Nucleic Acid topology, DNA Motif: A Critical Analysis

***Dr.Devadas Chetty, Assistant Professor of Zoology, HKE's SS Margol Degree College, Shahabad.**

Abstract

This paper attempts to study **Topological** characteristics of **DNA** and specifically **DNA motif** influence all major behavioarl patterns. Nucleic acids are naturally occurring chemical compounds that serve as the primary information-carrying molecules in cells and makeup the genetic material. Nucleic acids are found in abundance in all living things, where they create, encode, and then store information of every living cell of every life-form on Earth. In turn, they function to transmit and express that information inside and outside the cell nucleus to the interior operations of the cell and ultimately to the next generation of each living organism. The encoded information is contained and conveyed via the nucleic acid sequence, which provides the 'ladder-step' ordering of nucleotides within the molecules of RNA and DNA. They play an especially important role in directing protein synthesis.

Topology affects physical and biological properties of DNA and impacts fundamental cellular processes, such as gene expression, genome replication, chromosome structure and segregation. In all organisms DNA topology is carefully modulated and the supercoiling degree of defined genome regions may change according to physiological and environmental conditions. DNA topology is an intrinsic property of DNA molecules, and is controlled by the direct action of DNA topoisomerases. These are enzymes essential for proliferation and survival of all cells and are indeed important targets for chemotherapeutic drugs. In vitro, DNA topoisomerases are able to induce significant changes of linking number (ΔLk; see the Methods Section for definition of the topological parameters) of covalently closed DNA molecules, by either removing or introducing supercoils. Each enzyme has its own specificity, and can modify the ΔLk stepwise, leading to production of a range of differently supercoiled molecules. Elucidation of structural properties of DNA molecules with different topology may thus help to better understand genome functions. Therefore, more structural information is needed to fully understand the effect of capping interactions and the loops connecting the C-tracts on i-motif stability. In spite of the recent findings of i-motif ligands, the number of known i-motif-specific binders is very limited in comparison with G4 ligands. No three-dimensional structure of an i-motif/ligand complex has been determined yet. This will be a major achievement for the development of potential drugs based on i-motif recognition.

Key words: topological spaces, DNA, RNA, ligand complex, motif, enzyme

Introduction

Double-stranded nucleic acids are made up of complementary sequences, in which extensive Watson-Crick base pairing results in a highly repeated and quite uniform Nucleic acid double-helical three-dimensional structure. In contrast, single-stranded RNA and DNA molecules are not constrained to a regular double helix, and can adopt highly complex three-dimensional structures that are based on short stretches of intramolecular base-paired sequences including both Watson-Crick and noncanonical base pairs, and a wide range of complex tertiary interactions.

Nucleic acid molecules are usually unbranched and may occur as linear and circular molecules. For example, bacterial chromosomes, plasmids, mitochondrial DNA, and chloroplast DNA are usually circular double-stranded DNA molecules, while chromosomes of the eukaryotic nucleus are usually linear double-stranded DNA molecules. Most RNA molecules are linear, single-stranded molecules, but both circular and branched molecules can result from RNA splicing reactions. The total amount of pyrimidines in a double-stranded DNA molecule is equal to the total amount of purines. The diameter of the helix is about 20Å. The i-motif represents a paradigmatic example of the wide structural versatility of nucleic acids. In remarkable contrast to duplex DNA, i-motifs are four-stranded DNA structures held together by hemi- protonated and intercalated cytosine base pairs ($C:C^+$). First observed 25 years ago, and considered by many as a mere structural oddity, interest in and discussion on the biological role of i-motifs have grown dramatically in recent years. In this review we focus on structural aspects of i-motif formation, the factors leading to its stabilization and recent studies describing the possible role of i-motifs in fundamental biological processes.

In addition to the B-form DNA double helix (*1*), DNA can adopt a number of alternative non-B DNA structures (*2*,*3*) including G-quadruplex (G4s) (*4*) and intercalated motif (i-motif) structures (*5*) (Figure *1*). The in vivo existence of these structures in human cells remained uncertain until their recent visualization using antibody fragments that recognize them in a structure-specific manner (*6*,*7*). These findings provided key evidence that i-motif structures may be formed in regulatory regions of the human genome, and support the notion that G4 and i-motif structures may play complementary roles in the regulation of gene expression. In this review we focus on structural aspects of i-motif formation and recent studies that describe the possible role of i-motifs in cell biology. We will not extensively review the current uses of i-motif structures in nanotechnological applications, but we will provide a brief summary and refer the reader to excellent reviews on this subject (*8*–*10*).

