CMLS, Cell. Mol. Life Sci. 60 (2003) 862–870
1420-682X/03/050862-09
DOI 10.1007/s00018-003-3046-3


CMLS Cellular and Molecular Life Sciences

# Do natural DNA triple-helical structures occur and function in vivo?

*Dedicated to the memory of Professor Claude Hélène*

**R. Zain [a,*] and J-S Sun [b]**

[a] Department of Molecular Biology and Functional Genomics, Stockholm University, 10691 Stockholm (Sweden), Fax: +46 8 166488, e-mail: rula.zain@molbio.su.se
[b] USM0503 'Régulations et Dynamique des Génomes', Muséum National d'Histoire Naturelle, UMR8646 CNRS-MNHN, U565 INSERM, 43 rue Cuvier, 75231 Paris Cedex 05 (France)

**Abstract.** Formation of natural intramolecular triple-helical structures of DNA is still an intriguing research topic in view of the possible involvement of these structures in biological processes. The biochemical and biophysical properties of DNA triplex structures have been extensively studied, and experimental data show that H-DNA is likely to form in vivo and may regulate the expression of various genes. However, direct and unambiguous evidence of the possible biological roles of these structures is yet elusive. This review focuses on the basic facts that are in favor of, or against, the hypothesis of the presence and function of natural DNA triple-helical structures in vivo, and outlines the different methods and probes that have been used to support these facts.

**Key words.** Intramolecular structure; DNA; triple helix; H-DNA; oligopyrimidine•oligopurine.

## Introduction

Nucleic acids are polymorphic macromolecules and can adopt a variety of single-stranded, duplex and multi-stranded conformations, which in turn may provide important signals for the control of gene expression. Triple-helical structures of nucleic acids, DNA as well as RNA, can be formed inter- or intramolecularly. Although the first triple helix was found in polyribonucleotides [1], and despite the fact that RNA triple helices constitute a very challenging topic, this review will mainly deal with intramolecular DNA triple helices. Triple-helical structures of DNA are based on sequence-specific recognition of oligopyrimidine•oligopurine sequences of double-helical DNA, by a third strand oligonucleotide. This oligonucleotide winds around the double helix in its major groove and forms base triplets with the Watson-Crick base pairs, through Hoogsteen or reverse Hoogsteen hydrogen bonding with the oligopurine bases (figs. 1, 2).

The sequence-specific DNA recognition that leads to the formation of an oligonucleotide-directed triple helix [2, 3] has been largely exploited in order to down-regulate or up-regulate transcription of genes, induce directed mutagenesis, promote homologous recombination or to direct modification of genomic DNA at selected gene loci (see [4] and the references therein). Triple helix-forming oligonucleotides are therefore powerful gene-specific tools that can be employed in a wide range of applications in experimental biology and gene-based biotechnology and therapeutics.

Intramolecular triple helices, generally named H-DNA, were first found in plasmid DNA at the mirror repeats of oligopyrimidine•oligopurine sequences while subjected to physical constraints and/or low pH, (see [5] and the references therein). The plasmids underwent conformational rearrangement involving the disruption of half of the symmetry-related double helix and the folding back of one of the resulting single strands in order to form an intramolecular triple-helical structure (fig. 3). It is believed that the constraint in natural DNA, referred to as negative supercoiling, is the driving force of these rearrangements.

* Corresponding author.

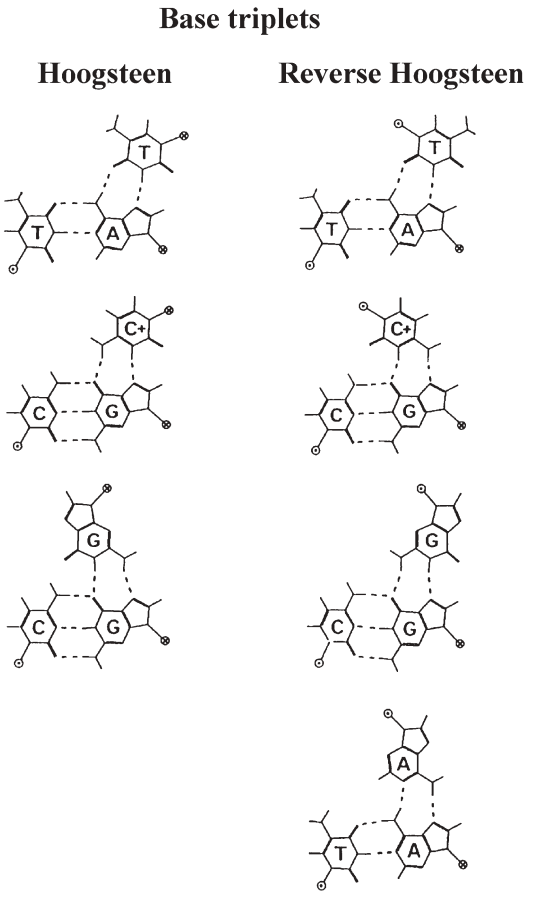


Figure 1. Base triplets formed by natural bases through either Hoogsteen (left column) or reverse Hoogsteen hydrogen bonds (right column).

Since DNA undergoes a characteristic conformational change upon H-DNA formation and forms a sharply bent double helix, which is often observed when DNA-regulatory proteins bind to DNA, it is tempting to postulate that H-DNA could act as a molecular switch to modulate gene expression in a structure-dependent manner in vivo, and that cellular proteins could specifically recognize triple-helical DNA structures and stabilize them. This speculation is supported by two facts: (i) mirror repeats of oligopyrimidine•oligopurine sequences have been found in several eukaryotic genomes, and were located in particular near regulating regions and recombination hot spots; (ii) DNA topological constraints, especially negative supercoiling, are an intrinsic and dynamic feature during the processing of DNA information such as replication, transcription and so on.

