y., Wang,L.C., Chen,L.K. and Pan,H.C. (2011) Long Tract of Untranslated CAG Repeats Is Deleterious in Transgenic Mice. *PLoS One*, **6**.
7. Li,L.B., Yu,Z.M., Teng,X.Y. and Bonini,N.M. (2008) RNA toxicity is a component of ataxin-3 degeneration in Drosophila. *Nature*, **453**, 1107–U1109.
8. Mykowska,A., Sobczak,K., Wojciechowska,M., Kozlowski,P. and Krzyzosiak,W.J. (2011) CAG repeats mimic CUG repeats in the misregulation of alternative splicing. *Nucleic Acids Res.*, **39**, 8938–8951.

9. Shieh,S.Y. and Bonini,N.M. (2011) Genes and pathways affected by CAG-repeat RNA-based toxicity in Drosophila. *Hum. Mol. Genet.*, **20**, 4810–4821.
10. Sathasivam,K., Neueder,A., Gipson,T.A., Landles,C., Benjamin,A.C., Bondulich,M.K., Smith,D.L., Faull,R.L., Roos,R.A., Howland,D. *et al.* (2013) Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 2366–2370.
11. Tsoi,H., Lau,C.K., Lau,K.F. and Chan,H.Y. (2011) Perturbation of U2AF65/NXF1-mediated RNA nuclear export enhances RNA toxicity in polyQ diseases. *Hum. Mol. Genet.*, **20**, 3787–3797.
12. Tsoi,H., Lau,T.C., Tsang,S.Y., Lau,K.F. and Chan,H.Y. (2012) CAG expansion induces nucleolar stress in polyglutamine diseases. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 13428–13433.
13. Fiszer,A. and Krzyzosiak,W.J. (2014) Oligonucleotide-based strategies to combat polyglutamine diseases. *Nucleic Acids Res.*, **42**, 6787–6810.
14. Gao,Z. and Cooper,T.A. (2013) Antisense oligonucleotides: rising stars in eliminating RNA toxicity in myotonic dystrophy. *Hum. Gene Ther.*, **24**, 499–507.
15. Krzyzosiak,W.J., Sobczak,K., Wojciechowska,M., Fiszer,A., Mykowska,A. and Kozlowski,P. (2012) Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. *Nucleic Acids Res.*, **40**, 11–26.
16. Gambari,R. (2014) Peptide nucleic acids: a review on recent patents and technology transfer. *Expert Opin. Ther. Patents*, **24**, 267–294.
17. Egholm,M., Buchardt,O., Christensen,L., Behrens,C., Freier,S.M., Driver,D.A., Berg,R.H., Kim,S.K., Norden,B. and Nielsen,P.E. (1993) PNA hybridizes to complementary oligonucleotides obeying the Watson–Crick hydrogen-bonding rules. *Nature*, **365**, 566–568.
18. Igloi,G.L. (1998) Variability in the stability of DNA-peptide nucleic acid (PNA) single-base mismatched duplexes: real-time hybridization during affinity electrophoresis in PNA-containing gels. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 8562–8567.
19. Jensen,K.K., Orum,H., Nielsen,P.E. and Norden,B. (1997) Kinetics for hybridization of peptide nucleic acids (PNA) with DNA and RNA studied with the BIAcore technique. *Biochemistry*, **36**, 5072–5077.
20. Nielsen,P.E. and Egholm,M. (1999) An introduction to peptide nucleic acid. *Curr. Issues Mol. Biol.*, **1**, 89–104.
21. Schwarz,F.P., Robinson,S. and Butler,J.M. (1999) Thermodynamic comparison of PNA/DNA and DNA/DNA hybridization reactions at ambient temperature. *Nucleic Acids Res.*, **27**, 4792–4800.
22. Demidov,V.V., Potaman,V.N., Frank-Kamenetskii,M.D., Egholm,M., Buchard,O., Sonnichsen,S.H. and Nielsen,P.E. (1994) Stability of peptide nucleic acids in human serum and cellular extracts. *Biochem. Pharmacol.*, **48**, 1310–1313.
23. Koppelhus,U. and Nielsen,P.E. (2003) Cellular delivery of peptide nucleic acid (PNA). *Adv. Drug Deliv. Rev.*, **55**, 267–280.
24. Hu,J., Matsui,M. and Corey,D.R. (2009) Allele-selective inhibition of mutant huntingtin by peptide nucleic acid-peptide conjugates, locked nucleic acid, and small interfering RNA. *Ann. N. Y. Acad. Sci.*, **1175**, 24–31.