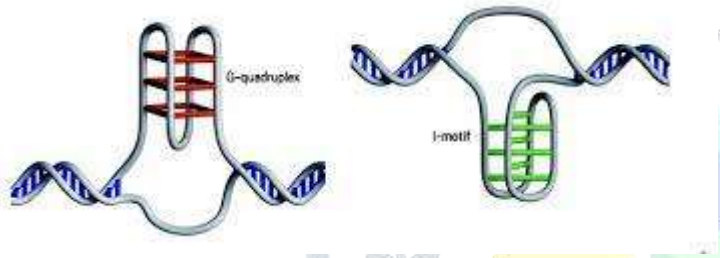
(A) $C:C^+$ base pair. (B) Stacking between intercalating $C:C^+$ base pairs. (C) Two views of the 3D structure of a tetramolecular i-motif (PDB 1YBR), face-to-face and back-to-back steps are highlighted and two hydrogen bonds between O4′ and H1′ are marked with black dashed lines. (D) Scheme representing the 3′E (left) and 5′E (right) intercalation topologies of an intermolecular i-motif structure.

Objective:

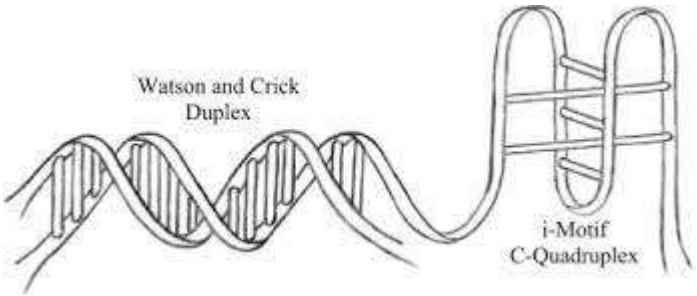
This paper intends to explore and analyze knotting and entanglement that have been observed in **DNA** and proteins, their enigmatic existence in RNA, **Nucleic acid as DNA motif**

DNA i-MOTIF-STRUCTURAL FEATURES

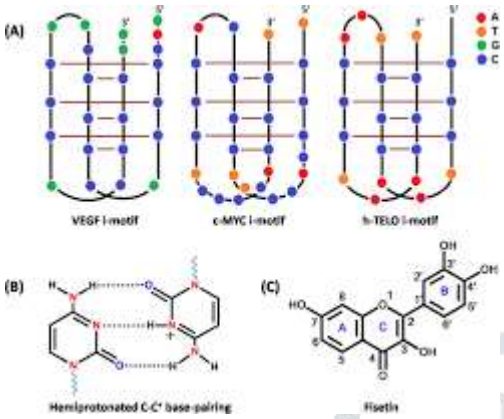
The first DNA i-motif was characterized by Gehring et al. for the hexamer sequence d(TCCCCC) forming an intercalated quadruple-helical tetramolecular structure under acidic conditions (5). It consists of two parallel-stranded duplexes intercalated in an antiparallel orientation and held together by hemi-protonated cytosine-cytosine^{+} ($C:C^{+}$) base pairs (Figure 1A and B) (5,11). Since this report, a number of i-motif structures has been determined by crystallographic and NMR methods (5,11). As in the case of G4 structures, i-motifs may fold in an intermolecular fashion from the association of two (dimers) or four (tetramers) separate DNA strands, or form an intramolecular structure (monomer) due to the spatial arrangement of four different C-tracts within the same strand.



i-Motifs have other very peculiar structural features. The distance between consecutive base pairs is 3.1 Å, and the right-handed helical twist angle is ∼12–20°, both significantly smaller than those measured in B-DNA (3.4 Å and 36°) (12). The intercalation of base pairs from two parallel duplexes leads to a structure with two major (wide) grooves and two minor (narrow) grooves (Figure 1C). The two minor grooves are extremely narrow (3.1 Å versus ∼5.7 Å for B-DNA), giving rise to a number of inter-strand short distances along the sugar phosphate backbones (12). These parameters result in destabilizing interactions due to close phosphate-phosphate distances, which are partially compensated by favorable inter-strand sugar-sugar contacts. Since the minor grooves are formed between antiparallel strands, sugar-sugar contacts occur in two alternative types: face-to-face (ff) and back-to-back (bb) steps. In the ff steps, the ring oxygens (O4′) face one another (5′-side); however, in the bb steps, the C3′-C4′ edges of the sugar moieties are oriented close to each other (3′-side, Figure).



