Translated from *Uspekhi Khimii* **58** 1318–1353 U.D.C. 577.113.7

Complexes of oligo(poly)nucleotides with structural anomalies

N G Dolinnaya, O I Gryaznova
Faculty of Chemistry, Lomonosov Moscow State University

ABSTRACT. The results of studies on the structure and properties of DNA-RNA hybrids and complexes of oligo(poly)nucleotides containing non-canonical base pairs or unpaired bases both within and at the ends of the double helix are surveyed. The methods used in the study of such systems are briefly characterised: X-ray diffraction analysis, NMR and UV spectroscopy, circular dichroism, scanning microcalorimetry, etc. A comparative analysis of the influence of the non-canonical pairs on the structure and the energetic and kinetic parameters of the formation and dissociation of the oligonucleotide complexes has been carried out. The question of the stability of the non-canonical pairs as a function of their nature and position in the double helix is considered. The mechanisms of the formation of the hydrogen bonds between the bases of non-complementary pairs are discussed.

The bibliography includes 171 references.

Contents

I. Introduction 758
II. Methods used to investigate double helix nucleic acids with structural anomalies 758
III. Hybrid DNA–RNA complexes 759
IV. Complexes of nucleic acids with non-canonical base pairs 763
V. Comparative analysis of the influence of the non-canonical pairs on the structure and energetic parameters of the formation of the oligomeric complexes 771
VI. Complexes of nucleic acids containing unpaired nucleotide residues 772

I. Introduction

The interest in the double helix complexes of nucleic acids with various structural anomalies arose in connection with the study of the mechanism of the transmission of genetic information. Thus it has been established that hybrid DNA–RNA duplexes are formed as intermediates in key biological processes—replication, transcription, and reverse transcription.[1,2] The correct "reading" of genetic information is known to be ensured by the complementation interactions of the nucleic acid (NA) bases and is controlled by the corresponding enzyme systems. However, the inclusion of "erroneous" nucleotides, forming non-complementary pairs, nevertheless occurs, although extremely rarely, leading to substitutional mutations.[3,4] On the other hand, non-canonical pairs are formed fairly frequently in codon-anticodon interactions (in the third position of the codon triplet)[5] and also on formation of the secondary or tertiary structures of different RNAs.[6–8] The development of the site-specific mutagenesis served as a powerful stimulus to the study of the spatial disposition and energetic parameters of the non-canonical pairs incorporated in the regular double helix structure. The use of the site-mutagenesis technique makes it possible to achieve point substitutions of nucleotides, deletions, and insertions in specified positions in DNA. The appearance of interest in the study of structures of this type coincided with the development of synthetic nucleotide chemistry and the improvement of modern methods for structural studies. This made it possible in recent years to obtain and to study in detail sets of DNA duplexes with non-canonical pairs and "superfluous" nucleotide units in specified positions. Such compounds are convenient models for the investigation of various cell processes—replication, transcription, translation, and mutagenetic processes. In addition, they are of interest as instruments for the study of enzymes of NA metabolism and the structure and conformational possibilities of NA. The available information about the thermodynamics and kinetics of the formation and dissociation of DNA-duplexes with non-canonical pairs is used widely for the diagnosis of genetic diseases.[9]

The data for the geometry of the pairing of certain types of non-complementary bases and their influence on the parameters of the double helix have been analysed in review articles published in 1987.[10,11] This review considers the structure of the DNA–RNA hybrid structures and NA complexes containing all the investigated combinations of the non-complementary pairs and "superfluous" nucleotide units; the influence of structural anomalies on the thermodynamic and kinetic parameters of complex formation and on the secondary structure of the double helices is considered. The structural anomalies induced by minor or modified bases are not included in the review.

II. Methods used to investigate double helix nucleic acids with structural anomalies

X-Ray diffraction analysis of single crystals (XRDA) and X-ray diffraction in oriented fibres (XRDF). The most complete information about the structure of oligonucleotide duplexes independent of conceptual models can be obtained with the aid of XRDA; high-resolution XRDA (~1.5–2 Å) makes it possible to determine the coordinates of all the non-hydrogen atoms, to observe experimentally the influence of modifications of the local and overall structures of the double helix, and to investigate its hydration shell.[10,12–14] XRDF is used to characterise the double and triple helices formed by polynucleotides[15] and yields much less information than XRDA.

High resolution NMR spectroscopy. Such spectroscopy has been widely used recently to investigate oligomeric duplexes.[16] From the ^{1}H NMR spectra, one can infer the local environment of each

proton, which makes it possible to follow the variation of the conformations of the individual nucleotide pairs as the temperature, pH, etc. are varied. Two groups of resonance proton signals are analysed.

(a) The signals of non-exchanging protons—H(5) and H(6) of the pyrimidine bases, H(2) and H(8) of the purine bases, and the protons of the CH_3 group of thymine and the carbohydrate rings. The thermal dissociation of the duplexes leads to a downfield shift of the signals. The thermodynamic parameters of the helix–coil transition can be calculated from the curves for the temperature dependence of the chemical shifts δ of these proton signals and it is possible to follow the dissociation of each nucleotide pair of the duplex.

(b) The signals of the slowly exchanging protons of the imino-groups N(1)–H of guanine and N(3)–H of thymine (uracil), participating in the formation of hydrogen bonds, are recorded in aqueous solutions. The presence of signals in the region of low fields indicates unambiguously the formation of hydrogen-bonded nucleotide pairs. With increase in temperature, an upfield shift, broadening, and disappearance of the signals take place owing to the increase in the relative content of the open-chain forms and the rapid exchange of the imino-group protons in the open-chain forms with water. The temperature of the disappearance of the signals is not related to the thermodynamic stability of the duplexes and merely yields a comparative characteristic of the stability of the individual nucleotide pairs. The study of the kinetics of the exchange of the imino-protons with water makes it possible to estimate the lifetimes τ of the individual base pairs in the helical complex. The activation energies and the rate constants for the dissociation and formation of individual nucleotide pairs are calculated from the temperature variation of τ.[17]

Measurement of the vicinal spin–spin coupling constants of the protons of the C(1′) and C(2′) atoms makes it possible to estimate the conformation of the carbon rings. Information about the distances between the atomic nuclei can be obtained with the aid of the nuclear Overhauser effect (nOe);[18] the nOe is observed for pairs of nuclei separated by distances not greater than 4 Å.

^{31}P NMR spectra contain data about the conformation of the sugar phosphate skeleton of the oligonucleotide constituting the double helix.

Scanning microcalorimetry (SMC). This is used to determine the thermodynamic parameters of the conformational transition.[19] The enthalpy of the helix–coil transitions ΔH^0 can be obtained directly from the curve for the temperature variation for the heat capacity of the system by calculating the area bounded by this curve. The van't Hoff enthalpy $\Delta H^0_{v\text{-}H}$ can be calculated simultaneously from the same experimental data.[20] It has been suggested that $\Delta H^0_{v\text{-}H}$ corresponds to the enthalpy of the helix–coil transition when the two states model is valid. The difference between the experimental ΔH^0 and $\Delta H^0_{v\text{-}H}$ is evidence that on fusion the system passes through a series of partially bound structures and makes it possible to estimate the length of the cooperatively melting section of the duplex.

UV spectroscopy. The curves for the temperature variation of the UV absorption by the system permit the determination of the melting point (T_m) of the double helix and the calculation of the thermodynamic parameters of the complex formation process. The temperature jump method,[21] based on the analysis of the change in the UV absorption after an abrupt increase in temperature, is widely used to obtain the kinetic characteristics of the helix–coil transition.

Circular dichroism (CD) and optical rotary dispersion (ORD). The interpretation of CD (ORD) data, based mainly on the empirical assignment of the spectra to the known conformations of the polymers, permits conclusions concerning the similarity of the geometrical parameters of the double helix structures and their inclusion in a particular study of Na forms ($\overline{A}$, $\overline{B}$, $\overline{Z}$).

Raman spectroscopy (RS). This spectroscopy makes it possible to estimate the secondary structure of double helix oligo(poly)nucleotides both in solution and in the crystalline states.[22]

Theoretical estimation of the conformation of the oligonucleotide duplexes. This is done by calculating the energies of the non-valence interactions by the method of atom–atom potential functions (AAPF).[23] The minimisation of the energy of the non-valence interactions in terms of different variables (parameters determining the geometry of the pairing of bases and the positions of the pairs relative to the axis of the helix, parameters of the helix, torsional angles, etc.) makes it possible to establish the most probable conformation of the non-canonical pairs and of the duplexes containing such pairs. It is noteworthy that the accuracy of calculations of this kind depends on the number of independent variables considered.

III. Hybrid DNA–RNA complexes

1. The secondary structure of the DNA–RNA hybrids

Natural DNA and synthetic polydeoxyribonucleotides are known to exist in aqueous solution and in fibres with a high relative moisture content in the form of the $\overline{B}$ family; forms belong to the $\overline{A}$ family are characteristic of the double helix RNA. Although both conformational families refer to right-hand helices, they differ sharply in their helix parameters. A significant factor determining the type of secondary structure is the conformation of the sugar: C(3′)-endo(*N*) in the $\overline{A}$ family and C(2′)-endo(*S*) in the $\overline{B}$ family. The only form in which the double helix RNA exists is the *A*-form. The conformational possibilities for DNA are much wider: depending on the conditions in the medium and the nucleotide sequence, the DNA duplexes can exist in the form of the $\overline{B}$ and $\overline{A}$ families, the alternating *B*-form, and the left-hand *Z*-form. It has also been shown that the local parameters of the nucleotide pairs vary within wide limits.[24]

By analysing the available literature data, it is possible to conclude that the DNA–RNA hybrids belong most typically to the $\overline{A}$-family of conformations. However, depending on the nucleotide composition and sequence and also on the complex formation conditions, the hybrid complexes can form structures of a new type—*B*- and *Z*-complexes as well as dimorphous complexes containing elements of both forms (Table 1).

We shall consider in greater detail some of the data presented in Table 1. The conformations of the hybrid heptameric complexes†

d(CTTTTTG) (GAAAAAC)r	d(CA_5G) + r(CU_5G)
"rA.dT" complex	"dA.rU" complex

† Here and henceforth the structure of the duplexes is formulated in the direction (5′–3′)–upper strand and (3′–5′)–lower strand.

were analysed simultaneously with the study of their ribo- and deoxyribo-analogues:[29,30]

d(CTTTTTG)	r(CUUUUUG)
(GAAAAAAC)d	(GAAAAAAC)r
"dA.dT" complex	"rA.rU" complex

Table 1. The structure of the DNA-RNA hybrid complexes.

Complexes	Form of helix	Method of analysis	Refs.
Polymeric DNA-RNA complexes with an irregular nucleotide sequence	*A*-DNA, *A*-RNA	XRDF, ORD	[25—27]
poly[d(A C)·poly[r(G U)]	closer to *A*-RNA	CD	[28]
poly[r(A C)·poly[d(G T)]	closer to *A*-RNA	CD	[28]
r(C A_5 G)·d(C T_5G)	*A*-RNA	CD, ^{1}H NMR	[29, 30]
d(C A_5 G)·2r(C U_5 G)	*A*-RNA	CD, ^{1}H NMR	[29, 30]
poly (dG)·poly(rC)	*A*-DNA*, *A*-RNA**	XRDF	[15]
poly(rG)·poly(dC)	*A*-DNA	XRDF	[15]
poly(dA)·poly(rU)	*A'*-RNA*, dimorphous structure**	XRDF	[15]
	A-RNA	CD	[31]
poly(dI) ***·poly(rC)	dimorphous structure	XRDF	[15]
poly(rI)·poly(dC)	*A*-DNA*, *A'*-RNA**	XRDF	[15]
poly(rA).poly(dT) (fibres)	*A'*-RNA*, dimorphous structure**	XRDF, ^{31}P NMR	[32, 33]
poly(rA).poly(dT) (in solution)	*B*-DNA,	^{1}H NMR	[34]
	dimorphous structure	CD	[31]
poly(dA)·2poly(rU)	*A*-RNA	XRDF	[35]
poly(rA)·2poly(dT)	dimorphous structure	CD	[31]
d(T C A C A T)·r(A U G U G A)	*B*-DNA	^{1}H NMR	[36]
poly(rG-dC)·poly(rG-dC)	*A*-RNA, *Z*-, *B*-form	CD, ^{31}P NMR	[37—39]
r(GCG)d(CGC)·r(GCG)d(CGC)	*A*-DNA	RS	[22]

*Relative humidity below 80%.
**Relative humidity above 80%.
***I = inosine.

On the basis of the spin–spin coupling constants for the protons of the carbohydrate residues, it has been established that the furanose rings in the "rA.rU" and "rA.dT" complexes are in the *N*-conformation (*A*-form), while the deoxyribose residues in "dA.dT" are in the *S*-conformation (*B*-form). The complex formation reaction between d(C A_5G) and r(C U_5G) leads to the preferential formation of the triple helix d(C A_5G).2r(C U_5G), which belongs to the *A*-family according to CD data.

Under the conditions of a high relative humidity, the polymeric complexes poly(dA).poly(rU) and poly(dI).poly(rC) (in the form of fibres) have unusual diffraction patterns representing a superposition of the reflections corresponding to the *A*- and *B*-forms of the helix.[15] It has been suggested that the carbohydrate residues of the polymeric strands in these complexes have different conformations—of the *N*-type for ribose and of the *S*-type for deoxyribose. A similar dimorphous structure of the DNA–RNA hybrid complex can play an important role in the functioning of the DNA- and RNA-polymerases.