Despite a large number of biophysical and biochemical studies on the formation of H-DNA in vitro and in living cells during the last two decades, there is still a 'mystery' around the biological roles played by these rather peculiar DNA structures in vivo. We here focus on the results that have been achieved in the search of DNA triple-helical structures in vivo and their possible biological roles.

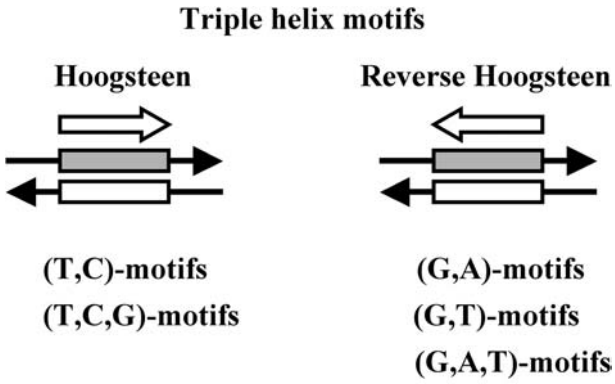


Figure 2. Classification of triple-helix motifs according to the use of base triplets in either Hoogsteen or reverse Hoogsteen configuration. Oligopyrimidine and oligopurine sequences are shown as white and gray boxes, respectively, and the third strand oligonucleotide as a white arrow. The 5′ to 3′ direction of each strand is shown by the sense of arrows. The third strand is parallel to the oligopurine strand in all Hoogsteen motifs, whereas it is antiparallel to the oligopurine strand in all reverse Hoogsteen motifs.

## H-DNA and *H-DNA

The structures of intramolecular DNA triplexes are formed at oligopyrimidine•oligopurine sequences under topological constraints. They can be divided into several different categories depending on the triplex structural motifs (fig. 2).

### H-DNA

Hypersensitivity toward S1 nuclease was earlier observed at the promoter region of some eukaryotic genes that were associated with oligopyrimidine•oligopurine sequences (see [6] and the references therein). Lyamichev et al. proposed a new DNA structure based on the observed hypersensitivity of supercoiled plasmids carrying oligopyrimidine•oligopurine sequences toward single-strand nuclease under acidic conditions, and on the conformational transition of DNA topoisomers unraveled by two-dimensional (2D) gel electrophoresis [7]. This new structure was called H-structure of DNA due to the observed requirement of protonation of the cytosines in the third strand, and consisted of one oligopyrimidine strand and one-half of an oligopurine strand, while the other half was left single stranded (fig. 3). Two isomers of this intramolecular structure can be formed where either the 5′ or the 3′ part of the oligopurine strand is single stranded. In this model, two different base triplets are then formed, T·AxT and C·GxC+, where thymines and cytosines form Hoogsteen hydrogen bonds with adenines and guanines, respectively (fig. 1). Furthermore, it was concluded that in order to form a stable intramolecular structure containing these base triplets, the oligopyrimidine•oligopurine sequences had to be organized as repeats with mirror symmetry [8]. Both homologous and

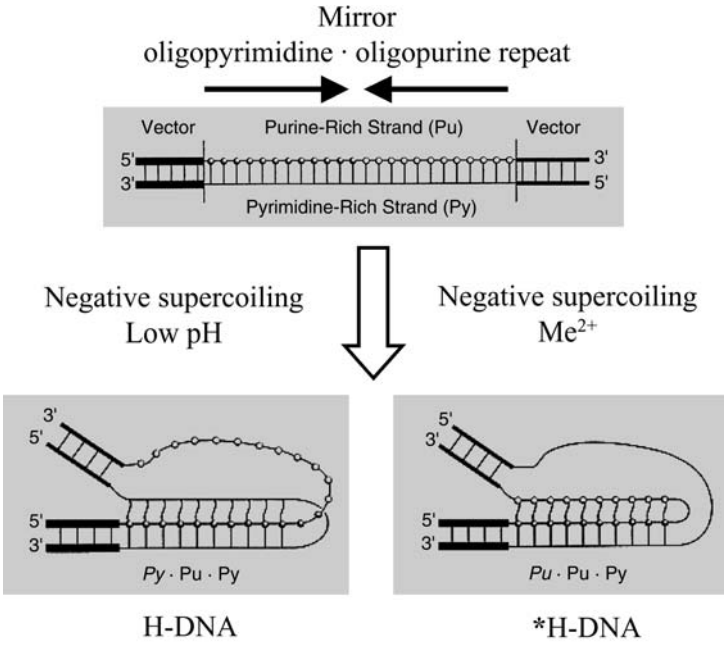


Figure 3. Schematic description of the formation of an intramolecular triple helix. Under topological constraints, half of the duplex is disrupted into oligopyrimidine and oligopurine single strands. One of them subsequently folds back to form an intramolecular triplex with half of the duplex, while the other half is left single stranded. As the central part of duplex acts as a linker, it can be any sequence of 3–10 bp. Different motifs, H- or *H-DNA, can be formed depending on pH and divalent cations.

mixed oligopyrimidine•oligopurine mirror repeats were found to form H-DNA structures, and hence these sequences were called H-palindromes.

***H-DNA**

*H-DNA, which was initially called H′-DNA, is a variant of H-DNA structure [9]. It is an intramolecular triple-helical structure of DNA and was first discovered when stretches of $d(C)_n \cdot d(G)_n$ in supercoiled plasmid DNA showed single-strand hyperactivity toward a single-strand specific reagent, namely chloroacetaldehyde. This structure was also confirmed by 2D gel electrophoresis at neutral pH in the presence of magnesium ions and was formed by the entire oligopurine strand and half of the oligopyrimidine strand leading to C·GxG base triplets (fig. 3) [10]. Subsequently, other sequences such as $d(C\text{-}T)_n \cdot d(G\text{-}A)_n$ proved to equally adopt an *H-DNA structure under negative supercoiling stress, consisting of C·GxG and T·AxA base triplets in reverse Hoogsteen configuration [11, 12]. *H-DNA, in contrast to H-DNA, could be formed in sequences that were not perfect oligopyrimidine•oligopurine mirror repeats due to the degeneracy of T·AxA and T·AxT base triplets in the same structure [13]. The formation of these intramolecular structures depends on the sequence and the relative stability of each isomer. Finally, it was also made clear that divalent cations, such as $Mg^{2+}$ and $Zn^{2+}$, were necessary for the formation of stable *H-DNA and that depending on the composition of *H-DNA sequences, different divalent cations were required [5].