Due to the spatial arrangement of C:C$^+$ base pairs, i-motif structures can be classified in two different intercalation topologies known as 3′E and 5′E (Figure) (13). When the outmost C:C$^+$ base pair is at the 3′-end, the structure is known as 3′E, while in the 5′E topology, the terminal C:C$^+$ base pair is at the 5′-end (14) (Figure 1D). In absence of additional interactions and for a given number of C:C$^+$ base pairs, the 3′E topology is more stable than the 5′E topology due to the extended sugar-sugar contacts along the narrow grooves (15).



Due to the requirement of hemi-protonated base pairs, it was thought that i-motif structures could only fold at acidic pH values; however, several recent studies have shown that i-motif structures can form at neutral pH depending on the sequence and environmental conditions (16,17). Thus, i-motif structures were observed at neutral pH and low temperatures under molecular crowding conditions (18,19), under negative superhelicity (20), in the presence of silver (21) or copper (I) cations (22), and inside silica nano-channels (23). Chemical modifications such as 2′-deoxy-2′-fluoro-arabinocytidine (2′F-araC, known as 2′F-ANA) also induce formation of DNA i-motif structures under neutral conditions (24,25).

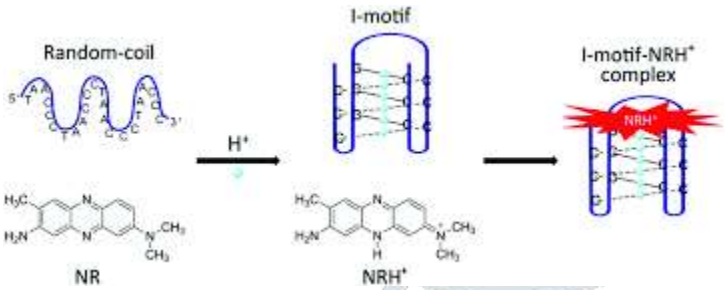
FACTORS AFFECTING THE STABILITY OF i-MOTIF STRUCTURES

Like other nucleic acid structures, i-motif stability depends on many factors, including sequence nature, temperature, and ionic strength. Unlike B-DNA or G-quadruplexes, in which π stacking interactions between sequential nucleobases play an essential role in their stability, the intercalative geometry between consecutive base pairs in i-motif structures gives rise to very little overlap between the six-membered aromatic pyrimidine bases (Figure *1B* and *C*). Although C:C$^+$ base pairs involve the favorable stacking of exocyclic carbonyls and amino groups in an antiparallel fashion, theoretical calculations showed that this and other favorable stacking interactions between consecutive C:C$^+$ base pairs barely compensate the electrostatic repulsion between their charged imino groups (*15*). Of particular relevance to understand the factors affecting i-motif stability are the multiple studies recently reported on chemically modified i-motif structures, which we discussed in detail below.

The C:C$^+$ base pair

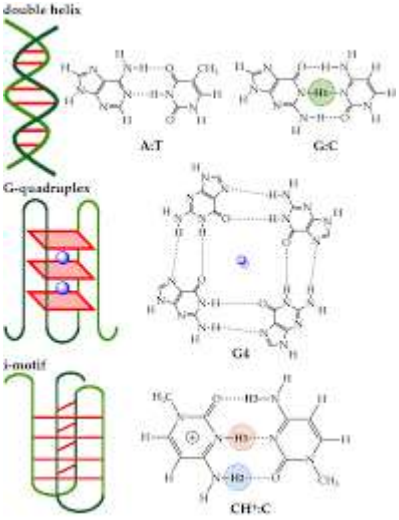
The hemi-protonated C:C$^+$ base pairs are the key interactions for i-motif stability. The three hydrogen bonds of the C:C$^+$ base pair confer a high stability. Computer calculations indicate that the base-pairing energy (BPE) for the C:C$^+$ base pair is 169.7 kJ/mol, higher than BPEs of canonical Watson-Crick G·C (96.6 kJ/mol) and neutral C·C (68.0 kJ/mol) (*26*). The central hydrogen bonding in a hemiprotonated C:C$^+$ (N3···H···N3) base pair has been described as a double-well potential where the proton delocalizes/oscillates between the two wells (*27*).