Interesting data have been obtained in studies of poly(rA).poly(dT). This ribo-deoxyribonucleotide hybrid is an analogue of the DNA-duplex poly(dA).poly(dT). The latter retains the *B*-form of the helix even under conditions where the natural non-regular DNA exists in the *A*-form. Apparently the "B affinity" is retained also in the hybrid complex despite the presence in the latter of the conformationally rigid polyribonucleotide chain. The poly(rA).poly(dT) fibres form a helix belonging to the $\overline{A}$ family only for a relative moisture content below 90%.[32] With increase in moisture content, there is a conformational transition to a structure similar to the *B*-form of DNA in terms of geometrical parameters but the structure is dimorphous according to the conformation of the carbohydrate residues.[32,33] It has been shown by ^{1}H NMR, using nOe studies, that in an aqueous buffer solution poly(rA).poly(dT) exists in the *B*-form and the carbohydrate residues in both strands have the same *S*-type conformation.[34] However, owing to the steric hindrance generated by the 2′-hydroxy-group of ribose, the geometry of the hybrid helix differs little from that of the classical *B*-form of the DNA. In particular, the inclination and twist of the bases as well as the distances between the bases and the axis of the helix have increased. It has been found that the *B*-form of the helix in solution is characteristic also of the duplex with the heterogeneous composition d(TCACAT).r(AUGUGA).[36]

The study of the structure of the double helix complexes whose composition includes covalently bound DNA and RNA fragments is of special interest. The mixed composition hybrid with the alternating ribodeoxyribonucleotide sequence poly(rG-dC).poly(rG-dC) is an analogue of the DNA-duplex poly(dG-dC).poly(dG-dC); the latter undergoes a transition into the left-hand *Z*-form at a high ionic strength.[40] At a moderate NaCl concentration, the structure of poly(rG-dC).poly(rG-dC) is analogous to *A*-RNA in terms of geometrical parameters and the conformation of the carbohydrate residues.[37] The transition of the hybrid helix to the *Z*-form occurs, in contrast to its DNA-analogue, both with increase and with decrease in the ionic strength with average *A–Z* transition points at 25 mM NaCl in the low salt region and at 1500 mM NaCl in the high salt region.[37,38] In the presence of tetrabutylammonium ions (> 30 mM) an additional conformational transition of the hybrid complex to the *B*-form is observed.[39]

The hexanucleotide mixed composition duplex

r(G C G)d(C G C)
(C G C)d(G C G)r

forms a helix in the *A*-form in the crystalline state according to RS data[22] and all the carbohydrate residues (both ribose and deoxyribose) have the same *N*-conformation. In aqueous solution, the geometry of the complex is in the main retained, but the deoxyribose rings of the terminal cytidine residues assume an *S*-type conformation, which is associated, according to the authors, with the lower stability of the end pairs and hence their greater conformational mobility.

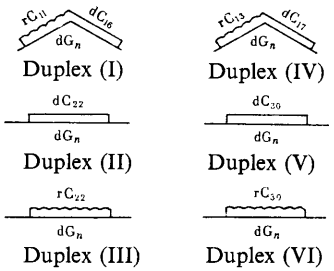
The secondary structure of the decameric duplex

1 2 3 4 5 6 7 8 9 10
r(G C G)d(T A T A—C G C)
(C G C—A T A T)d(G C G)r ,

in which the double helix tetranucleotide DNA fragment (d.d) is flanked by hybrid sections on both sides, has been investigated.[41,42] According to XRDA data, in the presence of a low water content, this duplex forms a helix which is similar, in terms of geometrical parameters, to the *A*-form of DNA with the same *N*-conformation of the carbohydrate ring.[41] According to the authors, the presence of hybrid sections at the end of the duplex stimulates the transition of the central d.d fragment to the *A*-form. In aqueous solution, the situation is somewhat different.[42] According to ^{1}H NMR data, which made it possible to estimate the ratio of the conformers for each nucleotide unit, all the ribose rings of the above duplex have the *N*-conformation. The existence of the *S*- and *N*-conformers is characteristic of all the deoxyribose residues located in both the hybrid and d.d sections of the duplex, the equilibrium constant

depending on the local environment of the individual nucleotide unit (Fig. 1). The carbohydrate residues dC (in the 8-position), dG (in the 9-position), and dC (in the 10-position), belonging to the hybrid fragment, contain respectively 52, 37, and 33% of the *N*-type conformers. This sequence in the variation of the percentage contents of the *N*-conformers is caused by the greater conformational freedom of the terminal deoxynucleotides. The data obtained permit the conclusion that a conformational transmission effect (the influence of the RNA fragment on the conformation of the carbohydrate rings of the DNA sequence linked to it) is observed with covalently bound DNA–RNA hybrid molecules. In the duplex considered, the effect is extended mainly to the two nucleotide residues dT (in the 4-position) and dA (in the 5-position) and to a lesser extent DT (in the 6-position) and dA (in the 7-position). Consequently there is a smooth transition from the *N*-conformation to the predominatly *S*-conformation. The occurrence of the stacking interaction between the bases along the entire molecule, with the predominance of the "stacking" characteristic of the *A*-form in r(GCG) and the "mixed" type in the d(TATACGC) section, has been demonstrated by ^{1}H NMR spectroscopy.

The structures of the hybrid polymeric-oligomeric duplexes $dG_n.rC_{11}dC_{16}$ and $dG_n.rC_{13}dC_{17}$ ($n > 70$) have been investigated.[43,44] The systems investigated are indicated below in a conventional schematic form.



It follows from the analysis of the ^{1}H NMR spectra of the duplexes (I)–(III) that the two fragments of the duplex (I) form helices with different conformations—the hybrid section of the molecule is in the *A*-form characteristic of the duplex (III) and the (d.d) section is in the *B*-form, like the duplex (II).[43] The CD spectra of the duplexes (IV)–(VI) have led to the analogous conclusion that the *A*- and *B*-forms coexist in the hybrid duplex (IV).[44] Analysis of the fusion curves (UV spectroscopy) shows that, despite the presence of two forms, the helix–coil transition in the duplex (IV) is cooperative and single-phase. According to the authors,[44] the existence of two adjoining conformations in the covalently bound hybrid duplexes (I) and (IV) leads to geometrical distortions—the bending of the axis of the helix at the boundary between the two forms; the transitional region between the *A* and *B* forms then includes only one nucleotide pair, belonging to the DNA fragment, and the phosphodiester linkages adjoining it. Analysis of the geometrical structure of the hybrid duplex (Fig. 2), obtained by computer modelling,[44,45] showed that in the region of the bend in the helix the complementation and interplanar interactions between the bases are retained.

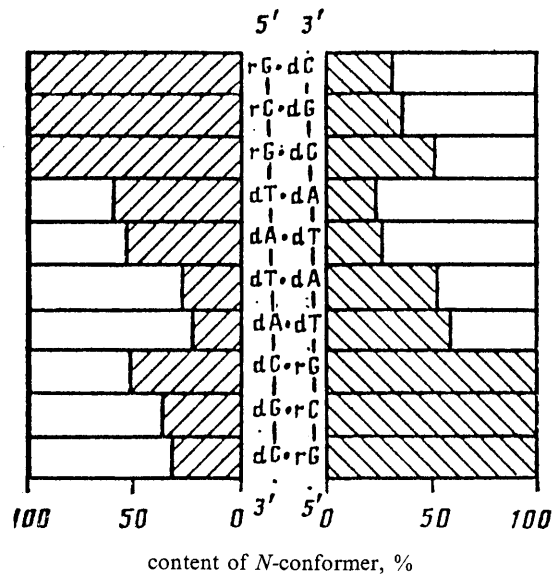


Figure 1. Conformation of the carbohydrate rings of the duplex formed by r(GCG)d(TATACGC) in aqueous solution (according to ^{1}H NMR spectroscopic data).[42]

The conclusions concerning the structure of the covalently bound DNA–RNA hybrid duplexes reached in this series of studies differ from the conclusions arrived at by Mellema et al.[42] for the

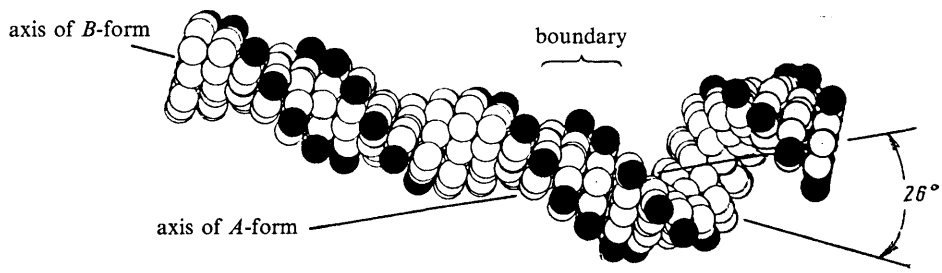


Figure 2. A molecular model of the complex $dG_n.rC_{11}dC_{16}$.[44]

decameric duplex, which can be accounted for by the differences between the lengths of the r.d and d.d sections, the nature of the nucleotide pairs (A.T or G.C) at the "junction" between these sections, and certain assumptions in the interpretation of the data.

2. Thermodynamic and kinetic aspects of the formation of DNA–RNA hybrid complexes

A comparative analysis of the thermodynamics and kinetics of the formation of hybrid double helices and of their DNA and RNA analogues has been performed on the model heptameric complexes "rA.dT", "dA.rU", "dA.dT", and "rA.rU" (see above).[17,30,46] The following duplex stability sequence has been established with the aid of UV and ^{1}H NMR spectroscopy: "dA.dT" > "rA.dT" > "rA.rU" » "dA.rU", which agrees with the stability series for polymeric complexes formed by poly(rA), poly(dA), poly(rU), and poly(dT) in analogous combinations.[47] Comparison of the equilibrium constants ($K_{25°}$) for the formation of the heptameric double helices (Table 2) shows that the duplex "dA.dT" is 10 and 4 times more stable than "rA.rU" and "rA.dT" respectively, while the hybrid duplex "dA.rU" is almost 200 times less stable than "rA.dT". It has been suggested[30] that the *S*-conformation is preferred for the dA clusters and the latter cannot therefore form stable DNA–RNA hybrids. According to the authors, the presence of such clusters facilitates the termination of transcription.

Table 2. Thermodynamic and kinetic parameters of the formation of hybrid oligomeric duplexes and the corresponding ribo- and deoxyribo-nucleotide analogues.

Parameter	«rA·dT»	«dA·rU»	«dA·dT»	«rA·rU»	Refs.
T_m, °C	22.9	<0	27.5	19,1	[30]
$K_{25°}$, litre mol^{-1}	$5{,}5\cdot10^3$	$<3\cdot10^1$	$1.8\cdot10^4$	$2.2\cdot10^3$	[30]
$\Delta H°$, kcal mol^{-1}	−48	—	−41	−44	[30]
$k_{p,12°}$, litre mol^{-1} s^{-1}	$6.6\cdot10^6$	—	$8.3\cdot10^{-6}$	$6.2\cdot10^6$	[46]
$E_{a,d}$, kcal mol^{-1}	—	—	43	39	[46]
$E_{a,r}$, kcal mol^{-1}	—	—	−0.5	0	[46]
$K_{d,12°}$, s^{-1}	50	—	9.6	38	[46]

Note. Complex formation conditions: 0.01 M phosphate buffer, pH 7, 0.1 mM EDTA, 0.2 M NaCl, $c_0 = 2 \times 10^{-4}$ M; the structures of the duplexes are shown in the text.

Kinetic parameters[17,46] have confirmed the thermodynamic data concerning the stability sequence of the duplexes "dA.dT", "rA.rU", and "rA.dT". The rate constant (K_r) for the formation of the duplex "dA.dT" is a little greater than the approximately equal corresponding constants for the duplexes "rA.rU" and "rA.dT" (Table 2). It has been suggested that this can be accounted for by the slower formation of the stacking-conformation in the ribonucleotide chain.[46,48] On the other hand, the approximate quality of the activation energies ($E_{a,r}$) for the formation of the above duplexes (Table 2) suggests identical mechanisms of the association of the ribo- and deoxyribo-nucleotide strands. The kinetic parameters characterising the dissociation of the double helices differ appreciably (Table 2). The rate constant (K_d) for the dissociation of the duplex "dA.dT" is appreciably lower than the corresponding values determined for the duplexes "rA.rU" and "rA.dT". This feature has been confirmed by the results of studies of the kinetics of the exchange of the imino-protons with water:[17] it has been established that the values of τ for the non-terminal pairs of the duplexes "rA.rU" and "rA.dT" are approximately the same and significantly lower than τ for "dA.dT".

According to ^{1}H NMR data, the end "fraying" effect, i.e. the rapid formation and breakdown of hydrogen bonds between the bases at the end of the duplex under the conditions of the thermodynamic stability of the double helix, is observed for these duplexes. It is essential to emphasise that the "fraying" of the end pairs is characteristic of RNA, DNA, and hybrid duplexes to approximately equal extents.

Overall, the data concerning the comparative stabilities of the hybrid, ribonucleotide, and deoxyribonucleotide complexes are ambiguous. For the majority of systems investigated, it has been shown that the hybrid duplexes are less stable than DNA and RNA of the same composition. For polymeric complexes, the difference is very small and amounts to 2–4 °C.[25,26,28,47,49] In oligomeric systems, the difference is more clear cut. Thus T_m for the complexes $rA_n.pdT_9$ (n = 6–9) is on average lower by 7 °C than T_m for $dA_n.pdT$ with the same n.[50]

On the other hand, other relative stabilities of the hybrid duplexes have been established in a number of studies. For example, the stability of the polymeric complexes, formed by different combinations of poly[d(r)I] and poly[d(r)C], varies in the sequence poly(rI).poly(rC) > poly(rI).poly(dC) > poly(dI).poly(dC) > poly(dI).poly(rC),[51] while for covalently bound DNA–RNA hybrids of the type $dG_n.rC_idC_k$, the following stability series has been determined: $dG_n.dC_{30} < dG_n.rC_{13}dC_{17} < dG_n.rC_{30}$.[44]

Analysis of the rates of exchange of the imino-proton duplexes poly(rA).poly(dT), poly(dA).poly(rU), poly(rA).poly(rU), and poly(dA).poly(dT) for deuterium[52] makes it possible to determine the ratios of the rates of dissociation of these complexes and hence their relative stabilities—poly(dA).poly(dT) > poly(dA).poly(rU) ⩾ poly(rA).poly(rU) > poly(rA).poly(dT). The relative stabilities of the hybrid complexes and their DNA and RNA analogues apparently depend on the nucleotide composition.

Generalising the literature data on the DNA–RNA hybrid complexes, the following conclusions can be reached:

1. The conformationally more rigid ribonucleotide chain has a dominant influence in the majority of systems investigated on the geometrical structure of the duplex as a whole and also on the conformation of the carbohydrate rings of the deoxyribonucleotide residues (complementary with respect to the ribo-fragment or covalently bound to it).