A direct imaging of H-DNA has been recently reported using atomic force microscopy [14]. Formation of H-DNA appeared in atomic force microscopy images as a clear protrusion with a different thickness than that of the DNA duplex. It resulted in a kink in the double-helix path consistent with the existing models. It is interesting to note that the kink formed an acute angle so that the flanking DNA regions were brought in close proximity. The mobility of flanking DNA arms is limited compared with that for cruciforms and three-way junctions. These structural properties of H-DNA may be important for promoter-enhancer interactions and other DNA transactions.

A different class of intramolecular triple-helical structures includes those arising from the folding of consecutive blocks of nucleotides along a single strand of DNA [15–20]. The structures formed were referred to as intrastrand triplexes, and the formation of the base triplets followed the same order as in the earlier mentioned structures, H- and *H-DNA. Unlike H- and *H-DNA, intrastrand triplex structures require three consecutive oligopyrimidine•oligopurine sequence domains of DNA with appropriate symmetry, and demand a higher level of negative supercoiling stress due to the disruption of three oligopyrimidine•oligopurine domains, instead of only one.

## Potential intramolecular triplex-forming sequences in genomes

Two bioinformatic analyses of prokaryotic genomic DNA as well as a set of eukaryotic genes have been reported. In the first, an overabundance of long (15–30 purines) contiguous oligopurine tracts was found in all eukaryotic sequences included in the study, starting from *Saccharomyces cerevisiae* and *Caenorhabditis elegans* through *Drosophila*, chicken, mouse and finally to human [21]. Whereas long oligopurine tracts were statistically overrepresented in all of these eukaryotic genomes, very long oligopurine stretches (>30 contiguous purines) occurred much more frequently in mouse and human than in the other eukaryotes. When prokaryotic genomic sequences of *E. coli* and *Bacillus subtilis* were analyzed, none of these trends were observed.

In the second analysis, 157 human genes totaling 1,086,110 bp, the complete yeast (*S. cerevisiae*) chromosome III, which contains 182 open reading frames (ORFs) (315,357 bp), and an *E. coli* sequence containing 350 ORFs (324,146 bp) were analyzed [22]. The occurrence of mirror repeats of DNA was compared in the three different organisms, and it was concluded that while both human and yeast genomes were highly enriched in the total number of mirror repeat sequences and in particular longer repeats (>12 bp), mirror repeats occurred only with random frequency in *E. coli*.

A recent and more detailed analysis of the complete yeast (*S. cerevisiae*) genome (12,294,083 bp, 6608 ORFs) of the occurrence of H- or *H-DNA, including all possible triplex motifs, with the length of mirror repeat >10 bp has been carried out [D. Polverari et al., unpublished]. An empirical rule based on the physical chemistry data of intermolecular triple helix formation was applied to eliminate the intramolecular triplexes that are unlikely to be formed due to weak stability. In total, 148 hits were found and classified into four categories according to gene organization: intron, exon, promoter region (defined as the region of 500 bp upstream of the first exon) and other regions. The highest occurrence of potential H- or *H-DNA sites was found in the promoter region (71 hits), followed by the exons (57 hits), whereas the intronic and other regions had 2 and 18 hits, respectively. In terms of frequency, the potential H- or *H-DNA sites in the promoter region (1/46,535) were estimated to be about 1.8-fold higher than the average occurrence of H-DNA in the yeast genome (1/83,068) and 3.3-fold higher than those in exons (1/152,903). This is in line with the early observation of hypersensitivity toward S1 nuclease at the promoter region of some eukaryotic genes that were associated with oligopyrimidine•oligopurine sequences.

Although the analyses presented above argue for the potential presence of H-like intramolecular triple helices in eukaryotic but not prokaryotic genomes, other types of intramolecular triplex structures might still be found in prokaryotes. One such example is the intrastrand triple-helical structure that was mentioned in the previous section. When the databases of several bacterial genomes were searched, potential intrastrand triplex (PIT) elements were found in *E. coli*, *Synechocystis* sp. and *Haemophilus influenza* [20]. Remarkably, PIT elements were detected as multiple copies of a particular class or sequence, and in *E. coli* up to 25 copies of a certain purine-motif PIT element were identified.

## Triplex-binding proteins

Sequence analyses of eukaryotic genomes have lent credence to the potential existence of intramolecular triplex structures of DNA in living cells. Yet it was very difficult to detect H-DNA (and *H-DNA) in vivo due to the lack of probes that would have access to DNA in the nucleus, bind and act specifically on the triple-helical structures potentially formed in living cells.

Identification of several proteins that interact with triple-helical DNA supports the argument that triplex DNA can be formed and stabilized upon binding of triplex-specific and single strand-specific proteins in vivo, and therefore might ultimately have biological implications. Both inter- and intramolecular triplexes with different sequences have been used to select triplex-binding protein(s) from HeLa cell nuclear extract [23, 24]. Using gel mobility shift assays, 2D gel electrophoresis and mass spectrometry, several proteins that showed affinity towards triplex structures were identified. Nevertheless, the authors concluded that the identified triplex-binding proteins were not entirely specific to triplexes but showed some affinity for other types of nucleic acid structures [25]. Moreover, the biological relevance of these proteins, besides their ability to bind triple-helical structures, has not yet been fully defined.