Leroy et al. estimated the proton-transfer rate to be 8×10^4 s^{-1} (*28*). The NMR structural study of an intramolecular telomeric i-motif (PDB code: 1ELN) showed that the C:C^+ base pairs are planar and the N3–N3 distance is around 2.6–2.8 Å (*29*). Most importantly, protonation at the N3 position produces a positively charged base pair. NMR spectroscopy and computational analyses performed by Lieblein et al. suggested that the N3···H^+···N3 bonds possess an asymmetric double-well potential and that the proton in one C:C^+ base pair tends to adopt a position that leads to the largest distance with respect to the proton of neighboring C:C^+ base pair (*27*).



The effect of chemical modifications in the C:C^+ base pairs has been investigated in different contexts. Wadkins et al. indicated that a cytosine modification might have different effects on i-motif stability depending on the environmental conditions (*30*). For instance, substitution of cytosine by its halogenated analogues such as 5-fluoro, 5-bromo and 5-iodo derivatives, increases stability of i-motifs at acidic conditions (*31*). Furthermore, cytosine methylation at position 5 leads to an increase in the pH of mid transition (i.e. pH_T) and T_m of i-motif structures (Δ pH_T = +0.11 for two sequences containing two 5-methylcytosine substitutions), while hydroxymethylation leads to a decrease in pH_T (ΔpH_T = –0.2) and T_m Nucleoside modifications introduced in i-motif structures. (A) Nucleobase modifications. (B) Sugar modifications and (C) A nucleoside analogue possessing both sugar and nucleobase modifications. d: deoxyribose, r: ribose and ara: arabinose sugar.

Recently, Waller's group investigated the effect of 2′-deoxyriboguanylurea (GuaUre-dR) (Figure 2A) on human telomeric i-motif formation (33). GuaUre-dR is a breakdown product of decitabine, a cytidine analogue that acts as an epimutagen and a chemotherapeutic agent. Despite the fact that this analogue can form a base pair with cytosine without protonation, the modified telomeric sequences exhibited a decrease in the pH_T (5.8) compared to the unmodified i-motifs (pH_T = 6.1) (33). Mir et al. studied the effect of pseudoiso-deoxycytidine (psC) (Figure 2A) on the stability of head-to-head and head-to-tail dimeric i-motif structures (34). An increase of the stability of the studied i-motif structures was observed when the neutral psC:C base pair was located at the end of a C:C^+ stack. However, protonated base pairs are required (psC:C^+ or psC:psC^+) when the psC modifications are located in central positions of the i-motif structures. The results from these two studies underline the importance of the electrostatic interactions in the i-motif stability, where the presence of positive charges in the core of the structures is a key factor for its stability.



Significant i-motif stabilizations have been reported recently by introducing phenoxazine 2′-deoxynucleotides (i-clamps) containing a C8-aminopropynol tether (Figure 2A). These nucleotides are able to form base pairs with protonated cytosines and, simultaneously, interact favorably with the phosphate backbone of the opposite strand. As in the case of psC substitutions, the effect depends on the position of the substitution, where more stabilization is observed when they are located at the 5′-end of the C-stacks (35).

Very recently, natural base lesions were introduced in the TAA loops and in the core cytosines of the human telomeric C-rich strand d(CCCTAA)$_3$CCCT (36). This study reveals that i-motifs containing apurinic sites (apA) and 8-oxoadenine (8oxoA) substituting adenine, and 5-hydroxymethyluracil (5hmU) (Figure 2A) substituting thymine exhibit thermal stabilities that depend on the position of these substitutions in the sequence. Thus, in comparison to the unmodified structure, ΔT_m values for i-motifs having apA substitutions ranges between −1.7 and +5.0°C, those having 8oxoA vary from −1.4 to +3.5°C, and ΔT_m for the ones containing 5hmU is +0.4°C on average. On the contrary, the presence of uracil instead of cytosine (due to the enzymatic or spontaneous deamination of cytosine) substantially reduces i-motif thermal stability. The extent of this destabilization depends on the position of the deaminated cytosines. Whereas the loss of outer cytosines produces a decrease of 11.6–17.0°C in melting temperatures, the destabilization produced by the substitution of inner cytosines is even more pronounced due to the loss of a C:C$^+$ pair intercalated between two inversely oriented pairs (36).