2. A dimorphous structure, i.e. different conformations of the ribo- and deoxyribo-nucleotide strands of the helix, is characteristic of a series of hybrid duplexes and triplexes formed by homopolymers.

3. According to thermodynamic and kinetic properties, the hybrid complexes are closer to the double helix RNA than to the DNA.

IV. Complexes of nucleic acids with non-canonical base pairs

In the analysis of the literature data on complexes of NA containing non-canonical base pairs, attention has been concentrated on the following questions:

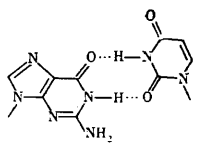
1. The mechanism of the pairing of the bases in the non-canonical pairs.

2. The influence of modification on the interplanar interactions of the bases and on the conformation of the sugar phosphate skeleton; the influence of the non-complementary pair on the secondary structure of the double helix.

3. The thermodynamic and kinetic parameters of the formation and dissociation of the modified oligomeric duplexes.

1. The dG.dT(G.U) pair

The possibility of the existence of the non-canonical G.U pair, stabilised by hydrogen bonds, was first postulated by Crick to account for the degeneracy of the genetic code.[53] He also put forward a mechanism of the formation of hydrogen bonds between the guanine and uracil residues:



G (*anti*) U (*anti*) .

Subsequently the study of various model systems, containing G.U(T) pairs, by modern experimental and theoretical methods confirmed the validity of the proposed mechanism.

Calculation of the conformations with the aid of AAPF and the construction of molecular models[23,54,55] showed that the bases of the dG.dT pair are in the usual tautomeric forms and have the *anti*-orientation relative to the carbohydrate rings. The inclusion of the dG.dT pair in the double helix does not reduce the interatomic contacts and induces only a small increase in the energy of the non-valence interactions and a slight deviation of the angles of the sugar phosphate skeleton from the standard values characteristic of the Watson–Crick helix.

An experimental study of the molecular structure of the dG.dT pair and of its influence on the conformation and the thermodynamic and kinetic parameters of the DNA double helix has been carried out by Patel and co-workers[56–60] in relation to the dodecameric duplex "GT":

```
  1 2  3  4 5 6 6′ 5′ 4′ 3′  2′ 1′
d(C G |T| G A A T T C |G| C G)
  G C |G| C T T A A G |T| G C)d
```

duplex "GT" .

The unmodified duplex "GC",[61] formed both in solution and in a crystalline state, having the helix in the *B*-form,[62,63] was used as a control:

```
  1 2 3 4 5 6 6′ 5′ 4′ 3′ 2′ 1′
d(C G C G A A T T C G C G)
 (G C G C T T A A G C G C)d
```

duplex "GC" .

The presence in the ^{1}H NMR spectrum of the duplex "GT" of two signals due to imino-protons corresponding to the dG.dT pair, which are located in a stronger field than the signals due to the dG.dC pair in the same position in the duplex "GC", indicates unambiguously the existence of two hydrogen bonds with carbonyl acceptor groups between the guanine and thymine residues. This confirms the structure of the dG.dT(G.U) pair presented above and rules out interactions with participation of rare tautomeric forms of the bases. The inclusion of the dG.dT pair leads to the breakdown of the interplanar interactions of the bases of this pair, in particular the thymine residue, with the neighbouring bases. According to ^{31}P and ^{1}H NMR data, the conformational perturbations introduced by the dG.dT pair are localised in the modifcation section and involve only the nearest nucleotide pair from the 3′-end.

The replacement of two dG.dC pairs in the duplex "GC" by the dG.dT pairs induces a decrease in the T_m of the duplex by 20 °C, but the enthalpy of the transition from the double helix state to the single helix state does not then change (see Table 4 below). Accoring to the authors, this is associated with a slight breakdown of the stacking interactions in the duplex "GT". The closeness of the ΔH^0 and the difference between the T_m for the duplexes "GC" suggest that the destabilising influence of the dG.dT pair is entropic in nature. It has been established on the basis of the difference between the values of $\Delta H^0_{v\text{-}H}$ and ΔH^0, determined from the heat absorption curves (SMC), that the cooperatively melting section, both in the duplex "GC" and in the duplex "GT", comprises 9 nucleotide pairs, i.e. it embraces the modification sections. According to ^{1}H NMR data, the end "fraying" effect, involving the nucleotide pairs in the 1- and 2-positions, is observed in the duplexes "GC" and "GT". The presence of the dG.dT pair increases this effect, as a result of which a broadening and disappearance of the signals due to the imino-protons of the dG.dC pairs (in the 1- and 2-positions) in the duplex "GT" takes place at lower temperatures than in the duplex "GC". Comparative analysis of the kinetic data for the exchange of the imino-protons of the duplexes "GT" and "GC" confirms this conclusion that the region corresponding to the spreading of the structural perturbations introduced by the dG.dT pair is limited only by the pair adjoining the non-canonical pair [at 35 °C, τ for the dG.dT pair and the neighbouring dG.dC pair (position 4) of the duplex "GT" are smaller by factors of 10 and 2 respectively than τ for the pairs in the same positions in the duplex "GC"]. The activation energy for the formation of the duplex "GT" has been determined by the temperature jump method[64] and proved to be much higher (22 kcal mol^{-1}) than the same parameter of the unmodified double helix.[65] On this basis, the authors suggested that the formation of the dG.dT and dG.dC pairs (in the 4-position) in the duplex "GT" can be accompanied by conformational adjustments requiring additional energy.

The influence of non-canonical pairs on the thermodynamic and kinetic parameters of the formation of the hexadecameric duplex has been investigated.[66] The hexadecanucleotide d(CAACTTGATATTAATA), which corresponds to the fragment of the bacteriophage ΦX174, was adopted as a basis. Together with the complementary sequence, it forms the protein binding section of the gene *A*. The structure of the initial complementary complex and the studied replacements of the canonical pairs by the dG.dT pair are presented below:

```
   1 2 3 4 5  6 7 8 9 10 11 12 13 14 15 16
d (C A A C T T G A T A T T A A T A)
  (G T T G A A C T A T A A T T A T) d
        T G          T
        G T          G
```

duplex "6TG" duplex "7GT" duplex "12TG" .

All the duplexes containing the dG.dT pair have T_m lower than that of the duplex "O" (see the duplexes 6–8 and 12, Table 4), the destabilising effect being greater for the substitution of the dG.dC pair by a non-complementary pair compared with the analogous substitution of the dA.dT pair and also when the substituted pair is located closer to the centre of the duplex. The dG.dT pair, substituting the dA.dT pair in the 12-position (the duplex "12TG"), indicates the smallest destabilisation. The smaller difference between the free energies of formation (ΔG) of the duplex "O" and the duplexes "6TG", "7GT", and "12TG" ($\Delta\Delta G \simeq 3$ kcal mol^{-1}, Table 4) indicates an insignificant disturbance of the stacking interactions of the bases by the dG.dT pair. The study of the kinetics of the formation and association of the duplexes "O" and "7GT" shows that the perturbations introduced by the G.T pair increase the rate constant for the dissociation of the duplex, while the formation rate constants remain almost the same. The destabilisation of the "7GT" duplex is apparently associated with the increase in the rate of dissociation.

Using the nOe technique, the A.T-rich duplex, containing the inner dG.dT pair, has also been investigated:[67]

d(AAATT|T̄|TCAAA)
(TTTAA|G̲|AGTTT)d .

Apart from determining the geometry of the dG.dT pairing, information was obtained in the above study about the degree of hydration of the small groove in the vicinity of the non-complementary pair.

A recent X-ray diffraction study of a synthetic dodecamer in the *B* form with two dG.dT pairs,

d(CGC|Ḡ|AATT|T̄|GCG)
(GCG|T̲|TTAA|G̲|CGC)d ,

confirmed that the structural changes induced by these non-complementary pairs are local in character and hardly disturb the overall conformation of the duplex.[68] Nevertheless, precisely this combination of bases (the dG.dT, dA.dC, and dA.dG pairs were compared) forms the most asymmetric pair (a characteristic feature of the Watson–Crick pairs is the presence of a twofold pseudoaxis located in the plane of the bases). This can be the cause of the effective repair of the dG.dT pair by the corresponding enzyme systems "recognising the deviations from the pseudosymmetry.[10,68] It is noteworthy that the dG.dT pair is also effectively discriminated by a restriction enzyme. Thus the replacement of the dG.dC pair by dG.dT or dG.dU in the recognition section of a series of restriction endonucleases leads to a sharp decrease and in some cases to a complete inhibition of the cleavage of the modified DNA duplexes.[69,70]

Two octameric duplexes having identical compositions but differing in the positions of two dG.dT pairs served as the object of study of the influence of the local environment of the dG.dT pairs on the conformation and stability of DNA in the *A* form:[12,71–73]

d(GG|Ḡ|GC|T̄|CC) d(GGG|Ḡ||T̄|CCC)
(CC|T̲|CG|G̲|GG)d (CCC|T̲|G̲|GGG)d .
I II

A characteristic property of the DNA duplexes containing GG and CC clusters is that in the crystalline state they form a helix in the *A* form.[74] It has been demonstrated by XRDA that the inclusion of the dG.dT pairs leads merely to a slight distortion of the sugar phosphate skeleton of the molecule and that the deformation generated is localised in the modification section and is limited to the nearest phosphodiester linkages. The stacking interactions in the region of the non-canonical pair are disturbed mainly by the decrease in the overlap of the bases of the dG.dT pair with those of the neighbouring pair from the 3′-end, while the "stacking" with the nearest "neighbour" in the 5′-position is retained in almost the same form as in the unmodified duplex. The thermodynamic stabilities of the crystal structures of the duplexes (I) and (II) differ appreciably—the crystals of the duplex (II) are stable at room temperature while those of the duplex (I) melt at 6 °C. According to the authors, one of the causes of the difference between the stabilities of the duplexes is the reduced contribution of the stacking interactions to the stabilisation of the structure of the duplex (I). In the duplex (I), the dG.dT pairs are separated by two dT.dC pairs and the stacking interactions are then disturbed in four out of the seven sites of the interplanar contacts between the base pairs, while in the duplex (II), where the dG.dT pairs are located next to each other, such perturbations occur at only three sites. The different destabilisations of the duplexes (I) and (II) can also be associated with the different states of the hydration shells. In particular, it has been shown that in the duplex (I) there are no bridging hydrogen bonds stabilising the dG.dT pairs. On the basis of these data, it is concluded that the consecutively arranged dG.dT pairs destabilise the double helix less than the isolated pairs.[12,71–73] The influence of the dG.dT pairs on the structure of DNA in the *Z* form has been investigated by the XRDA method in relation to the hexameric duplex[75]

d(C|Ḡ|CG|T̄|G)
(G|T̲|GC|G̲|C)d .

Despite the smaller flexibility of the sugar phosphate skeleton of *Z*-DNA compared with *B*-DNA, the non-complementary pair is incorporated in the *Z*-helix virtually without distorting it. It has been shown that the alternation of the conformations of the glycosidic linkage (*syn*-, *anti*-) and of the furanose ring (*N*- and *S*-) is virtually preserved in the nucleotide residues of the duplex, although in the dG.dT pair the deoxyribose of guanosine has the non-characteristic C(4′)-*endo*-conformation. The dG.dT pair, incorporated in the *Z*-helix, not only does not impair the interplanar interactions, but actually stabilises them to some extent, because a more complete overlap of the guanine and thymine residues with the bases of the neighbouring pairs is observed compared with the dG.dC pair located in the same position in the unmodified duplex. It has been established that the hydration shell of the duplex plays an important role in stabilising the dG.dC pair. Bridging hydrogen bonds are formed with participation of water molecules between N(2) and O(6) of guanine and between O(2) and O(4) of thymine respectively, compensating partly for the absence of such a bond in the dG.dT pair compared with the dG.dC pair.

It is noteworthy that the structure of the hydration shell in the vincinity of the non-complementary pair hardly depends on the form of DNA (*A*, *B*, or *Z*).[10,68]

We shall also note briefly certain studies in which the influence of the dG.dT pairs on oligomeric duplexes has been analysed. According to ^{1}H NMR data, the dinucleotides pd(GT) and pd(TG) do not form complexes in solution under conditions where the self-complementary dinucleotides are associated with formation of mini-duplexes.[76,77] On the other hand, the polymer poly[d(GT)] can exist in solution in the form of a double helix.[78–80] This has been demonstrated by UV, CD, and ^{1}H NMR spectroscopy, and also on the basis of the binding of ethidium bromide.

Data characterising the destabilising influence of the dG.dT pairs on oligomeric duplexes have been published.[50,81,82]

A series of studies have been devoted to the ribonucleotide complexes containing the G.U pair. According to XRDA data, the terminal G.U pairs in the tetrameric duplex

r(G GC U)
(U CG G)r

hardly disturb the stacking interactions of the bases.[83]

Table 3. Thermodynamic parameters of the varied nucleotide pair in the octadecameric DNA duplex.[127]

Varied base pair (X.Y)	T_m^* of duplex, °C	ΔH(X.Y), kJ mol^{-1}	ΔS(X.Y), kJ mol^{-1} K^{-1}	Type of structure of X.Y pair
G·C	69.6	70.6	0.208	Watson–Crick
C·G	68.1	68.3	0.208	"
A·T	67.7	67.8	0.208	"
T·A	64.9	55.5	0.208	"
G·T	60.1	30.6	0.116	weak
T·G	60.0	60.8	0.208	bound
A·G	59.3	59.1	0.208	bound
T·T	58.7	0	0.031	open-chain
C·T	58.6	0	0.031	open-chain
G·G	56.7	57.0	0.208	bound
G·A	56.6	30.6	0.128	weak
T·C	56.0	0	0.035	open-chain
A·A	55.6	55.6	0.208	bound
A·C	54.4	30.6	0.133	weak
C·A	53.2	54.1	0.208	bound
C·C	51.5	0	0.045	open-chain

* T_m was measured in 1 mM cacodylate buffer (pH 7.1) containing 0.5 M NaCl, 0.1 M EDTA, and nucleotide at a concentration of 0.3 A_{260}/ml.