Nelson et al. identified an additional triplex-binding protein from *S. cerevisiae* [26]. This protein was purified from whole-cell yeast extract using affinity chromatography and a psoralen-stabilized intermolecular triple-helix DNA. Electrophoretic mobility shift assay (EMSA) confirmed that the protein had a high binding preference toward purine-motif triplexes, and sequence analyses revealed it as being a product of the *STM1* gene. The *STM1* gene has been identified as a multicopy suppressor of mutations in several genes involved in mitosis. In order to identify additional genes that encode triplex-binding proteins, Musso et al. screened an *S. cerevisiae* genomic library using a psoralen cross-linked triplex probe and southwestern methods [27]. These experiments indicated that a second gene, *CDP1*, also encoded a triplex-binding protein. However, the screening identified only a small

portion of this protein, and no other yeast genes were found using this method, including the *STM1*.

A triplex-binding protein has also been identified in *Drosophila*. Jimenez-Garcia et al. demonstrated that the GAGA factor of *Drosophila*, which is a sequence-specific DNA-binding protein, could bind triple-helical structures of DNA in vitro using EMSA and footprinting by DNase I and dimethyl sulfate (DMS) [28]. GAGA showed a similar affinity and specificity to intermolecular triplexes as to canonical double-stranded DNA. GAGA-binding sites are found at the promoter of several genes, and they consist of repeated $d(GA/TC)_n$ which could potentially form triple-helical structures. Since the GAGA factor participates in the regulation of expression of several genes in *Drosophila*, it was suggested that its interaction with triplex DNA could play a role at that level. However, additional evidence demonstrating the presence of a GAGA-triplex complex in vivo is still necessary.

Rao and Craig, on the other hand, reported that TnsC, a protein encoded by the bacterial transposon Tn7, was able to recognize and bind inter- as well as intramolecular triple-helix structures formed in plasmid DNA in vitro [29, 30]. Interestingly, TnsC has been shown to control the target site of the transposon, leading it to preferentially insert adjacent to the formed inter- or intramolecular pyrimidine-motif triplex structure. Although these experiments outlined, in a remarkable way, the interaction of the protein TnsC with triple-helical structures of DNA, as well as the consequence that it has on targeting the transposition event, it remains to be shown how these findings could provide evidence of the formation of triplexes in vivo and even more their potential involvement in different biological processes.

## Monitoring the formation of intramolecular triplex structures in vivo and their biological effects

A more direct in vivo probing of the formation of triple-helical structures was achieved by immunodetection using triplex antibodies. Agazie et al. prepared triplex-specific monoclonal antibodies, Jel 318 and Jel 466, by immunizing mice with a triplex-forming sequence of DNA [31, 32]. Binding of the two antibodies to chromosomes and cell nuclei was demonstrated by immunofluorescence, and it was inhibited by the addition of competing triplex DNA [33]. The effect of the two triplex-specific antibodies on total replication and transcription in isolated intact nuclei was measured and estimated to ~20% inhibition. Furthermore, cell proliferation was suppressed when the antibodies were incorporated into cultured mammalian cells. More recently, immunodetection by triplex-specific antibodies was combined with fluorescence in situ non-denaturating hybridization (N-FISH) in order to detect triplex-forming DNA in nuclei in fixed cells under mild conditions [34]. Employment of single-strand DNA fluorescent probes that could be complementary to the single-stranded region of intramolecular triplexes led to the detection of sequence-dependent foci-type signals that overlapped with, or were closely related to, triplexes that were immunolocalized by the triplex-specific antibodies. These results demonstrated that triple-helical structures could form in vivo, and several models of high-order structures of chromatin in living cells have been suggested, where intramolecular as well as transmolecular triplex structures could take part.

DNA sequences with the potential to form intramolecular triplex structures are overrepresented in the genome of several eukaryotes. However, for experimental convenience, the first studies of the formation of triplex DNA in living cells were performed in *E. coli*. DNA plasmids that contained potential triplex-forming inserts were transformed to bacterial cells, and different chemical probes such as osmium tetraoxide, trimethylpsoralene and chloroacetaldehyde were used to investigate the formation of triplex structures [35–37]. Both H- and *H-DNA were shown to exist in bacterial cells, although the experimental conditions were not strictly physiological.

*E. coli* was furthermore used to monitor the effect of potential triplex sequences of DNA on biological processes such as transcription, replication and recombination. Plasmids carrying triplex-forming stretches of DNA and reporter genes were transformed to bacterial cells, and the expression of the reporter gene was analyzed. When an artificial oligopyrimidine•oligopurine repeat capable of forming an intramolecular triplex under physiological conditions was inserted within the coding region of the *lacZ* gene, substantial reduction of gene expression was observed in *E. coli* [38, 39]. On the other hand, insertion of an intramolecular triplex-forming sequence upstream of the $\beta$-lactamase promoter led to an increase of its expression activity [40], while insertion between the $\beta$-lactamase promoter and the coding sequence led to a strong down-regulation of transcription of the $\beta$-lactamase gene [41]. This latest effect was achieved mainly when cells were allowed to grow in the presence of a triplex-stabilizing compound, namely a benzopyridoindole derivative, indicating the involvement of a triplex structure in the modulation of expression. Similar experiments were carried out in mouse and yeast cell lines harboring plasmids with triplex-forming sequences [42, 43].

The formation of intramolecular triple-helix structures of DNA has been shown to cause arrests of DNA polymerase in vitro and was hence suspected to disrupt DNA replication in vivo [44]. Single-stranded DNA fragments or negatively supercoiled plasmids were used in order to examine the impact of DNA triplex structures on the arrest of DNA polymerization [45, 46]. When naturally occurring oligopyrimidine•oligopurine mirror repeats of

varying length were inserted in supercoiled plasmid DNA, a length-dependent DNA polymerization arrest was detected. Since the stability of a triplex structure is dependent on, among other factors, the length of its sequence, these results suggest the involvement of an intramolecular triplex in the interruption of polymerization, although direct evidence of this type of mechanism has been difficult to provide.