Sugar and phosphate backbone modifications

Sugar modifications

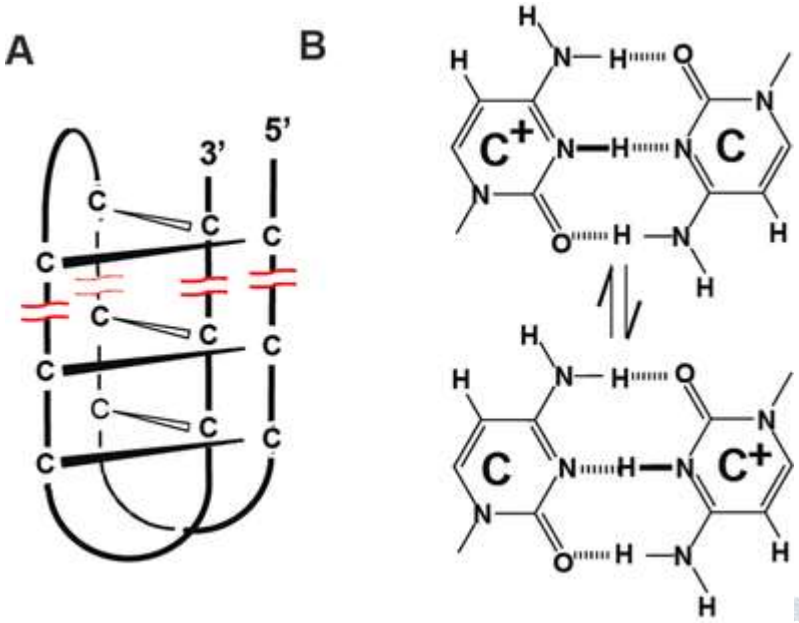
The i-motif folding architecture is destabilized by the close inter-strand phosphorus-phosphorus distances (5.9 Å) along the narrow groove. This destabilizing factor is compensated by an inter-strand favorable interaction network between the deoxyribose sugar moieties along the minor groove (12). Additional stabilizing interactions along the narrow groove results from C-H1′···O4′ interactions within each pair of antiparallel strands (Figure 1C). Moreover, studies have shown intra-nucleoside hydrogen bond between O4′ and H6, and O4′ bonding simultaneously with H1′ and H4′ (12).

Sugar-backbone chemical modifications have shown to be very useful to assess these interactions and modulate i-motif stability. These include cytidines from RNA (37,38), 2′F-RNA (39), locked (LNA) (40) and unlocked (UNA)

nucleic acids (41), 2′-arabinonucleic acids (ANA) (42), acyclic threoninol nucleic acids (aTNA) (43), and more recently, 2′-fluoro-arabinonucleic acids (2′F-ANA) (24,25,44) (Figure 2B).

Interestingly, substituting the 2′-OH in RNA with its 2′-arabinose epimer leads to stable i-motif structures since the OH group is placed in the wide major groove (42). The different stability observed between riboses and arabinoses confirms the critical significance of sugar–sugar contacts in the minor groove on the stabilization of i-motifs. We recently demonstrated that incorporating 2′F-araC (Figure 2B) modifications in i-motif structures leads to significant stabilization over a wide pH range. 2′F-araC is one of very few chemical modifications that stabilize i-motif structures at neutral conditions not only at the ends of the C-tracts, but also in central positions (24). Furthermore, the 2′F-araC modification stabilized intermolecular centromeric and intramolecular telomeric i-motifs in all the positions tested (24). Despite the fact that the nucleoside 2′F-araC exhibits a lower pK_a compared to deoxycytidine (3.9 versus 4.4, respectively), the pH_T in the modified i-motif structures was remarkably higher (+0.7 for centromeric sequences and +0.8 for telomeric sequences), allowing for the observation of these structures at neutral pH. NMR structural determination revealed that the 2′F-araC residues adopt a C2′-endo sugar pucker, instead of the C3′-endo conformation that is usually found in unmodified structures, with the fluorine atom oriented in the major groove. Therefore, 2′F-araC modifications do not perturb the hydrogen-bonding network that provides stability to the structure, but instead lead to additional electrostatic interactions that are absent in the unmodified structure. The above results allow the utilization of i-motif structures in several applications, most importantly in biological assays that require physiological temperature and pH conditions. Following these interesting observations, Aviñó et al. investigated the effect of (2′S)-2′-deoxy-2′-C-methyl-cytidine units (C_{Me}Up) (Figure 2B) on telomeric i-motif structures (46). This modification adopts mainly a C3′-endo sugar pucker with the methyl group at C2′ in the ‘arabino’ (or β) orientation. C_{Me}Up was tolerated in i-motif structures; however, stabilization was less pronounced compared to 2′F-araC. This result confirms the role of favorable electrostatic interactions induced by the electronegativity of the fluorine atom in the enhanced stability provoked by 2′F-araC.