The influence of the G.U pair and of its local environment on the stability of pentanucleotide double helices has been investigated.[84,85] The series of duplexes investigated are presented below:

r(C A G U G) (G U U A C)r duplex "5GU"	r(A G G C U) (U C U G A)r duplex "5*GU"
r(C A A U G) (G U U A C)r duplex "5AU"	r(A G A C U) (U C U G A)r duplex "5*AU"
r(C A G U G) (G U C A C)r duplex "5GC"	r(A G G C U) (U C C G A)r duplex "5*GC"
series I	series II .

In the series I duplexes, the central pair is flanked by the A.U pairs, while in the series II duplexes it is flanked by G.C pairs. It has been shown that all the series II complexes are more stable than the corresponding series I complexes. According to ^{1}H NMR data, the T_m of the G.U pair in the duplex "5GU" is lower by 4 °C than the T_m of the remaining pairs, i.e. the G.U pair is a region of local instability of the duplex, while in the duplex "5*GU" the G.U pair has the same T_m as the remaining pairs. The substitution of the central complementary pair by G.U in the series I duplexes reduces their stability the destabilisation being greater on replacing the G.C pair (the T_m of the duplexes "5GU", "5GC", and "5AU" are respectively 23, 38, and 28 °C). In the series II duplexes, the destabilising influence of the G.U pair is much smaller and is manifested solely on substitution of the G.C pair; the replacement of A.U by G.U hardly affects the stability of the double helix (the T_m of the duplexes "5*GU", "5*GC", and "5*AU" are respectively 44, 54, and 45 °C). According to the authors,[85] the different effects accompanying the inclusion of the G.U pair between the A.U or G.C pairs are associated with the fact that the interplanar interactions between the bases of the G.U pair and the pairs flanking it are disturbed to a greater extent in the duplex "5GU" than in the duplex "5*GU".

A comparative analysis of the thermodynamic parameters of the formation of oligomeric duplexes containing G.U pairs and of the unmodified duplexes containing A.U pairs in the same positions (see the duplexes 34–49 in Table 4) has been carried out[86] and it has been established that the substitution of the A.U pair by G.U reduces the stability of the complex, such decrease depending on its composition, nucleotide sequence, and the position of the G.U pair in the oligomer chain and its local environment. The destabilisation of the complexes induced by such substitution is mainly enthalpic in nature, although a decrease in the stability of duplexes containing adjoining G.U pairs is associated with the unfavourable alteration of the entropy of complex formation (Table 4).

The destabilisation effect of the replacement of the canonical pairs by the G.U pair has also been pointed out in a number of others studies.[5,87–89] However, the formation of the G.U pair is thermodynamically more favourable than the appearance of other non-canonical pairs under the same conditions. Thus the pentaribonucleotide AGGCU forms on self-association the duplex "GU" with two G.U pairs and not the duplex "GG" with a single G.G pair:[85]

r(A G GC U) (U CG G A)r duplex "GU"	r(A G G C U) (U C G G A)r duplex "GG"

A similar conclusion concerning the preferential formation of the G.U pairs before the G.G pairs was reached by Uhlenbeck et al.[90]

The study of the influence of the terminal G.U pairs on the thermal stability of RNA duplexes showed that the contribution of the free energy of the G.U pair, flanking the tetrameric GC-containing core, is approximately equal to the contribution of the terminal A.U pair and is smaller by ~1 kcal mol^{-1} than the contribution of the G.C pair.[91] Other terminal non-complementary pairs (A.A, A.C, A.G, and U.U) are less stable than the corresponding Watson–Crick pairs. Furthermore, it has been shown that the terminal G.U pair is more stable if it contains 5′-G and 3′-U but not conversely.[91]

The thermal stabilities and conformational characteristics of the polymeric complexes of the ribo- and deoxyribo-series, consisting of the G.U or dG.dT pairs, i.e. poly[r(GU)] and poly[d(GT)], have been compared on the basis of CD data.[79] It was established that the DNA duplex is formed at a higher ionic strength (0 °C, 0.1 M NaCl) than its ribo-analogue (0 °C, 0.02 M NaCl). In the NaCl concentration range from 0.1 to 1 M, the poly[d(GT)]$_2$ helix undergoes conformational rearrangements (within the limits of the *B*-form), which entails a smooth variation of the CD spectrum. The relative stabilities of the duplexes depend on the ionic strength of the buffer: the ribonucleotide duplex is more stable in 0.1 M NaCl, the T_m for poly[r(GU)] and poly[d(GT)] being 14 and 3 °C respectively, while the deoxy-analogue is more stable in 1 M NaCl (the T_m are 23 and 30 °C). However, these duplexes are much less stable than the duplexes formed by the self-complementary polymers—poly[r(AU)] and poly[d(AT)] (the T_m for the latter in 0.1 M NaCl are 65 and 57 °C respectively).

Table 4. Thermodynamic parameters of the formation of oligonucleotide complexes with non-canonical base pairs and unpaired nucleotide units.*

No.	Oligomeric complex	Conventional designation of complex	Complex formation conditions	T_m, °C	$-\Delta H^0$, kcal mol^{-1}	$-\Delta S^0$, e.u.	$-\Delta G^0$ (t, °C) kcal mol^{-1}	$\Delta\Delta G^{0****}$ (t, °C), kcal mol^{-1}	Refs.
1	d(CGCGAATTCGCG) (GCGCTTAAGCGC)d	«GC»	0.01 M phosphate buffer, pH 7.0, 0.1 mM EDTA, 0.1 M NaCl, c_0^{**} = 0.65 mM	71.3	102 (74)**	—	—	—	[61]
2	d(CG[T]GAATTC[G]CG) (GC[G]CTTAAG[T]GC)d	«GT»	ditto	51.3	106 (76)**	—	—	—	[56]
3	d(CG[A]GAATTC[G]CG) (GC[G]CTTAAC[A]GC)d	«GA»	»	55.0	—	—	—	—	[99]
4	d(CG[C]GAATTC[A]CG) (GC[A]CTTAAG[C]GC)d	«AC»	»	40.0	—	—	—	—	[92]
5	d(CGCAGAATTC–GCG) (GCG–CTTAAGACGC)d	«A»	»	52.0	104 (70)**	—	—	—	[129]
6	d(CAACTTGATATTAATA) (GTTGAACTATAATTAT)d	«O»	0.025 M phopsphate–1 mM cacodylate buffer, pH 7.0, 0.1 mM EDTA, 1 M Na^+, c_0 = 17.5 μM	52.0	102.0	289	15.8 (25°)	—	[66]
7	d(CAACT[T]GATATTAATA) (GTTGA[G]CTATAATTAT)d	«6TG»	ditto	45.1	92.6	230	13.2 (25°)	2.6 (25°)	[66]
8	d(CAACTTG[A]TATTAATA) (GTTGAA[T]TATAATTAT)d	«7GT»	»	42.9	91.3	264	12.7 (25°)	3.1 (25°)	[66]
9	d(CAACTTG[A]TATTAATA) (GTTGAAC[A]ATAATTAT)d	«8AA»	»	42.1	92.0	269	12.6 (25°)	3.2 (25°)	[66]
10	d(CAACTTG[A]TATTAATA) (GTTGAAC[C]ATAATTAT)d	«8AC»	»	40.6	90.9	265	12.1 (25°)	3.7 (25°)	[66]
11	d(CAACTTGA[T]ATTAATA) (GTTGAACT[C]TAATTAT)d	«9TC»	»	43.0	98.4	286	13.1 (25°)	2.7 (25°)	[66]
12	d(CAACTTGATAT[T]AATA) (GTTGAACTATA[G]TTAT)d	«12TG»	»	46.9	95.5	274	14.0 (25°)	1.8 (25°)	[66]
13	d(CAAAAAAG) (GTTTTTTC)d	—	0.01 M phosphate buffer, pH 7.0 0.1 mM EDTA, 1 M NaCl, c_0 = 0.2 mM	37	59	172	7.8 (25°)	—	[138]
14	d(CAAACAAAG) (GTTT–TTTC)d	«C»	ditto	22	53	161	4.9 (25°)	—	[138]
15	d(CCCAGGG) (GGGTCCC)d	—	0.01 M phosphate buffer, c_0 = 0.4 mM for self-complementary oligonucleotides and 0.1 mM for non-self-complementary oligonucleotides	48	63	177	10.0 (25°)	—	[103]
16	d(CCCGGG) (GGGCCC)d	—	ditto	43	62	176	9 (25°)	—	[103]
17	d(CCC[T]GGG) (GGG[T]CCC)d	—	»	31	60	179	6.7 (25°)	—	[103]
18	d(CCC[A]GGG) (GGG[A]CCC)d	—	»	27	60	180	5.9 (25°)	—	[103]
19	d(GATGGGCAG) (CTACCCGTC)d	«N»	0.01 M phosphate buffer, c_0 = 11.52 μM	41.53	68.04	191	—	—	[144]
20	d(GATGGG–CAG) (CTACCCCGTC)d	«C_N»	c_0 = 10.14 μm	29.21	60.39	174	—	—	[144]
21	d(GATGGG–CAG) (CTACCCAGTC)d	«A_N»	c_0 = 7.20 μM	23.15	51.07	146	—	—	[144]
22	d(CAAACAAAG) (GTTTGTTTC)d	—	0.01 M phosphate buffer, c_0 = 0.4 mM	48	64.5	183	10.1 (25°)	—	[110]
23	d(CAAAGAAAG) (GTTTCTTTC)d	—	ditto	50	62.8	179	9.5 (25°)	—	[110]
24	d(CAAAAAAAG) (GTTTTTTTC)d	—	»	45	68.0	196	9.6 (25°)	—	[110]
25	d(CAAATAAAG) (GTTTATTTC)d	—	»	41	58.6	168	8.5 (25°)	—	[110]
26	d(CAAA[T]AAAG) (GTTT[G]TTTC)d	—	»	31	55.6	165	6.5 (25°)	—	[110]
27	d(CAAA[G]AAAG) (GTTT[T]TTTC)d	—	»	27	46.7	137	5.8 (25°)	—	[110]
28	d(CAAA[G]AAAG) (GTTT[G]TTTC)d	—	0.01 phosphate buffer, c_0 = 0.4 mM	30	53.5	158	6.3 (25°)	—	[110]
29	d(CAAA[G]AAAG) (GTTT[A]TTTC)d	—	ditto	30	52.6	156	6.2 (25°)	—	[110]
30	d(CAAA[A]AAAG) (GTTT[G]TTTC)d	—	»	24	39.9	116	5.3 (25°)	—	[110]
31	d(CAAA[C]AAAG) (GTTT[T]TTTC)d	—	»	24	53.2	161	5.3 (25°)	—	[110]
32	d(CAAA[T]AAAG) (GTTT[C]TTTC)d	—	»	22	50.0	151	5.0 (25°)	—	[110]
33	d(CAAA[A]AAAG) (GTTT[A]TTTC)d	—	»	21	36.9	107	5.0 (25°)	—	[110]

Table 4 (continued)

No.	Oligomeric complex	Conventional designation of complex	Complex formation conditions	T_m, °C	$-\Delta H^0$, kcal mol^{-1}	$-\Delta S^0$, e.u.	$-\Delta G^0$ (t, °C) kcal mol^{-1}	$\Delta\Delta G^{0****}$ (t, °C) kcal mol^{-1}	Refs.
34	r(A U G C A U)p p(U A C G U A)r	—	0.01 M phosphate buffer, pH 7.0, 0.1 mM EDTA, 1 M NaCl, c_0 = 0.1 mM	30.2	38.4	108.4	—	—	[86]
35	r(A U G C G U)p p(U G C G U A)r	—	ditto	20.3	24.3	64.5	—	0.52 (37°)	[86]
36	r(A U G C G C A U)p p(U A C G C G U A)r	—	»	60.5	38.4	108.4	—		[86]
37	r(A U G C G C G U)p p(U G C G C G U A)r	—	»	58.7	54.4	145.8	—	0.88 (37°)	[86]
38	r(A U A C G U A U) (U A U G C A U A)r	—	»	42.0	55.6	158.1	—		[86]
39	r(A U G C G U A U)p p(U A U G C G U A)r	—	»	34.6	44.7	127.1	—	1.3 (37°)	[86]
40	r(A U G U A C A U)p p(U A C A U G U A)r	—	0.01 M phosphate buffer, pH 7.0, 0.1 mM EDTA, 1 M NaCl, c_0 = 0.1 mM	41.8	53.7	152.1	—	—	[86]
41	r(A U G U G C A U)p p(U A C G U G U A)r	—	ditto	39.7	54.2	155.0	—	0.32 (37°)	[86]
42	r(G C U A G C) (C G A U C G)r	—	»	49.3	63.0	177.1	—	—	[86]
43	r(G C U G G C) (C G G U C G)r	—	»	41.7	59.1	169.5	—	1.42 (37°)	[86]
44	r(G A C G U C) (C U G C A G)r	—	»	46.3	57.3	161.0	—	—	[86]
45	r(G G C G U C) (C U G C G G)r	—	»	29.6	37.7	106.1	—	2.58 (37°)	[86]
46	r(C U G C A G)p p(G A C G U C)r	—	»	45.4	54.5	153.0	—	—	[86]
47	r(C U G C G G) (G G C G U C)r	—	»	27.1	38.7	110.5	—	2.72 (37°)	[86]
48	r(C A G C U G)p p(G U C G A C)r	—	»	43.6	50.2	140.0	—	—	[86]
49	r(C G G C U G) (G U C G G C)r	—	»	35.8	44.3	125.0	—	1.08 (37°)	[86]

*The parameters of the formation of the unmodified complexes are quoted for comparison.
**The values of $\Delta H^0_{v\text{-}H}$ are given in brackets.
***c_0 is the concentration of oligonucleotides.
****$\Delta\Delta G^0$ (t, °C) is the difference between the values of ΔG^0(t, °C) for the fully complementary and defective complexes.

2. The A.C pair

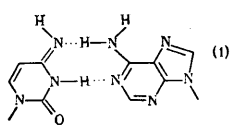
The structure of the dA.dC pair and its influence on the conformation and stability of the DNA duplex "AC"

1 2 3 4 5 6 6′ 5′ 4′ 3′ 2′ 1′
d(C G C G A A T T C A C G)
(G C A C T T A A G C G C)d

duplex "AC"

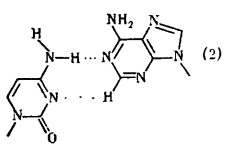
have been investigated.[92]

The theoretically possible ways in which the bases can be paired are illustrated below:

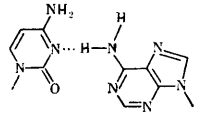

(1)

C (*anti*).A (*anti*)

;

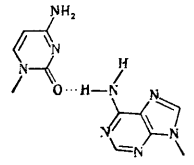

(2)

C (*anti*).A (*anti*)


(3)

C (*anti*).A(*anti*)

;

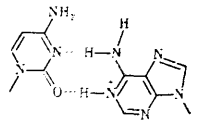

(4)

C (*anti*).A(*anti*)

.