Most of the naturally occurring mirror-repeat sequences used in the previously mentioned study were polypyrimidine•polypurine tracts localized in the intron 21 of the *PKD1* gene. Intron 21 contains a 2500-bp polypyrimidine•polypurine sequence where 23 mirror repeats have been found [47]. Instability of the *PKD1* gene was associated with autosomal dominant polycystic kidney disease (ADPKD). 2D-gel electrophoresis analyses of a DNA plasmid containing the entire *PKD1* intron 21 revealed that it underwent structural transitions under acidic conditions and negative supercoiling [48]. In addition, P1 nuclease mapping of the same plasmid identified four single-stranded regions under the same conditions where the structural transitions were observed, negative supercoiling and low pH. A recent examination on the role of DNA topology on the stability of the *PKD1* gene was carried out using different *E. coli* strains with mutations in the nucleotide excision repair, the topoisomerase I and/or the gyrase gene. It was shown that the polypurine•polypyrimidine sequences present in the intron 21 of this gene were able to adopt secondary DNA structures under negative supercoiling conditions in *E. coli*. In addition, these structures proved to interact with several cellular factors including the nucleotide excision repair, which could explain the instability of the *PKD1* gene and its tendency to undergo mutations [49].

A further example of the possible biological role of triple-helical structures of DNA has emerged from studies concerning the regulation of the *frataxin* gene. Expansion of the trinucleotide (GAA/TCC) repeats present in the first intron of the *frataxin* gene has been associated with Friedreich's ataxia (FREDA) disease [50, 51] and the reduction of transcription of this gene [52–54]. The oligopyrimidine•oligopurine character of these expanded repeat sequences (>100 triplets) provides them with the potential to form a triple-helical structure. GAA/TCC repeats of various lengths were cloned in the intron of a reporter gene and transcription was measured in plasmid transfected cell lines [52]. Expanded repeats were found to inhibit transcription of the reporter gene and a length-dependent inhibition of plasmid replication was also observed. Factors inhibiting the formation of a stable triplex structure during transcription, such as the presence of a complementary oligonucleotide, or a non-acidic medium, proved to increase the yield of full-length transcript in vitro [55]. An intramolecular-triplex structure formed by expanded triplet repeats could clearly trigger the inhibition of transcription of the *frataxin* gene though this effect has not yet been confirmed in vivo.

Oligopyrimidine•oligopurine sequences are present in the vicinity of the promoter of several eukaryotic genes such as *MUC1* [56], *C1 inhibitor* [57], chicken malic enzyme [58–60], human γ-globin [61, 62], human c-*ets*-2 [63], mouse vascular smooth muscle α-actin [64, 65], mouse c-ki-*ras* [66–68], *Hmga2* gene [69] and several others [5, 70, 71]. Functional analysis of these sequences revealed in most cases that they were essential for transcriptional activity and highly capable of forming intramolecular triplex structures in vitro. However, contradictory results concerning the impact of an intramolecular triplex structure in the regulation of transcription were obtained. For instance, when activity of the *C1 inhibitor* gene promoter was examined using mutation analysis, protein binding and transient transfection assay, it was suggested that formation of an H-DNA structure was not involved in basal promoter activity under the employed experimental conditions [72, 73]. Another related study was performed on the expression of the glycoprotein MUC1 and demonstrated that some of the oligopyrimidine•oligopurine mirror-repeat elements located at the promoter of this gene were single-strand sensitive and able to form intramolecular triplex structures in vitro. Nevertheless, deletion and mutation experiments of this sequence in transient transfected cell lines did not support the involvement of a DNA secondary structure in regulating the transcription of the *MUC1* gene. Interestingly, a recent DNase I mapping in chromatin from *MUC1* transgenic mice and human cell lines identified a new DNase I hypersensitive site, very close to one of the earlier defined oligopyrimidine•oligopurine mirror repeat elements [74]. Since this latter element has not been investigated yet, it is especially valuable to examine whether it will form a triplex structure, under negative supercoiling conditions similar to those occurring upstream in chromatin during transcription.

Recent studies on the regulation mechanism of nuclear factor *Hmga2*, involved in development and tumorigenesis, resulted in the identification of a 60-bp oligopyrimidine•oligopurine mirror repeat sequence in the promoter region. Using supercoiled plasmid DNA and S1 nuclease sensitivity experiments, it was shown that this element could adopt a single-strand conformation. This element had a homology with a region present in several growth-related genes and showed an affinity toward a single-strand polypyrimidine binding protein that binds also to similar elements in the c-Ki-*ras* and c-*myc* P1 promoters [75]. Since this region of oligopyrimidine•oligopurine also contained a binding site for transcription factor Sp1, several mutated plasmids carrying either a single- or double-strand protein-binding site were used in order to establish the regulating mechanism of this promoter. Transfection assays in eukaryotic cell lines indicated that a sec-

ondary DNA structure, which included a single-strand sequence and was formed in the oligopyrimidine•oligopurine region, contributed to a positive transcriptional activity. Based on its ability to form a triplex DNA structure in vitro and the impact of the single-strand sensitivity site and the triplex-forming sequence on transcriptional activity, it is tempting to suggest that a triplex-directed mechanism could be involved in the regulation of transcription exerted by this and maybe other related promoters.

All the above-described studies that examined the role of triplex structures in gene regulation have independently indicated the possible participation of triple-helical structures as regulatory elements, either as enhancers or repressors. Nevertheless, the establishment of a detailed mechanism of an intramolecular triplex-directed regulation of gene expression is still very challenging.