In conclusion, most sugar modifications destabilize i-motif structures, regardless of whether or not they favor the C3′-endo sugar conformation. Substituents oriented toward the compact minor groove cause steric clashes that destabilize the structure. However, chemical modifications that preserve sugar-sugar contacts across the minor groove, such as arabinose sugars, are well tolerated. Of particular relevance is the effect of 2′F-araC substitutions, which lead to stable structures at pH 7.



In humans, a 171 bp DNA called alphoid satellite is tandemly repeated along the centromeres (Figure). The alphoid satellite is an AT rich sequence that frequently contains a 17-bp GC rich segment whose sequence has two variants known as CENP-B box and A box (Figure 6A). Interestingly, the CENP-B box, the sequence specifically bound by centromeric protein CENP-B, is absent in lower primates and in human Y chromosome. Gallego et al. found that whereas the G-rich sequence of CENP-B box folds into a structure stabilized by canonical base pairs, its complementary C-rich sequence forms a dimeric i-motif at acidic pH (66,102). Likewise, the C-rich sequences of the two variants of A box were reported to fold into dimeric i-motif structures

Conclusion

Human chromosome showing the centromere in grey and the structure of the human centromeric alpha satellite DNA. The 171 bp monomers (green arrows) are AT-rich sequences that can contain either A box (red and black) or CENP-B box (orange and black). The monomers are tandemly repeated and form a higher order repeat, which in turn appears repeated along the centromere. (B) Schematic and cartoon structures of the dimeric i-motifs formed by truncated A box (PDB ID: 2MRZ) and CENP-B box (PDB ID: 1C11). Cytosines are represented in grey, thymines in green, adenines in magenta and guanines in blue. (C) Proposed model for nucleosome organization at the centromere.

The i-motifs formed by C-rich sequences of A and CENP-B boxes (orange) would maintain the structural organization of the centromere. However, in the genome and with the exception of chromosome ends, the formation of G4 and i-motif structures is compromised by hybridization between complementary strands to form a duplex DNA.

Several factors have been found to affect this competence. The first G4/i-motif-duplex interconversion studies were carried out by Phan and Mergny. They inferred that 1:1 mixtures of the C-rich $d((C_3TA_2)_3C_3T)$ and G-rich $d(AG_3(T_2AG_3)_3)$ telomeric sequences at acidic pH (<5) and in the presence of KCl produced predominantly i-motif and G4, respectively (115). However, at pH 7.0 and 100 mM NaCl the duplex formed by the hybridization of both sequences was the predominant species. Several groups have investigated the factors influencing interconversion kinetics and determined that the sequence, the experimental conditions (i.e. ionic strength, temperature, and pH) (116,117), and the incorporation of chemical modifications all play a significant role in favoring the tetraplex structures over duplexes and vice versa.