The ^{1}H NMR spectrum of the duplex "AC" contains no signals due to the exchanging protons of the dA.dC pair and Scheme (1), postulating the involvement of the cytosine enolic form in the formation of the hydrogen bond, does not therefore reflect the real structure. The absence of an nOe between the H(2) atom of the adenine in the dA.dC pair and the imino-protons flanking the defect in the dG.dC pairs rules out the orientation of the bases illustrated in Scheme (2). The authors suggest that the dA.dC pair is stabilised by one hydrogen bond between the N(6)-imino proton of adenine and the N(3) atom [Scheme (3)] or the O(2) atom [Scheme (4)] of

cytosine. However, one cannot rule out the possibility of the protonation of the N(1) atom of adenine, which leads to the formation of a second hydrogen bond between the $N^+(1)$–H group of adenine and the O(2) atom of cytosine [Scheme (5)]:



C(*anti*).A^+(*anti*)

The location of the H(6) and H(5) signals of the cytosine in the dG.dC pair adjoining the non-canonical pair (4-position) of the duplex "AC" in a stronger field compared with the signal of the corresponding protons in the duplex "GC" (see above) is evidence of the disturbance of the stacking interactions in the region of the non-canonical pair. It has been suggested that one or both bases of the dA.dC pair are outside the helix. Significant distortions of the sugar phosphate skeleton of the duplex "AC" around the dA.dC pair have been observed by ^{31}P NMR.

The replacement of two dG.dC pairs in the duplex "GC" by the dA.dC pairs reduces the T_m of the duplex by 30 °C (Table 4). According to ^{1}H NMR data, the presence of the dA.dC pair in the third position in the duplex "AC" results in the "fraying" of the dG.dC pairs in the first and second positions of this duplex at lower temperatures than in the duplex "GC". The study of the kinetics of the exchange of the imino-protons of the duplex "AC" with water established τ for the dG.dC pair in the fourth position and for the dA.dT pairs in the fifth and sixth positions are smaller by factors of 3 and 2 respectively than τ for the analogous pairs of the duplex "GC". Consequently, the destabilising influence of the dA.dC pairs is not restricted to the neighbouring pair but involves the entire molecules of the duplex "AC".

However, different conclusions were reached concerning the character of the influence of the non-complementary dA.dC pair on the local and overall DNA structure after a deeper study of the structure of the duplex "AC" and its isomer

d(CG|A|GAATTC|C|CG)
(GC|C|CTTAAG|A|GC)d

by ^{1}H and ^{31}P NMR spectroscopy and after the estimation of the interatomic distances from nOe data.[93] Thus it is clear that the A and C residues are incorporated in the double helix, participating in stacking interactions with the neighbouring bases and inducing only local conformational rearrangements which do not extend to the central AATT segment. The results obtained do not permit a choice between the two A.C pairing schemes (3) and (5).[93]

This interpretation agrees well with the conclusions reached in a study[94] where the defective duplex

d(CGC|C|GC|A|GC)
(CG|A|CG|C|CGC)d

containing two dA.dC pairs was investigated by NMR spectroscopy and studies[95,96] devoted to the XRDA of the duplex "CA" whose nucleotide sequence is similar to that of the duplex "AC":

d(CGC|A|AATT|C|GCG)
(GCG|C|TTAA|A|CGC)d
duplex "CA".

It has been established that, when the dA.dC pair is integrated in the DNA, the stacking interaction with the bases of the neighbouring pairs and the conformation of the sugar phosphate skeleton are only insignificantly disturbed. According to XRDA data, Scheme (5), involving the protonation of adenine, is preferred for the A.C pairing; the important role of bridging hydrogen bonds, with participation of the water molecules incorporated in the hydration shell of the duplex, in the stabilisation of the dA.dC pair has been emphasised.

A comparative analysis of the formation of complexes by 11-, 14-, and 17-unit oligonucleotides with the DNA ΦX174 and a mutant DNA in which a G–A point substitution had been carried out, has been performed.[97] The modified complexes investigated are indicated below:

d(AACA|C|CCTATG)
(...CTTTGT|A|GGATACCC...)d
complex "11AC"

d(AACA|C|CCTATGGGA)
(...CTTTGT|A|GGATACCCTCG...)d
complex "14AC"

d(AACA|C|CCTATGGGAGCG)
(...CTTTGT|A|GGATACCCTCGCTTT...)d
complex "17AC".

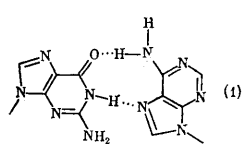
(The complexes with the non-mutant DNA containing the complementary dG.dC pair are designated respectively by "11GC", "14GC", and "17GC".) It was demonstrated that the degree of binding of all three oligonucleotides to the mutant DNA is much weaker than to the natural DNA. Furthermore, the duplexes containing the dA.dC pair are much less stable than their complementary analogues. The replacement of one dG.dC pair by dA.dC is equivalent as regards destabilisation to the removal of three nucleotide pairs (the temperatures of the half-dissociation of the complexes "17AC" and "14GC" and also "14AC" and "11GC" are 43 and 41 °C and 32 and 33 °C respectively).

The destabilising influence of the dA.dC pairs on the DNA duplexes has also been noted in other studies.[66,81,98]

The replacement of dG.dC pair by dA.dC in the recognition section of restriction endonucleases SmaI, KpnI leads to the complete inhibition of these enzymes.[70]

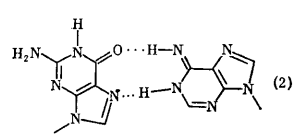
3. The purine – purine pairs

The following schemes for the formation of hydrogen bonds in the dG.dA pairs have been analysed in the literature:



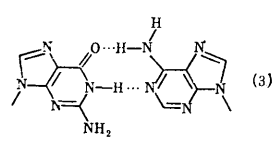
(1)

G (*anti*).A (*syn*)



(2)

G (*syn*).A (*anti*)



(3)

G (*anti*).A (*anti*).

Theoretical analysis of the conformations of the DNA duplexes showed that the structures (1) and (3) correspond to minimum energies of non-valence interactions.[54,55]

The relative disposition of the bases in the dG.dA pair has been studied experimentally in three model systems. The structure of the dodecameric duplex "GA" has been investigated by NMR spectros-

copy in solution:[99]

1 2 3 4 5 6 6′ 5′ 4′ 3′ 2′ 1′
d(C G | Ā | G A A T T C | Ḡ | C G)
(G C | G | C T T A A G | A | G C)d
duplex "GA".

The nOe between the H(2) atom of the adenine in the dG.dA pair and the imino-protons of the neighbouring dG.dC pairs indicates the *anti*-conformation of adenosine [Scheme (2) and (3)]. The absence of an nOe between the H(2) and H(8) atoms of adenine and guanine respectively rules out the relative disposition of the bases illustrated in Scheme (2). On the basis of these data, the authors conclude that Scheme (3) actually obtains. A similar conclusion concerning the relative orientations of the bases in the dG.dA pair was reached in a study[100] where a decameric duplex containing two adjoining dG.dA pairs was investigated by ^{1}H NMR:

d(C C A A | Ḡ | | Ā | T T G G)
(G G T T | A | | G | A A C C)d.

The same duplex has been investigated by XRDA.[101] It was found that in the crystalline stage too both nucleotide residues of the non-canonical pair are in the *anti*-conformation, the bases being retained by two hydrogen bonds [Scheme (3)]. A distinct propeller disposition of the bases (the bases are rotated relative to one another by 24.8°) is characteristic of the dG.dA pair in this conformation, which facilitates pairing. Despite the fact that the distance between the C(1′) atoms of adenine and guanine is greater by 2 Å than in the Watson–Crick pair, the block consisting of two dG.dA pairs is fairly effectively incorporated in the double helix. The small changes in the individual torsional angles of the sugar phosphate skeleton observed under these conditions are within the limits characteristic of canonical pairs.

A different geometry of the dG.dA pairing with the *syn*-conformation of adenine [Scheme (1)] arises in the dodecameric duplex "AG", whose primary structure is similar to that of the duplex "GA":[13,102]

d(C G C | Ḡ | A A T T | Ā | G C G)
(G C G | A | T T A A | G | C G C)d
duplex "AG".

This investigation was also carried out by XRDA. It is significant that the incorporation in *B*-DNA of the pair dG.dA in which the nucleotide residues have respectively the *anti*- and *syn*-conformations is accompanied by a minimal adjustment of the parameters of the sugar phosphate skeleton. It has been postulated that the realisation of a particular pairing scheme [(1) or (3)] may depend on local conditions, because, according to theoretical calculations, the energy difference between these two states of the dG.dA pair is small.[101]

It has been established that the dG.dA pair, both in the state (1) and in the state (3), is incorporated in the double helix virtually without disturbing the interplanar interactions.[13,99–102]

The destabilising influence of the dG.dA pair is shown by the decrease in the T_m of the duplex "GA" by 17 °C compared with the duplex "GC" (Table 4). Data for the exchange of the imino-protons of the duplex "GA" with water have shown that the perturbations introduced by the dG.dA pair are local in character and involve only the neighbouring dG.dC pair in the fourth position.[99]

Experimental data for the molecular structure of the dA.dA pair have been obtained recently in a NMR study of the DNA duplex:[103]

d(C C C | Ā | G G G)
(G G G | A | C C C)d.

It has been shown that stable hydrogen bonds are not formed between the bases of the non-complementary pair. The incorporation of the dA.dA pair in the double helix is achieved via the stacking interaction of the adenine bases with the neighbouring bases and is accompanied by the displacement of the adenine residues from the plane in order to reduce the overlap of the opposite imino-groups and also by an appreciable distortion of the sugar phosphate skeleton.

The structure of the G.G pair in the NA duplexes has not been investigated by experimental methods.

Purine–purine pairs induce a marked destabilisation of the double helices (Table 4). Comparison of the thermal stabilities of the oligomer-polymeric complexes $dG_n.dC_{12}A_mC_x$ and $dG_n.dC_{10}G_mC_x$ (n = 100, m = 0–6, x = 5–41) having different numbers of consecutively arranged dG.dA or dG.dG pairs shows[104] that the destabilisation of the duplexes is proportional to the number of non-canonical pairs. The inclusion of one dG.dA or dG.dG pair induces a decrease in the T_m complex corresponding to the removal of on average three dG.dC pairs. However, the presence of the dG.dA or dG.dG pairs in the central section of the complex does not disturb the cooperative nature of fusion. Such non-canonical pairs are apparently incorporated in the regular double helix and participate in stacking interactions both with one another and with neighbouring dG.dC pairs.

The influence of the local environment on the destabilising effect of the A.A and G.G pairs has been demonstrated.[84,85] According to ^{1}H NMR spectroscopic data, pentaribonucleotide complexes of the type

r(CA | X̄ | UG)
(GU | X | AC)r

with central A.A and G.G pairs flanked by A.U pairs are not formed. On the other hand, the relatively stable duplex "5*AA" (T_m = 26 °C), in which the A.C pair is surrounded by G.C pairs, is formed:

r(A G | Ā | C U)
(U C | A | G A)r
duplex "5*AA".

4. Pyrimidine–pyrimidine pairs

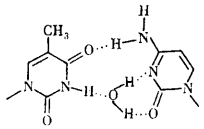
Conformational analysis of the DNA duplexes using AAPF showed that the formation of a dimer stabilised by a hydrogen bond is possible for each pyrimidine–pyrimidine pair.[23,54,55]

The molecular structure of the pyrimidine–pyrimidine pair has been little investigated by experimental methods. The results of a NMR spectroscopic study of the dodecameric duplex "CT"[105]

1 2 3 4 5 6 6′ 5′ 4′ 3′ 2′ 1′
d(C G | C̄ | G A A T T C | T̄ | C G)
(G C | T | C T T A A G | C | G C)d
duplex "CT"

established that the dC.dT pair is retained by two hydrogen bonds: a full hydrogen bond between the NH_2 group of cytosine and the O(4) atom of thymine and a weak hydrogen bond between the

N(3)–H group of thymine and the electron-accepting groups of cytosine with participation of a water molecule:



T (*anti*).C (*anti*).

The replacement of two dG.dC pairs in the duplex "GC" (see above) by the dC.dT pairs reduces T_m by 30 °C (Table 4), the destabilising influence of the dC.dT pair involving all the inner base pairs of the duplex "CT". According to data for the rate of exchange of the imino-protons of the duplex "CT" with water, the lifetimes of the dG.dC (in the 4-position) and dA.dT (in the 5- and 6-positions) pairs are respectively smaller by a factor of 10 and 6 than the values of τ for the analogous pairs in the duplex "GC".

The destabilisation of the oligomeric complexes induced by the replacement of dA.dT by the dC.dT pair is also indicated by the results of other studies[50,66] (see the duplexes 6 and 11 in Table 4).