## Perspectives

After more than a decade of investigation, the accumulated data strongly suggest that the intramolecular DNA triplexes may exist in vivo and be involved in a number of DNA information flow processes in a family of genes that accommodate the appropriate oligopyrimidine•oligopurine mirror repeat sequences in their regulatory region. However, it remains to establish unambiguously that these intramolecular DNA triple-helical structures can act as molecular switches to modulate gene expression and other DNA metabolism events in a structure-dependent manner, in addition to the well-established sequence-specific regulation. The fast and still accelerating pace of deciphering the genomic information of a large number of organisms will, together with the development of functional genomics approaches and triplex-specific probes, certainly contribute to further elucidate the formation of DNA triple-helical structures in vivo and their biological relevance.

*Acknowledgement.* The authors thank Mr D. Polverari for communication of the unpublished data on recent sequence analyses of yeast genome, Dr S. Sigurdsson for critical reading of the manuscript and Dr H. Amiri for valuable discussions. Acknowledgment is due to the Swedish Research Council for financial support.

1 Felsenfeld G., Davies D. R. and Rich A. (1957) Formation of a three-stranded polynucleotide molecule. J. Am. Chem. Soc. **79:** 2023–2024
2 Le Doan T., Perrouault L., Praseuth D., Habhoub N., Decout J. L., Thuong N. T. et al. (1987) Sequence-specific recognition, photocrosslinking and cleavage of the DNA double helix by an oligo-[alpha]-thymidylate covalently linked to an azidoproflavine derivative. Nucleic Acids Res. **15:** 7749–7760
3 Moser H. E. and Dervan P. B. (1987) Sequence-specific cleavage of double helical DNA by triple helix formation. Science **238:** 645–650
4 Giovannangeli C. and Helene C. (2000) Triplex-forming molecules for modulation of DNA information processing. Curr. Opin. Mol. Ther. **2:** 288–296
5 Mirkin S. M. and Frank-Kamenetskii M. D. (1994) H-DNA and related structures. Annu. Rev. Biophys. Biomol. Struct. **23:** 541–576
6 Wells R. D., Collier D. A., Hanvey J. C., Shimizu M. and Wohlrab F. (1988) The chemistry and biology of unusual DNA structures adopted by oligopurine.oligopyrimidine sequences. FASEB J. **2:** 2939–2949
7 Lyamichev V. I., Mirkin S. M. and Frank-Kamenetskii M. D. (1985) A pH-dependent structural transition in the homopurine-homopyrimidine tract in superhelical DNA. J. Biomol. Struct. Dyn. **3:** 327–338
8 Mirkin S. M., Lyamichev V. I., Drushlyak K. N., Dobrynin V. N., Filippov S. A. and Frank-Kamenetskii M. D. (1987) DNA H form requires a homopurine-homopyrimidine mirror repeat. Nature **330:** 495–497
9 Kohwi Y. and Kohwi-Shigematsu T. (1988) Magnesium ion-dependent triple-helix structure formed by homopurine-homopyrimidine sequences in supercoiled plasmid DNA. Proc. Natl. Acad. Sci. USA **85:** 3781–3785
10 Panyutin I. G., Kovalsky O. I. and Budowsky E. I. (1989) Magnesium-dependent supercoiling-induced transition in (dG)n.(dC)n stretches and formation of a new G-structure by (dG)n strand. Nucleic Acids Res. **17:** 8257–8271
11 Bernues J., Beltran R., Casasnovas J. M. and Azorin F. (1990) DNA-sequence and metal-ion specificity of the formation of *H-DNA. Nucleic Acids Res. **18:** 4067–4073
12 Bernues J., Beltran R., Casasnovas J. M. and Azorin F. (1989) Structural polymorphism of homopurine-homopyrimidine sequences: the secondary DNA structure adopted by a d(GA.CT)22 sequence in the presence of zinc ions. EMBO J. **8:** 2087–2094
13 Dayn A., Samadashwily G. M. and Mirkin S. M. (1992) Intramolecular DNA triplexes: unusual sequence requirements and influence on DNA polymerization. Proc. Natl. Acad. Sci. USA **89:** 11406–11410
14 Tiner W. J. Sr, Potaman V. N., Sinden R. R. and Lyubchenko Y. L. (2001) The structure of intramolecular triplex DNA: atomic force microscopy study. J. Mol. Biol. **314:** 353–357
15 Sklenar V. and Feigon J. (1990) Formation of a stable triplex from a single DNA strand. Nature **345:** 836–838
16 Haner R. and Dervan P. B. (1990) Single-strand DNA triple-helix formation. Biochemistry **29:** 9761–9765
17 Chen F. M. (1991) Intramolecular triplex formation of the purine.purine.pyrimidine type. Biochemistry **30:** 4472–4479
18 Macaya R. F., Schultze P. and Feigon J. (1992) Sugar conformations in intramolecular DNA triplexes determined by coupling constants obtained by automated simulation of P. COSY cross peaks. J. Am. Chem. Soc. **114:** 781–783
19 Hoyne P. R., Gacy A. M., Mcmurray C. T. and Maher L. J. 3rd (2000) Stabilities of intrastrand pyrimidine motif DNA and RNA triple helices. Nucleic Acids Res. **28**: 770–775
20 Hoyne P. R., Edwards L. M., Viari A. and Maher L. J. (2000) Searching genomes for sequences with the potential to form intrastrand triple helices. J. Mol. Biol. **302:** 797–809
21 Behe M. J. (1995) An overabundance of long oligopurine tracts occurs in the genome of simple and complex eukaryotes. Nucleic Acids Res. **23:** 689–695
22 Schroth G. P. and Ho P. S. (1995) Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA. Nucleic Acids Res. **23:** 1977–1983
23 Guieysse A. L., Praseuth D. and Helene C. (1997) Identification of a triplex DNA-binding protein from human cells. J. Mol. Biol. **267:** 289–298
24 Musso M., Nelson L. D. and Van Dyke M. W. (1998) Characterization of purine-motif triplex DNA-binding proteins in HeLa extracts. Biochemistry **37:** 3086–3095

25 Guillonneau F., Guieysse A. L., Le Caer J. P., Rossier J. and Praseuth D. (2001) Selection and identification of proteins bound to DNA triple-helical structures by combination of 2D-electrophoresis and MALDI-TOF mass spectrometry. Nucleic Acids Res. **29:** 2427–2436