References

1. "What is DNA". What is DNA. Linda Clarks. Retrieved 6 August 2016.
2. Bill Bryson, A Short History of Nearly Everything, Broadway Books, 2015.p. 500.
3. Dahm R (January 2008). "Discovering DNA: Friedrich Miescher and the early years of nucleic acid research". Human Genetics. 122 (6): 565–81. doi:10.1007/s00439-007-0433-0. PMID 17901982. S2CID 915930.
4. Cox M, Nelson D (2008). Principles of Biochemistry. Susan Winslow. p. 288. ISBN 9781464163074.
5. "DNA Structure". What is DNA. Linda Clarks. Retrieved 6 August 2016.
6. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. (February 2001). "Initial sequencing and analysis of the human genome" (PDF). Nature. 409 (6822): 860–921. Bibcode:2001Natur.409..860L. doi:10.1038/35057062. PMID 11237011.
7. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. (February 2001). "The sequence of the human genome". Science. 291 (5507): 1304–51. Bibcode:2001Sci...291.1304V. doi:10.1126/science.1058040. PMID 11181995.
8. Budowle B, van Daal A (April 2009). "Extracting evidence from forensic DNA analyses: future molecular biology directions". BioTechniques. 46 (5): 339–40, 342–50. doi:10.2144/000113136. PMID 19480629.
9. Elson D (1965). "Metabolism of Nucleic Acids (Macromolecular DNA and RNA)". Annual Review of Biochemistry. 34: 449–86. doi:10.1146/annurev.bi.34.070165.002313. PMID 14321176.
10. Dahm R (January 2008). "Discovering DNA: Friedrich Miescher and the early years of nucleic acid research". Human Genetics. nih.gov. 122 (6): 565–81. doi:10.1007/s00439-007-0433-0. PMID 17901982. S2CID 915930.
11. Brock TD, Madigan MT (2009). Brock biology of microorganisms. Pearson / Benjamin Cummings. ISBN 978-0-321-53615-0.
12. Hardinger, Steven; University of California, Los Angeles (2011). "Knowing Nucleic Acids" (PDF). ucla.edu.

13. Mullis, Kary B. The Polymerase Chain Reaction (Nobel Lecture). 1993. (retrieved December 1, 2010) http://nobelprize.org/nobel_prizes/chemistry/laureates/1993/mullis-lecture.html
14. Verma S, Eckstein F (1998). "Modified oligonucleotides: synthesis and strategy for users". Annual Review of Biochemistry. 67: 99–134. doi:10.1146/annurev.biochem.67.1.99. PMID 9759484.
15. Gregory SG, Barlow KF, McLay KE, Kaul R, Swarbreck D, Dunham A, et al. (May 2006). "The DNA sequence and biological annotation of human chromosome 1". Nature. 441 (7091): 315–21. Bibcode:2006Natur.441..315G. doi:10.1038/nature04727. PMID 16710414.
16. Todorov TI, Morris MD (April 2002). National Institutes of Health. "Comparison of RNA, single-stranded DNA and double-stranded DNA behavior during capillary electrophoresis in semidilute polymer solutions". Electrophoresis. nih.gov. 23 (7–8): 1033–44. doi:10.1002/1522-2683(200204)23:7/8<1033::AID-ELPS1033>3.0.CO;2-7. PMID 11981850.
17. Margaret Hunt; University of South Carolina (2010). "RN Virus Replication Strategies". sc.edu.
18. McGlynn P, Lloyd RG (August 1999). "RecG helicase activity at three- and four-strand DNA structures". Nucleic Acids Research. 27 (15): 3049–56. doi:10.1093/nar/27.15.3049. PMC 148529. PMID 10454599.
19. Stryer, Lubert; Berg, Jeremy Mark; Tymoczko, John L. (2007). Biochemistry. San Francisco: W.H. Freeman. ISBN 978-0-7167-6766-4.
20. Rich A, RajBhandary UL (1976). "Transfer RNA: molecular structure, sequence, and properties". Annual Review of Biochemistry. 45: 805–60. doi:10.1146/annurev.bi.45.070176.004105. PMID 60910.
21. Watson JD, Crick FH (April 1953). "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid". Nature. 171 (4356): 737–8. Bibcode:1953Natur.171..737W. doi:10.1038/171737a0. PMID 13054692. S2CID 4253007.
22. Ferré-D'Amaré AR, Doudna JA (1999). "RNA folds: insights from recent crystal structures". Annual Review of Biophysics and Biomolecular Structure. 28: 57–73. doi:10.1146/annurev.biophys.28.1.57. PMID 10410795.
23. Alberts, Bruce (2008). Molecular biology of the cell. New York: Garland Science. ISBN 978-0-8153-4105-5.
24. Gilbert, Walter G. 1980. DNA Sequencing and Gene Structure (Nobel Lecture) http://nobelprize.org/nobel_prizes/chemistry/laureates/1980/gilbert-lecture.html
25. Sanger, Frederick. 1980. Determination of Nucleotide Sequences in DNA (Nobel Lecture) http://nobelprize.org/nobel_prizes/chemistry/laureates/1980/sanger-lecture.html
26. NCBI Resource Coordinators (January 2014). "Database resources of the National Center for Biotechnology Information". Nucleic Acids Research. 42 (Database issue): D7-17. doi:10.1093/nar/gkt1146. PMC 3965057. PMID 24259429.