The existence of U.U pairs stabilised by a hydrogen bond has not been demonstrated experimentally. UV-photohydration and photodimerisation data for the polymeric complexes poly(U).poly(A,U) have led to the conclusion[106] that the uridine residues in poly(A.U) are outside the helix and do not participate in the stabilisation of the secondary structure. It has been demonstrated on molecular models that the formation of such inner loops does not induce appreciable structural changes in the double helix.[107] The increase in the content of uridine residues in the copolymer poly(A.U) diminishes the stability of the complex poly(U).poly(A,U),[88,108,109] while the polynucleotide poly(A,U), containing 49 mol% of uridine does not enter at all into complementation interactions with polyuridylic acid.[107]

The smallest, compared with other nucleotides, capacity of uridine for stacking interactions, which play an important role in the stabilisation of double-stranded structures, is the reason why the pentanucleotide r(AGUCU) does not form a duplex in solution, in contrast to the pentamers containing A, G, or C in the third position.[85]

The geometry of the dT.dT pair has been discussed.[110] According to ^{1}H NMR spectroscopic data, this pair is not stabilised by hydrogen bonds,[103,111] although it is effectively incorporated in the double helix with minimum distortions of the conformations of the bases and of the sugar phosphate skeleton.[103,111,112] The study of the characteristic features of the molecular structure of the dT.dT pair in the duplex

d(CCC|T|GGG)
(GGG|T|CCC)d

showed that equilibrium between two or more conformations is characteristic of thyine residues in the temperature range 10 to 20 °C.[110]

The presence of the dT.dT pair in the duplex "TT"

d(ATCCTA|T|TAGGAT)
(TAGGAT|T|ATCCTA)d

duplex "TT".

reduces its T_m by 11 °C compared with the dodecameric duplex

d(ATCCTATAGGAT)
(TAGGATATCCTA)d

Furthermore the dT.dT pair forms a local unstable section in which the double helix dissociation process is initiated on raising the temperature simultaneously with fusion, which begins at the ends. A similar conclusion was reached[113] in a study of the duplex

d(CGCGA|T|TCGCG)
(GCGCT|T|AGCGC)d .

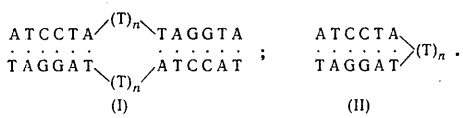
The dependence of the stability of the non-complementary dT.dT pair on the nature of the neighbouring pairs is illustrated by the following data.[103,110] The dT.dT pair has the greatest destabilising effect when it replaces the central dA.dT pair in the

—ATA—
—TAT—

sequence and has the smallest effect on replacement of the dA.dT pair in the sequence

—CAG—.
—GTC—

It has been shown in a series of studies[114,115] by NMR and temperature jump methods that the oligonucleotides d[ATCCTA(T)$_n$.TAGGAT] (n = 0–7), the primary structure of which contains palindrome elements, can form in solution both defective double helices (I) and monomolecular "pins" (II):



(I) (II)

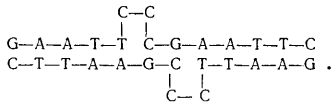
Oligomers containing more than three non-complementary thymine units form "pins" exclusively, while the coexistence of both forms is characteristic of oligonucleotides with $n \leqslant 3$. A decrease in the nucleotide concentration and in the ionic strength of the solution promotes the preferred formation of "pins". According to ^{1}H NMR spectroscopic data, the stabilisation of loops in the "pin" structures containing less than three paired thymine residues is achieved as a result of the decomposition of the nucleotide pairs adjoining the oligo(dT) sequence. The most stable structures are those with loops containing 4–5 nucleotide units. The enthalpies of dissociation of the base pairs in the double-stranded section of the molecule show that the pairs closest to the loop are more stable than the terminal pairs, i.e. the loop stabilises to some extent the double helix fragments. By measuring the nOe, it has been established that the stabilisation is achieved as a result of the "stacking" of the first (from the 5′-end) unpaired thymine residue and the base of the neighbouring nucleotide pair.

The formation of "pin" structures analogous to (II) has also been detected for the oligonucleotides d[CGCGCG(T)$_n$CGCGCG] (n = 4, 5)[116,117] and d(CGCGATTCGCG).[113]

Experimental evidence for the formation of hydrogen bonds between the bases of the C.C pair has not so far been obtained. According to many workers, the absence of hydrogen bonds from the C.C pair and also the incomplete "stacking" of the unpaired cytosine and the bases of the neighbouring nucleotide pairs lead to both cytidine residues "sticking out" from the helix, i.e. to the formation of an inner loop. The destabilising effect of the C.C pair is due to the same factors.[84,85,90,118]

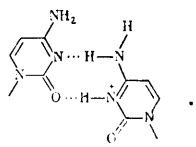
It has been established in the study of the formation of complexes by the oligonucleotides ($A_6C_mU_6$) (m = 1–8)[20,119] that the oligomers with m = 1–3 form a bimolecular helix with an inner loop and that, when m = 4–8, the "pin" structure becomes thermodynamically

more favourable. Among the "pins" formed by oligoribonucleotides, the one whose loop consists of six nucleotide units has the maximum stability. It has been established from ^{1}H NMR and nOe data[120] that the oligonucleotide d(GAATTCCCGAATTC) forms an unusual bimolecular structure containing non-coinciding loops:



The protonation of one of the cytosine residues of the C.C pair promotes the formation of hydrogen bonds between the bases. Thus, according to CD data, the oligomer d($C_4A_4T_4C_4$) forms a duplex with terminal dC.dC$^+$ pairs.[121] The presence of the central d(A_4T_4) section promotes the formation of dC.dC$^+$ pairs between antiparallel chains (in earlier studies[122,123] it was postulated that the C.C$^+$ pairs can arise only between parallel strands) and that this process occurs at higher pH than, for example, the association of dC_8 with d$(C^+)_8$.

The following structure has been proposed for the dC.dC$^+$ pair, which can be incorporated, according to the authors,[81] in the double helix of DA without causing significant changes in the geometrical parameters:



C (*anti*).C$^+$ (*anti*)

It has been suggested that the dC.dC$^+$ pairs are also formed on self-association of poly[d(CT)] and poly[d(AC)].[124]

V. Comparative analysis of the influence of the non-canonical pairs on the structure and energetic parameters of the formation of the oligomeric complexes

(1) The pyrimidine–purine non-complementary pairs as well as the dG.dA pair, incorporated in nucleotide double helices, are retained by two hydrogen bonds. The dA.dA pair is not stabilised by hydrogen bonds. There is no rigorous experimental evidence concerning the molecular structure of the dG.dG pair and the pyrimidine–pyrimidine pairs. The bases in the non-complementary pairs are in the usual tautomeric forms and as a rule in the *anti*-conformation [the dG(*anti*).dA(*syn*) pair is an exception].

(2) The replacement of the A.T and G.C pairs by non-canonical pairs alters the conformation of the sugar phosphate skeleton.

(3) Depending on the nature of the heterocyclic bases in the non-canonical pairs and on their immediate environment, the bases of such pairs can be incorporated in the double helix, participating in the stacking interaction, or can be displaced from the helix, forming a loop.

(4) The smallest perturbations of a local character are induced by the G.U, dG.dT, and dG.dA pairs, which coexist with the complementary pairs within the double helix.

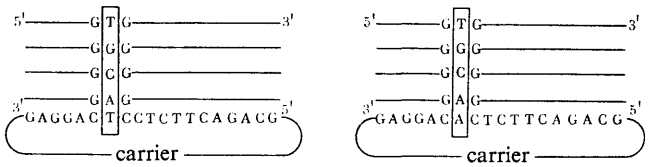
(5) The dA.dA and dC.dT pairs exert the most extensive perturbing influence. The bases of these pairs are arranged, at least partially, outside the double helix.

(6) The replacement of the Watson–Crick pairs by non-canonical pairs reduces the stability of the duplex. The degree of destabilisation depends on the length of the modified duplex, the nature of the non-canonical and replaced (G.C or A.T) pairs, the relative number of "erroneous" pairs and their positions in the duplex, and the local environment, as happens, for example, in the series of DNA duplexes[110]

d(CAAA|X|AAAG)
(GTTT|X′|TTTC)d

X, X′ = A, C, G,T .

The study of their melting profiles yielded the following stability series (the X.X′ pairs are indicated): dG.dT > dG.dG > dG.dA > dC.dT > dA.dA = dT.dT > dA.dC = dC.dC. Computer analysis of 100 t-RNA made it possible to estimate the relative stabilities of the non-complementary pairs in DNA: G.U ≃ G.G > C.C, U.U, C.A, A.A > G.A, U.C.[125] Measurement of the rates of dissociation of a series of oligomer–polymer DNA duplexes fixed on a dry agarose gel, i.e.



led to the derivation of the following series based on decreasing thermal stabilities as a function of the nature of the varied nucleotide pair: dT.dA ≃ dA.dT > dG.dT ⩾ dG.dA > dA.dA ≃ dT.dT ≃ dC.dA > dC.dT.[126] The thermodynamic parameters of the formation of oligomeric complexes with non-canonical base pairs are summarised in Table 4.

(7) In the general case, the destabilisation of the modified duplex compared with the "regular" duplex is associated with a decrease in the enthalpy of complex formation (as a result of the disturbance of the stacking interactions in the region of the non-canonical pair) overlapping the favourable change in entropy (as a result of the greater conformational mobility of the "erroneous" pairs).

The comparative characteristics of the influence of all the possible combinations of non-canonical pairs on the thermodynamic parameters of the complex formation process have been obtained[127] in relation to the DNA duplex

d(CGTCGTTT|X|ACAACGTCG)
(GCAGCAAA|Y|TGTTGCAGC)d

X, Y = A, C, G, T .

Only the components of ΔH^0 (X.Y) and ΔS^0 (X.Y) (calculated per nucleotide pair) characterising the stacking interactions of the non-canonical pair with neighbouring pairs (the "virtual stacking" approximation) were analysed (Table 3).

The data obtained led to a hypothesis concerning the structural features of the non-complementary pairs. Three main types of conformations of such pairs have been considered.

(a) Bound pairs (the so called "wobble" pairs). They are characterised by the same value of ΔS and somewhat reduced values of ΔH compared with the Watson–Crick pairs, they have paired bases, and are incorporated in a double helix.

(b) Open chain pairs. They are characterised by low values of ΔH and ΔS; such pairs form loops.

(c) Weak nucleotide pairs. They have intermediate (between extremum values) thermodynamic parameters and form intermediate

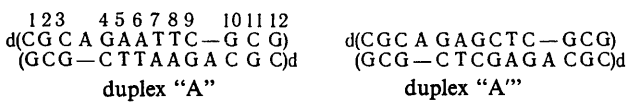
structures of different types. A direct correspondence has not been found between the structural features of the pairs and their destabilising influence. It is of interest that the T_m of duplexes with the non-canonical pairs X.Y and Y.X, differing only in the relative disposition of the bases in the complementary chains, are not the same although the environment of these pairs and hence the interplanar contacts are not altered. This is apparently associated with the effect of the long-range interactions of the non-canonical pairs.

Consideration of the formation and reparation of the non-complementary pairs has led to the conclusion that the reparation systems eliminate with the greatest efficiency the most frequently encountered pairs. Thus the repair system of *E. coli* effectively "mends" the dT.dG, dG.dG, and then dC.dA pairs and to a much lesser extent the dG.dA, dA.dG, dT.dT, dC.dC, dT.dC, and dC.dT pairs.[128] It has been suggested that the recognition of the "erroneous" pairs by reparation enzymes is not associated with the degree of destabilisation but is determined by the orientation of the bases relative to the sugar phosphate residue and the profile of the functional groups in the grooves within the double helix.[10,127]

VI. Complexes of nucleic acids containing unpaired nucleotide residues

1. Inner unpaired units

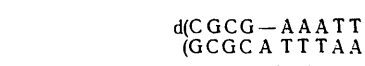
Analysis of the structure of the DNA duplexes "A" and "A'",[14,57–59,64,129,130] whose composition includes adenosine residues without a partner on the complementary chain, i.e.



showed that the unpaired purine base is incorporated in the helix and particpates in the stacking interaction with the neighbouring dG.dG pairs. However, the inclusion of a "superfluous" unit is acompanied by an increase in the distance between the G.C pairs and hence leads to the disturbance of the "stacking" of the bases of the unmodified strand of the duplex. Furthermore, a distortion of the sugar phosphate skeleton as well as a change in the length and torsional angles of the P–O bonds of the phosphate section opposite the "superfluous" adenosine have been observed in the modification section (^{1}H NMR spectroscopic data).

The presence of an unpaired residue greatly destabilises the duplex (Table 4). The similarity of the values of ΔH^0 for the duplexes "GC" (see above) and "A" [the disturbance of the stacking interactions between the G(3) and C(4) bases on the unmodified strand is compensated by the additional interplanar interactions of the "superfluous" adenosine] indicates the entropic nature of the destabilising effect. On the basis of the difference between $\Delta H_{v\text{-}H}$ and ΔH^0, determined by the SMC method for the duplex "A", it has been established that the cooperatively melting section comprises 9 ± 1 nucleotide pairs, i.e. includes the "superfluous" unit. According to data for the rate of exchange of the imino-protons of the duplex "A" with water, the destabilising influence of the unpaired adenosine involves all the inner pairs of the duplex.

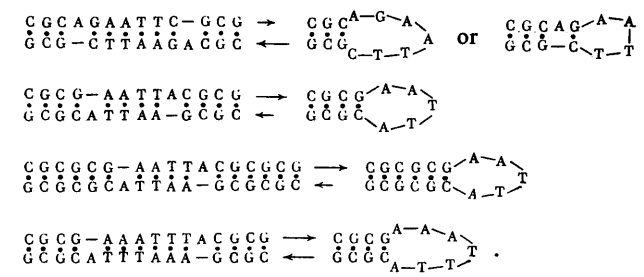
An NMR spectroscopic study of a longer sequence containing the unpaired adenosine residue, i.e.



also showed that the "superfluous" base is integrated in the double helix.[131] However, this entails not an increase in the distance between the bases on the opposite strand, as in the case of the duplex "A", but some bending of the helix. Since the defects are arranged symmetrically on different strands of the duplex, the helix acquires an S shape. The bases of the unmodified strand opposite the unpaired residue are then in fairly effective contacts.

It is of interest that in the crystalline state the conformation with the disposition of the unpaired adenosine outside the helix is characteristic of both duplexes—"A"[132] and "A'".[133,134] Despite the palindrome sequence, both "superfluous" bases are oriented in different ways (probably owing to the influence of the crystal packing). Furthermore, the local structure in the region of the defect is different for the "A" and "A'" duplexes. Thus in solution the unpaired adenosine residue is incorporated in the double helix, while in the crystalline state it "sticks out". The data obtained indicate the conformation flexibility of DNA duplexes modified in this way and their ability to pass from one conformational state to another when the external conditions are altered.