26 Nelson L. D., Musso M. and Van Dyke M. W. (2000) The yeast STM1 gene encodes a purine motif triple helical DNA-binding protein. J. Biol. Chem. **275:** 5573–5581

27 Musso M., Bianchi-Scarra G. and Van Dyke M. W. (2000) The yeast CDP1 gene encodes a triple-helical DNA-binding protein. Nucleic Acids Res. **28:** 4090–4096

28 Jimenez-Garcia E., Vaquero A., Espinas M. L., Soliva R., Orozco M., Bernues J. et al. (1998) The GAGA factor of *Drosophila* binds triple-stranded DNA. J. Biol. Chem. **273:** 24640–24648

29 Rao J. E. and Craig N. L. (2001) Selective recognition of pyrimidine motif triplexes by a protein encoded by the bacterial transposon Tn7. J. Mol. Biol. **307:** 1161–1170

30 Rao J. E., Miller P. S. and Craig N. L. (2000) Recognition of triple-helical DNA structures by transposon Tn7. Proc. Natl. Acad. Sci. USA **97:** 3936–3941

31 Lee J. S., Burkholder G. D., Latimer L. J., Haug B. L. and Braun R. P. (1987) A monoclonal antibody to triplex DNA binds to eucaryotic chromosomes. Nucleic Acids Res. **15:** 1047–1061

32 Agazie Y. M., Lee J. S. and Burkholder G. D. (1994) Characterization of a new monoclonal antibody to triplex DNA and immunofluorescent staining of mammalian chromosomes. J. Biol. Chem. **269:** 7019–7023

33 Agazie Y. M., Burkholder G. D. and Lee J. S. (1996) Triplex DNA in the nucleus: direct binding of triplex-specific antibodies and their effect on transcription, replication and cell growth. Biochem. J. **316 ( Pt 2):** 461–466

34 Ohno M., Fukagawa T., Lee J. S. and Ikemura T. (2002) Triplex-forming DNAs in the human interphase nucleus visualized in situ by polypurine/polypyrimidine DNA probes and antitriplex antibodies. Chromosoma **111:** 201–213

35 Karlovsky P., Pecinka P., Vojtiskova M., Makaturova E. and Palecek E. (1990) Protonated triplex DNA in *E. coli* cells as detected by chemical probing. FEBS Lett. **274:** 39–42

36 Ussery D. W. and Sinden R. R. (1993) Environmental influences on the in vivo level of intramolecular triplex DNA in *Escherichia coli*. Biochemistry **32:** 6206–6213

37 Kohwi Y., Malkhosyan S. R. and Kohwi-Shigematsu T. (1992) Intramolecular dG.dG.dC triplex detected in *Escherichia coli* cells. J. Mol. Biol. **223:** 817–822

38 Sarkar P. S. and Brahmachari S. K. (1992) Intramolecular triplex potential sequence within a gene down regulates its expression in vivo. Nucleic Acids Res. **20:** 5713–5718

39 Brahmachari S. K., Sarkar P. S., Raghavan S., Narayan M. and Maiti A. K. (1997) Polypurine/polypyrimidine sequences as cis-acting transcriptional regulators. Gene **190:** 17–26

40 Kato M. and Shimizu N. (1992) Effect of the potential triplex DNA region on the in vitro expression of bacterial beta-lactamase gene in superhelical recombinant plasmids. J. Biochem. **112:** 492–494

41 Duval-Valentin G., De Bizemont T., Takasugi M., Mergny J. L., Bisagni E. and Helene C. (1995) Triple-helix specific ligands stabilize H-DNA conformation. J. Mol. Biol. **247:** 847–858

42 Kohwi Y. and Kohwi-Shigematsu T. (1991) Altered gene expression correlates with DNA structure. Genes Dev. **5:** 2547–2554

43 Maiti A. K. and Brahmachari S. K. (2001) Poly purine.pyrimidine sequences upstream of the beta-galactosidase gene affect gene expression in *Saccharomyces cerevisiae*. BMC Mol. Biol. **2:** 11

44 Rao B. S. (1996) Regulation of DNA replication by homopurine/homopyrimidine sequences. Mol. Cell. Biochem. **156:** 163–168

45 Krasilnikov A. S., Panyutin I. G., Samadashwily G. M., Cox R., Lazurkin Y. S. et al. (1997) Mechanisms of triplex-caused polymerization arrest. Nucleic Acids Res. **25:** 1339–1346

46 Potaman V. N. and Bissler J. J. (1999) Overcoming a barrier for DNA polymerization in triplex-forming sequences. Nucleic Acids Res. **27:** e5

47 Van Raay T. J., Burn T. C., Connors T. D., Petry L. R., Germino G. G., Klinger K. W. et al. (1996) A 2.5 kb polypyrimidine tract in the PKD1 gene contains at least 23 H-DNA-forming sequences. Microb. Comp. Genomics **1:** 317–327

48 Blaszak R. T., Potaman V., Sinden R. R. and Bissler J. J. (1999) DNA structural transitions within the PKD1 gene. Nucleic Acids Res. **27:** 2610–2617

49 Bacolla A., Jaworski A., Connors T. D. and Wells R. D. (2001) Pkd1 unusual DNA conformations are recognized by nucleotide excision repair. J. Biol. Chem. **276:** 18597–18604

50 Campuzano V., Montermini L., Molto M. D., Pianese L., Cossee M., Cavalcanti F. et al. (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science **271:** 1423–1427

51 Cossee M., Schmitt M., Campuzano V., Reutenauer L., Moutou C., Mandel J. L. et al. (1997) Evolution of the Friedreich's ataxia trinucleotide repeat expansion: founder effect and premutations. Proc. Natl. Acad. Sci. USA **94:** 7452–7457