25. Hu,J., Matsui,M., Gagnon,K.T., Schwartz,J.C., Gabillet,S., Arar,K., Wu,J., Bezprozvanny,I. and Corey,D.R. (2009) Allele-specific silencing of mutant huntingtin and ataxin-3 genes by targeting expanded CAG repeats in mRNAs. *Nat. Biotechnol.*, **27**, 478–484.
26. Kiliszek,A. and Rypniewski,W. (2014) Structural studies of CNG repeats. *Nucleic Acids Res.*, **42**, 8189–8199.
27. Kabsch,W. (2010) Xds. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.
28. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–325.
29. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
30. Langer,G.G., Hazledine,S., Wiegels,T., Carolan,C. and Lamzin,V.S. (2013) Visual automated macromolecular model building. *Acta Crystallogr. D Biol. Crystallogr.*, **69**, 635–641.
31. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
32. Murshudov,G.N., Skubak,P., Lebedev,A.A., Pannu,N.S., Steiner,R.A., Nicholls,R.A., Winn,M.D., Long,F. and Vagin,A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 355–367.
33. Evans,P.R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 282–292.
34. Sheldrick,G.M. (2008) A short history of SHELX. *Acta Crystallogr. A*, **64**, 112–122.
35. Rasmussen,H., Liljefors,T., Petersson,B., Nielsen,P.E., Liljefors,T. and Kastrup,J.S. (2004) The influence of a chiral amino acid on the helical handedness of PNA in solution and in crystals. *J. Biomol. Struct. Dyn.*, **21**, 495–502.
36. Lu,X.J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
37. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
38. DeLano,W.L. (2002) *The PyMOL Molecular Graphics System*. DeLano Scientific, Palo Alto, CA, USA.
39. Leontis,N.B. and Westhof,E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.
40. Saenger,W. (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag, New York.
41. Kiliszek,A., Kierzek,R., Krzyzosiak,W.J. and Rypniewski,W. (2009) Structural insights into CUG repeats containing the 'stretched U-U wobble': implications for myotonic dystrophy. *Nucleic Acids Res.*, **37**, 4149–4156.
42. Nielsen,P.E. (1993) Peptide nucleic acid (PNA): a model structure for the primordial genetic material? *Orig. Life Evol. Biosph.*, **23**, 323–327.
43. Banack,S.A., Metcalf,J.S., Jiang,L., Craighead,D., Ilag,L.L. and Cox,P.A. (2012) Cyanobacteria produce N-(2-aminoethyl)glycine, a backbone for peptide nucleic acids which may have been the first genetic molecules for life on Earth. *PLoS One*, **7**, e49043.
44. Rasmussen,H., Kastrup,J.S., Nielsen,J.N., Nielsen,J.M. and Nielsen,P.E. (1997) Crystal structure of a peptide nucleic acid (PNA) duplex at 1.7 A resolution. *Nat. Struct. Biol.*, **4**, 98–101.
45. He,W., Hatcher,E., Balaeff,A., Beratan,D.N., Gil,R.R., Madrid,M. and Achim,C. (2008) Solution structure of a peptide nucleic acid duplex from NMR data: features and limitations. *J. Am. Chem. Soc.*, **130**, 13264–13273.
46. Yeh,J.I., Pohl,E., Truan,D., He,W., Sheldrick,G.M., Du,S. and Achim,C. (2010) The crystal structure of non-modified and bipyridine-modified PNA duplexes. *Chemistry*, **16**, 11867–11875.
47. Haaima,G., Rasmussen,H., Schmidt,G., Jensen,D.K., Kastrup,J.S., Stafshede,P.W., Norden,B. and Nielsen,P.E. (1999) Peptide nucleic acids (PNA) derived from N-(N-methylaminoethyl)glycine. Synthesis, hybridization and structural properties. *N. J. Chem.*, **23**, 833–840.
48. Eldrup,Anne B., Nielsen,Bettina B., Haaima,G., Rasmussen,H., Kastrup,Jette S., Christensen,C. and Nielsen,Peter E. (2001) 1,8-Naphthyridin-2(1H)-ones − Novel Bicyclic and Tricyclic Analogues of Thymine in Peptide Nucleic Acids (PNAs). *Eur. J. Org. Chem.*, **2001**, 1781–1790.