The study of the conformational transitions in a series of imperfect palindromes with inserts comprising superfluous adenosine residues has shown[135,136] that the distance between the insert and the "central symmetry" determines the equilibrium between the duplex and the single-stranded "pin". The theoretically possible conformations of the oligodeoxyribonucleotides investigated are illustrated in the scheme below:



It has been established on the basis of UV and NMR spectroscopic and SMC data that the introduction of additional adenosine units does not induce a change in the enthalpy of complex formation regardless of their positions in the sequence.

The DNA duplex[137]



served as a model of a double helix with a "superfluous" guanosine residue.[137] A three-dimensional model of the duplex "G" has been constructed from ^{1}H NMR spectroscopic data and by means of computer modelling (incremental approach). The series of structures differing in the length of the double helix sequence flanking the defect proved to be very similar. The duplex "G" is a double helix belonging to the $\overline{B}$-family; the axis of the helix is bent by 18–23°. As was to be expected, the unpaired residue is incorporated in the double helix.

The structure and stability of the DNA duplex "C", containing a superfluous cytidine residue, have been investigated:[138]

1 2 3 4 5 6 7 8
d(C A A A C A A A G)
(G T T T — T T T C)d
duplex "C".

The occurrence of an nOe duplex "C" between the imino-protons of the dA.dT pairs (in the 4- and 5-positions) and also the absence of an upfield shift of the signls of the non-exchanging protons of the unpaired cytosine as the temperature reduced indicate that the "superfluous" cytosine residue "sticks out" from the helix and does not participate in the stacking interaction with the bases of neighbouring pairs. Cytosine is apparently "expelled" by the neighbouring purine bases, which are known to possess an enhanced capacity for stacking interactions. The position of cytosine outside the helix destabilises the duplex (the duplexes 13 and 14 in Table 4 and in the paper of Aboul-Ela et al.[110]), the destabilising effect being enthalpic in nature in this instance.

The question arises whether the unpaired cytosine always occupies a position outside the helix. It has been found that its conformation depends on temperature. Thus a study by two-dimensional NMR spectroscopy and the analysis of nOe parameters of the DNA duplex

d(C C G C G A A T T C — C G G)
(G G C — C T T A A G C G C C)d

showed[139] that at low temperatures (0 °C) the "superfluous" cytosine is outside the helix, while the adjoining bases participate in the stacking interaction. With increase in temperature, the conformational equilibrium shifts towards the formation of structures in which the cytosine residue is built into the duplex (> 40 °C). It is noteworthy that the disturbances in the sugar phosphate skeleton extend to a region remote from the site of the defect.

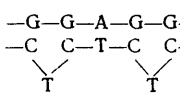
Even more interesting results have been obtained in a study of the conformational characteristics of other unpaired pyrimidine bases—thymine and uracil—in the duplexes "T"[140] and "U":[141]

d(C T G G T G C G C)	r(C U G G U G C G C)
(G A C C — C G C G)d	(G A C C — C G C G)r
duplex "T"	duplex "U" .

These duplexes, which have a similar sequence, differ only in the nucleotide residues belonging to the deoxyribo- or ribo-series. The set of NMR spectroscopic data, the nOe parameters, the temperature dependence of the chemical shifts of the proton signals, etc. show that the unpaired thymine residues are built into the duplex "T". According to the results of computer modelling, the insertion of the "superfluous" thymine does not induce appreciable changes in the structure of *B*-DNA under these conditions: the base pairing is not disturbed and there are no appreciable distortions in the torsional angles. We may note that, when the more bulky adenine residue is intercalated, the structural perturbations are much more pronounced.

In contrast to thymine residues, the uracil in the duplex "U" is preferentially located outside the helix. The regular structure of *A*-RNA is then hardly disturbed. The authors suggest that the difference in the behaviour of these bases in the DNA and RNA duplexes can be accounted for by the different tendencies of thymine and uracil to participate in stacking interactions.

If the unpaired thymine residues are not unique but are encountered regularly and frequently in the duplex, then a position outside the helix is more characteristic of them. Such a conformation has been postulated for the polymeric duplex poly[d(GGAA)].poly[d(TC)][142] from data for the formation of cytosine photodimers:



It has been established that the "superfluous" thymine residues do not disturb the cooperative nature of the fusion but diminish greatly the stability of the defective helix compared with the regular poly[d(GGA)].poly[d(TCC)].

A similar conclusion that the unpaired thymine residues are located outside the helix has been reached on the basis of a study of the duplex poly[d(GA)].poly[d(TTC)].[143]

Interesting data for the stability and dynamics of a series of duplexes with a nucleotide sequence corresponding to the mutation "hot spot" with a shift of the frame have been obtained by NMR spectroscopy:[144]

d(G A T G G G C A G)	d(G A T G G G — C A G)	d(G A T G G G — C A G)
(C T A C C C G T C)d	(C T A C C C C G T C)d	(C T A C C C A G T C)d .
duplex "N"	duplex "C_N"	duplex "A_N"

It was found that the unpaired adenine exerts a mainly local influence on the stability of the nucleotide pairs (the duplex "A_N"), while the delocalisation of the destabilising effect along the G.C track of the duplex "C_N" as a result of the migration of the defect is characteristic of the unpaired cytosine:

d(G A T G G G — C A G) → d(G A T G G — G C A G) →
(C T A C C C C G T C)d ← (C T A C C C C G T C)d ←
→ d(G A T G — G G C A G) → d(G A T — G G G C A G)
← (C T A C C C C G T C)d ← (C T A C C C C G T C)d
duplex "C_N" .

The greater thermal stability of the duplex "C_N" compared with the duplex "A_N" (ΔT_m = 5.7 °C) has been explained by the authors in terms of the positional degeneracy of the defect (unpaired cytosine) and the associated increase in entropy. It is noteworthy that the results of the above study[144] conflict with the results and conclusions of a number of other studies[57,64,129] as regards a series of aspects associated with the character of the influence of the "superfluous" adenosine on the stability of the duplex pairs and the enthalpy of complex formation.

In the next study of the above series, the conformational dynamics of the "superfluous" guanine residue, located in the G.C track of the analogous duplex, was investigated.[145] The presence of free imino-protons of the unpaired guanine made it possible to follow the migration of the defect. It was discovered that the "superfluous" unit occupies all possible positions, although it occurs more frequently at the ends of the track. One cannot say with certainty that the rate of migration of the defect is the same for oligopurine and oligopyrimidine sequences. As regards the location of the "superfluous" base relative to the duplex, experimental data show that the unpaired purine bases are incorporated in the double helix, while the pyrimidine bases are involved in equilibrium between the *endo*-helical and *exo*-helical conformations.

Overall, one can say that the ways in which DNA neutralises perturbations induced by unpaired bases are fairly varied and depend on the local "stacking" of the bases, the nature of the superfluous unit, the flexibility of the sugar phosphate skeleton, and the nucleotide sequence.

2. Oligonucleotide duplexes flanked by unpaired units

The influence of the terminal unpaired nucleotides on the stability of the NA double helices has been investigated mainly on short ribonucleotide duplexes.[146–159] Analysis of the available literature (Table 5) permits the following conclusions.

Table 5. Thermodynamic parameters of the formation of oligonucleotide complexes flanked by unpaired nucleotide residues.*

No.	Complex	Complex formation conditions	T_m °C	$-\Delta H^{0}$**, kcal mol^{-1}	$-\Delta S^{0}$, e.u.	$-\Delta G^{0}(t, °C)$, kcal mol^{-1}	Refs.
1	r(C C G G) (G G C C)r	0.01 M cacodylate buffer, pH 7.0, 1 mM EDTA, 1 M NaCl, c_0 = 0.1 mM	27.1	34.7	97.1	4.41 (39°)	[152]
2	r(A C C G G)p p(G G C C A)r	ditto	36.4	39.7	109.9	5.44 (39°)	[152]
3	r(C C G G U)p p(U G G C C)r	»	37.8	42.8	119.3	5.60 (39°)	[152]
4	r(C C G G A)p p(A G G C C)r	»	44.6	47.3	130.4	6.57 (39°)	[152]
5	r(G C C G G)p p(G G C C G)r	0.01 M phosphate buffer, pH 7.0, 0.5 mM EDTA, 1 M NaCl, c_0 = 0.1 mM	31.0	34	93	—	[158]
6	r(U C C G G)p p(G G C C U)r	ditto	29.7	37	104	—	[158]
7	r(C C G G C)p p(C G G C C)r	»	34.4	40.3	113	—	[154]
8	r(C C G G G)p p(G G G C C)r	»	46.5	48.3	133	—	[154]
9	r(G G C C) (C C G G)r	»	35.0	36.0	98	—	[154]
10	r(A G G C C)p p(C C G G A)r	»	38.3	37.9	103	—	[158]
11	r(C G G C C)p p(C C G G C)r	»	39.3	35.1	94	—	[158]
12	r(U G G C C)p p(C C G G U)r	»	34.5	35.6	97	—	[158]
13	r(G G C C C)p p(C C C G G)r	»	47.5	41.7	112	—	[154]
14	r(G G C C U)p p(U C C G G)r	»	51.5	48.3	130	—	[154]
15	r(G G C C G)p p(G C C G G)r	»	57.5	52.3	140	—	[154]
16	r(G G C C A)p p(A C C G G)r	»	58.0	51.5	137	—	[154]
17	r(G C G C) (C G C G)r	0.005 M phosphate buffer, pH 7.0, 0.5 mM EDTA, 1 M NaCl, c_0 = 1 mM	27.4	32.3	89.0	—	[157]
18	r(G G C G C)p p(C G C G G)r	ditto	28.8	35.3	98.7	—	[157]
19	r(U G C G C) (C G C G U)r	»	31.5	29.2	77.6	—	[157]
20	r(G C G C A)p p(A C G C G)r	»	51.9	50.0	135.7	—	[157]
21	r(G C G C C)p p(C C G C G)r	»	41.3	39.2	106.2	—	[157]
22	r(G C G C G)p p(G C G C G)r	»	50.7	47.9	129.5	—	[157]
23	r(G C G C U)p p(U C G C G)r	»	45.5	45.5	124.6	—	[157]
24	d(G A A T T C G C G) (G C G C T T A A G)d	0.01 M phosphate buffer, pH 7.0, 0.1 mM EDTA, 1 M NaCl, c_0 = 0.7 mM	46.3	60 (35)	—	—	[155]
25	d(G G A A T T C C) (C C T T A A G G)d	ditto	52.5	59 (59)	—	—	[155]

* Data for the fully complementary complexes are quoted for comparison.

** The values of ΔH^{0}_{v-H} are given in brackets.

*** c_0 is the oligonucleotide concentration.

The flanking of the double helix by an unpaired terminal residue increases the stability of the duplex.

Stabilisation is achieved as a result of additional interplanar interactions of the unpaired residue with the terminal paired base of the opposite strand; this entails a decrease in the end pair "fraying" effect.

The increase in stability is due to the favourable change in the enthalpy of complex formation (the contribution by additional "stacking") which exceeds the unfavourable entropic effect (the decrease in the degree of "freedom" of the pendant residue) compared with the fully complementary duplex.

The following sequence in the stabilisation by unpaired residues has been established: G ≃ A > U ≃ C, i.e. purine bases are more effective than pyrimidine bases. A similar dependence of the stability of the duplexes also on the nature of the base involved in overlapping with the unpaired residue has also been observed.

In helices of the ribo-series, the 3'-terminal base increases stability to a greater extent than the base located at the 5'-end. On the other hand, in deoxyribonucleotide duplexes, the 5'-terminal unit makes a greater contribution to stabilisation,[112] which is apparently associated with the differences between the structures of the *A*- and *B*-forms.

Conversely, the presence of more than one unpaired residue at the ends of the duplex reduces the stability of the double helix.[155]

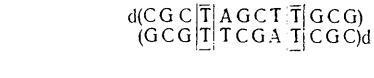
- - - o0o - - -

During the preparation of the manuscript for the press, a number of communications on the subject of the review were published.

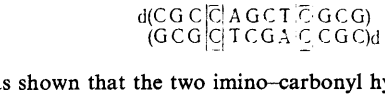
It was shown that[160] by the RS method that the non-complementary G.T pair induces a greater disturbance of *B*-DNA (solution) than of *Z*-DNA (crystalline state). Arguments are put forward in support of the hypothesis that this method in fact yields most information required to determine the geometry of the phosphodiester linkages near the non-complementary pairs.

A comparative study of the fine structure of the DNA duplexes in the *A*-form, one of which contains two adjacent G.T pairs while the others are unmodified analogues, has been carried out by XRDA (1.7 Å).[161] The thermal stability of imperfect and fully complementary double helices has also been studied (UV spectroscopy).

The structural features of the pyrimidine–pyrimidine non-complementary pairs C.C and T.T forming part of the composition of the 12-unit DNA duplexes have been studied by NMR spectroscopy:[162]



and



It was shown that the two imino–carbonyl hydrogen bonds, stabilising the T.T pair, are weaker than the corresponding bonds in the G.T pair. As regards the C.C pair, the experimental data did not permit a correct choice between several pairing schemes. It was only evident that one of the bases of the pair must be protonated.

A study has been made[163] of the stability of the formation of the $C.C^{+}$ pairs in 2-, 3-, or 6-unit cytosine tracts localised within an antiparallel DNA duplex. It had been shown earlier that such pairs are formed at the ends of an antiparallel duplex.[121] According to CD data, the pairing of cytosine residues does not occur in any of the duplexes studied.

The interrelation between the thermodynamic parameters of the fusion of defective duplexes, formed by the matrix DNA and decanucleotide primers containing non-complementary bases at the 3'-end (variants of the non-complementary pairs—G.T, C.T, and T.T), on the one hand, and the accuracy of the DNA replication, on the other, has been analysed.[164] The causes of the ambiguous relation are discussed.

The influence of the nucleotide sequence of the matrix DNA on the type of the non-complementary pair has been observed in the DNA-polymerase reaction.[165] The frequency of errors counted was artificially increased by using only three out of the four deoxyribonucleotide triphosphates.

The additional possibility of the formation of non-canonical pairs has been discussed in a theoretical study.[166] According to calculations of the energies of the intermolecular interactions (AAPF), in alternative pairing schemes one of the bases is in the *syn*-orientation and a water molecule participates in the formation of hydrogen bonds. According to the authors, such pairs are closer to the Watson–Crick pairs in terms of geometrical parameters.