52 Ohshima K., Montermini L., Wells R. D. and Pandolfo M. (1998) Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication in vivo. J. Biol. Chem. **273:** 14588–14595

53 Bidichandani S. I., Ashizawa T. and Patel P. I. (1998) The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. Am. J. Hum. Genet. **62:** 111–121

54 Grabczyk E. and Usdin K. (2000) The GAA*TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. Nucleic Acids Res. **28:** 2815–2822

55 Grabczyk E. and Usdin K. (2000) Alleviating transcript insufficiency caused by Friedreich's ataxia triplet repeats. Nucleic Acids Res. **28:** 4930–4937

56 Hollingsworth M. A., Closken C., Harris A., Mcdonald C. D., Pahwa G. S. and Maher L. J. 3rd (1994) A nuclear factor that binds purine-rich, single-stranded oligonucleotides derived from S1-sensitive elements upstream of the CFTR gene and the MUC1 gene. Nucleic Acids Res. **22:** 1138–1146

57 Davis T. L., Firulli A. B. and Kinniburgh A. J. (1989) Ribonucleoprotein and protein factors bind to an H-DNA-forming c-myc DNA element: possible regulators of the c-myc gene. Proc. Natl. Acad. Sci. USA **86:** 9682–9686

58 Xu G. and Goodridge A. G. (1996) Characterization of a polypyrimidine/polypurine tract in the promoter of the gene for chicken malic enzyme. J. Biol. Chem. **271:** 16008–16019

59 Xu G. and Goodridge A. G. (1999) Function of a C-rich sequence in the polypyrimidine/polypurine tract of the promoter of the chicken malic enzyme gene depends on promoter context. Arch. Biochem. Biophys. **363:** 202–212

60 Xu G. and Goodridge A. G. (1998) A CT repeat in the promoter of the chicken malic enzyme gene is essential for function at an alternative transcription start site. Arch. Biochem. Biophys. **358:** 83–91

61 Ulrich M. J., Gray W. J. and Ley T. J. (1992) An intramolecular DNA triplex is disrupted by point mutations associated with hereditary persistence of fetal hemoglobin. J. Biol. Chem. **267:** 18649–18658

62 Bacolla A., Ulrich M. J., Larson J. E., Ley T. J. and Wells R. D. (1995) An intramolecular triplex in the human gamma-globin 5′-flanking region is altered by point mutations associated with hereditary persistence of fetal hemoglobin. J. Biol. Chem. **270:** 24556–24563

63 Mavrothalassitis G. J., Watson D. K. and Papas T. S. (1990) The human ETS-2 gene promoter: molecular dissection and nuclease hypersensitivity. Oncogene **5:** 1337–1342
64 Sun S., Stoflet E. S., Cogan J. G., Strauch A. R. and Getz M. J. (1995) Negative regulation of the vascular smooth muscle alpha-actin gene in fibroblasts and myoblasts: disruption of enhancer function by sequence-specific single-stranded-DNA-binding proteins. Mol. Cell. Biol. **15:** 2429–2436
65 Becker N. A., Kelm R. J. Jr, Vrana J. A., Getz M. J. and Maher L. J. 3rd (2000) Altered sensitivity to single-strand-specific reagents associated with the genomic vascular smooth muscle alpha-actin promoter during myofibroblast differentiation. J. Biol. Chem. **275:** 15384–15391
66 Hoffman E. K., Trusko S. P., Murphy M. and George D. L. (1990) An S1 nuclease-sensitive homopurine/homopyrimidine domain in the c-Ki-ras promoter interacts with a nuclear factor. Proc. Natl. Acad. Sci. USA **87:** 2705–2709
67 Pestov D. G., Dayn A., Siyanova E., George D. L. and Mirkin S. M. (1991) H-DNA and Z-DNA in the mouse c-Ki-ras promoter. Nucleic Acids Res. **19:** 6527–6532
68 Raghu G., Tevosian S., Anant S., Subramanian K. N., George D. L. and Mirkin S. M. (1994) Transcriptional activity of the homopurine-homopyrimidine repeat of the c-Ki-ras promoter is independent of its H-forming potential. Nucleic Acids Res. **22:** 3271–3279
69 Rustighi A., Mantovani F., Fusco A., Giancotti V. and Manfioletti G. (1999) Sp1 and CTF/NF-1 transcription factors are involved in the basal expression of the Hmgi-c proximal promoter. Biochem. Biophys. Res. Commun. **265:** 439–447
70 Santra M., Danielson K. G. and Iozzo R. V. (1994) Structural and functional characterization of the human decorin gene promoter. A homopurine-homopyrimidine S1 nuclease-sensitive region is involved in transcriptional control. J. Biol. Chem. **269:** 579–587
71 Ko J. L. and Loh H. H. (2001) Single-stranded DNA-binding complex involved in transcriptional regulation of mouse mu-opioid receptor gene. J. Biol. Chem. **276:** 788–795
72 Zahedi K., Bissler J. J., Prada A. E., Prada J. A. and Davis A. E. (1999) The promoter of the C1 inhibitor gene contains a polypurine.polypyrimidine segment that enhances transcriptional activity. J. Immunol. **162:** 7249–7255
73 Zahedi K., Prada A. E., Mulligan A., Prada J. A. and Davis A. E. 3rd (2002) Normal transcription of the C1 inhibitor gene is dependent upon a polypurine-polypyrimidine region within the promoter. Inflammation **26:** 183–191
74 Shiraga T., Smith D., Nuthall H. N., Hollingsworth M. A. and Harris A. (2002) Identification of two novel elements involved in human MUCI gene expression in vivo. Mol. Med. **8:** 33–41
75 Rustighi A., Tessari M. A., Vascotto F., Sgarra R., Giancotti V. and Manfioletti G. (2002) A polypyrimidine/polypurine tract within the Hmga2 minimal promoter: a common feature of many growth-related genes. Biochemistry **41:** 1229–1240