The influence of the terminal non-complementary pairs (A.C, A.A, and A.G) on the thermal stability of oligoribonucleotide duplexes has been discussed.[167]

The dimorphous structure of poly(rA).poly(dT) has been confirmed by Raman spectroscopy.[168]

Comparative studies have been made of the thermal stabilities and of the secondary structures (by UV and CD spectroscopy) for a series of oligomeric DNA, RNA, and hybrid RNA.DNA duplexes.[169] It has been shown that the ribonucleotide duplex is the most stable, the hybrid duplexes occupying an intermediate position. The *A*-form of the double helix is characteristic of the latter but it is not fully identical with *A*-RNA.

The influence of two unpaired thymine residues, flanking the hexanucleotide duplexes $[d(GC)_3]_2$ or $[d(CG)_3]_2$ at the 3'- or 5'-end, on the thermodynamic and structural characteristics of the double helices has been studied.[170] It has been observed with the aid of UV, NMR, and CD spectroscopy and by computer modelling that the presence of thymidine residues stabilises the double helix as a result of the stacking interaction of the unpaired bases, the effect being greater in the case of 5'-terminal substituents.

The individual thermodynamic contriution of the four-unit loop and of the unpaired base in the double helix section has been determined for 14-pin ribonucleotide structures.[171] It has been shown that the destabilising effect of the unpaired purine is greater than that of pyrimidine and depends on the nucleotide sequence of the section adjoining the defect.

References

1. A Sugino, S Hirose, R Okazaki *Proc. Nat. Acad. Sci. USA*, 1972, **69**(7) 1863
2. J D Watson *Molecular Biology of the Gene* (Translated into Russian from English), Izd. Mir, Moscow, 1978, p. 276
3. L L Gerchman, D B Ludlum *Biochim. Biophys. Acta*, 1973, **308** 310
4. R S Lasken, M F Goodman *J. Biol. Chem.*, 1984, **259**(18) 11491
5. O C Uhlenbeck, J Baller, P Doty *Nature*, 1970, **225**(5232), 508
6. G J Quigley, N G Seeman, A H-J Wang, et al. *Nucleic Acids Res.*, 1975, **2**(12) 2329
7. **R Mikelsaar** *J. Theor. Biol.*, **1981, 92 163**
8. W Taub, J L Sussman *Nucleic Acids Res.*, 1982, **10**(8) 2701
9. R B Wallace, A B Studencki, A Murasugi *Biochimie*, 1985, **67** 755
10. O Kennard *Nucleic Acids and Molecular Biology* (Edited by F Eckstein, D M J Lilley), Springer-Verlag, Berlin–Heidelberg, 1987, Vol. 1, p. 25
11. D J Patel, L Shapiro, D Hare *Nucleic Acids and Molecular Biology* (Edited by F Eckstein, D M J Lilley), Springer-Verlag, Berlin–Heidelberg, 1987, Vol. 1, p. 70
12. T Brown, O Kennard, G Kneale, D Rabinovich, *Nucleosides, Nucleotides*, 1985, **4**(1–2) 255
13. T Brown, W N Hunter, G Kneale, O Kennard *Proc. Nat. Acad. Sci. USA*, 1986, **83**(8) 2402
14. M A Saper, H Eldar, K Mizuuchi, et al., *J. Mol. Biol.*, 1986, **188**(1) 111
15. S Arnott, R Chandrasekaran, R P Millane, H-S Park *J. Mol. Biol.*, 1986, **188**(4) 631
16. N G Dolinnaya, E S Gromova *Usp. Khim.*, 1983, **52**(1) 138 [English translation in *Russ. Chem. Rev.*, 1983(1)]
17. A Pardi, I Tinoco *Biochemistry*, 1982, **21**(19) 4686
18. J Tropp, A G Redfield *Biochemistry*, 1981, **20**(8) 2133
19. S V Matveev, V V Filimonov, P L Privalov *Mol. Biol.*, 1982, **16**(6) 1234
20. J Gralla, D M Crothers *J. Mol. Biol.*, 1973, **73**(4) 483
21. M E Craig, D M Crothers, P Doty *J. Mol. Biol.*, 1971, **62**(2) 383
22. J M Benevides, A H-J Wang, A Rich, et al. *Biochemistry*, 1986, **25**(1) 41
23. V P Chuprina, V I Poltev *Nucleic Acids Res.*, 1983, **11**(15) 5205
24. V L Florent'ev *Bioorg. Khim.*, 1982, **8**(7) 885
25. G Milman, R Langridge, M J Chamberlin *Proc. Nat. Acad. Sci. USA*, 1967, **57**(6) 1804
26. S Heguchi, M Tsuboi, Y Iitaka *Biopolymers*, 1969, **7**(6) 909
27. M-I B Tunis, J E Hearst *Biopolymers*, 1968, **6**(8) 1218
28. D M Gray, R L Ratliff *Biopolymers*, 1975, **14**(3) 487
29. A Pardi, F H Martin, I Tinoco *Biochemistry*, 1981, **20**(14) 3986
30. F H Martin, I Tinoco *Nucleic Acids Res.*, 1980, **8**(10) 2295
31. H T Steely, D M Gray, R L Ratliff *Nucleic Acids Res.*, 1986, **14**(24) 10071
32. S B Zimmerman, B H Pheiffer *Proc. Natl. Acad. Sci. USA*, 1981, **78**(1) 78
33. H Shindo, U Matsumoto *J. Biol. Chem.*, 1984, **259**(14) 8682
34. G Gupta, M H Sarma, R H Sarma *J. Mol. Biol.*, 1985, **186**(2) 463
35. S Arnott, P J Bond, E Selsing, P J C Smith *Nucleic Acids Res.*, 1976, **3**(10) 2459
36. D G Reid, S A Salisbury, et al. *Eur. J. Biochem.*, 1983, **135**(2) 307
37. H-Y Wu, M J Behe *Proc. Nat. Acad. Sci. USA*, 1984, **81**(23) 7284
38. M J Behe *Biopolymers*, 1986, **25**(3) 519
39. S D Jayasena, M J Behe *Nucleic Acids Res.*, 1987, **15**(9) 3907
40. R E Dickerson, H R Drew, B N Conner, et al. *Science*, 1982, **216** 475
41. A H-J Wang, S Fujii, J H van Boom, et al. *Nature*, 1982, **299**(5884) 601
42. J-R Mellema, C A G Haasnoot, G A van der Marel, et al. *Nucleic Acids Res.*, 1983, **11**(16) 5717
43. E Selsing, R D Wells, T A Early, D R Kearns *Nature*, 1978, **275**(5677) 249

44. E Selsing, R D Wells *J. Biol. Chem.*, 1979, **254**(12) 5410
45. E Selsing, R D Wells, C I Alden, S Arnott *J. Biol. Chem.*, 1979, **254**(12) 5417
46. J W Nelson, I Tinoco *Biochemistry*, 1982, **21**(21) 5289
47. M Riley, B Maling, M I Chamberlin *J. Mol. Biol.*, 1966, **20**(2) 359
48. T G Dewey, D H Turner *Biochemistry*, 1979, **18**(26) 5757
49. M Chamberlin, P Berg *J. Mol. Biol.*, 1964, **8**(2) 297
50. S Gillam, K Waterman, M Smith *Nucleic Acids Res.*, 1975, **2**(5) 625
51. E J O'Brien, A W MacEwan *J. Mol. Biol.*, 1970, **48**(2) 243
52. H Takashima, M Nakanishi, M Tsuboi *Biochemistry*, 1985, **24**(18) 4823
53. F H C Crick *J. Mol. Biol.*, 1966, **19**(2) 548
54. J W Keepers, P Schmidt, T L James, P A Kollman *Biopolymers*, 1984, **23**(12) 2901
55. V P Chuprina, V I Poltev *Nucleic Acids Res.*, 1985, **13**(1) 141
56. D J Patel, S A Kozlowski, L A Marky, et al. *Biochemistry*, 1982, **21**(3) 437
57. A Pardi, K M Morden, D J Patel, I Tinoco *Biochemistry*, 1982, **21**(25) 6567
58. D J Patel, A Pardi, K Itakura *Science*, 1982, **216**(4546) 581
59. L A Marky, S A Kozlowski, K J Breslauer *Biophys. J.*, **33**(2), Pt. 2 318a
60. D Hare, L Shapiro, D J Patel *Biochemistry*, 1986, **25**(23) 7445
61. D J Patel, S A Kozlowski, L A Marky, et al. *Biochemistry*, 1982, **21**(3) 428
62. R E Dickerson, H R Drew *J. Mol. Biol.*, 1981, **149**(4) 761
63. D R Hare, D E Wemmer, S-H Chou, et al. *J. Mol. Biol.*, 1983, **171**(3) 319
64. Y G Chu, I Tinoco *Biopolymers*, 1983, **22**(4) 1235
65. D Porschke, M Eigen *J. Mol. Biol.*, 1971, **62**(2) 361
66. N Tibanyenda, S H de Bruin, C A G Haasnoot, et al. *Eur. J. Biochem.*, 1984, **139**(1) 19
67. E Quignard, G V Fazakerley, G van der Marel, et al. *Nucleic Acids Res.*, 1987, **15**(8) 3397
68. W N Hunter, T Brown, G Kneale, et al. *J. Biol. Chem.*, 1987, **262**(21) 9962
69. P J Greene, M S Poonian, A L Nussbaum, et al. *J. Mol. Biol.*, 1975, **99**(2) 237
70. S Shenoy, K Daigle, K C Ehrlich, et al. *Nucleic Acids Res.*, 1986, **14**(11) 4407
71. T Brown, O Kennard, G Kneale, D Rabinovich *Nature*, 1985, **315**(6020) 604
72. G Kneale, T Brown, O Kennard, D Rabinovich *J. Mol. Biol.*, 1985, **186**(4) 805
73. W N Hunter, G Kneale, T Brown, et al. *J. Mol. Biol.*, 1986, **190**(4) 605
74. M McCall, T Brown, O Kennard *J Mol. Biol.*, 1985, **183**(3) 385
75. P S Ho, C A Frederick, G J Quicley, et al. *Embo J.*, 1985, **4**(13A) 3617
76. T R Krugh, M A Young *Biochem. Biophys. Res. Commun.*, 1975, **62**(4) 1025
77. M A Young, T R Krugh *Biochemistry*, 1975, **14**(22) 4841
78. A G Lezius, E Domin *Nature New Biology*, 1973, **244**(136) 169
79. D M Gray, R L Ratliff *Biopolymers*, 1977, **16**(6) 1331
80. T A Early, J Olsmted, D R Kearns, A G Lezius *Nucleic Acids Res.*, 1978, **5**(6) 1955
81. T A Millican, G A Mock, M A Chauncey, et al. *Nucleic Acids Res.*, 1984, **12**(19) 7435
82. S A Salisbury, N A Anand *J. Chem. Soc., Chem. Commun.*, 1985, 985
83. H Mizuno, K Tomita, E Nakagawa, et al. *J. Mol. Biol.*, 1981, **148**(1) 103
84. P J Romaniuk, D W Hughes, R J Grégoire, et al. *Biochemistry*, 1979, **18**(23) 5109
85. D Aikema, P A Hader, R A Bell, T Neilson *Biochemistry*, 1982, **21**(9) 2109
86. N Sugimoto, R Kierzek, S M Freier, D H Turner *Biochemistry*, 1986, **25**(19) 5755
87. S K Podder *Nature New Biology*, 1971, **232**(30) 115
88. K F Bautz, F A Bautz *Proc. Nat. Acad. Sci. USA*, 1964, **52**(6) 1476
89. T Ackermann, V Gramlich, H Klump, et al. *Biophys. Chem.*, 1979, **10**(3) 231
90. O C Uhlenbeck, F H Martin, P Doty *J. Mol. Biol.*, 1971, **57**(2) 217
91. S M Freier, R Kierzek, M H Caruthers, et al. *Biochemistry*, 1986, **25**(11) 3209
92. D J Patel, S A Kozlowski, S Ikuta, K Itakura *Biochemistry*, 1984, **23**(14) 3218
93. X Gao, D J Patel *J. Biol. Chem.*, 1987, **262**(35) 16973
94. M H Sarma, G Gupta, R H Sarma, et al. *Biochemistry*, 1987, **26**(24) 7707
95. W N Hunter, T Brown, O Kennard *Nucleic Acids Res.*, 1987, **15**(16) 6589
96. W N Hunter, T Brown, N N Anand, O Kennard *Nature*, 1986, **320**(6062) 552
97. R B Wallace, I Shaffer, R F Murphy, et al. *Nucleic Acids Res.*, 1979, **6**(11) 3543
98. D H Evans, A R Morgan *Nucleic Acids Res.*, 1986, **14**(10) 4267
99. D J Patel, S A Kozlowski, S Ikuta, K Itakura *Biochemistry*, 1984, **23**(14) 3207
100. L-S Kan, S Chandrasegaran, S M Pulford, P S Miller *Proc. Nat. Acad. Sci.* USA, 1983, **80** 4263
101. G G Privé, U Heinemann, S Chandrasegaran, et al., Science, 1987, **238**(4826) 498
102. W N Hunter, T Brown, O Kennard *J. Biomol. Struct. Dynam.*, 1986, **4**(2) 173
103. F N Arnold, S Wolk, P Cruz, I Tinoco *Biochemistry*, 1987, **26**(13) 4068
104. J B Dodgson, B D Wells *Biochemistry*, 1977, **16**(11) 2367
105. D J Patel, S A Kozlowski, S Ikuta, K Itakura *Fed. Proc.*, 1984, **43**(11) 2663
106. A J Lomant, J R Fresco *J. Mol. Biol.*, 1973, **77**(3) 345
107. J R Fresco, M B Alberts *Proc. Nat. Acad. Sci. USA*, 1960, **46**(3) 311
108. R F Steiner *J. Biol. Chem.*, 1960, **235**(10) 2946
109. A J Lomant, J F Fresco *Biopolymers*, 1973, **12**(8) 1889
110. F Aboul-ela, D Koh, I Tinoco, F H Martin *Nucleic Acids Res.*, 1985, **13**(13) 4811
111. A G Cornelis, J H J Haasnoot, J F den Hartog, et al. *Nature*, 1979, **281**(5728), 235
112. J-R Mellema, R van der Woerd, G A van der Marel, et al. *Nucleic Acids Res.*, 1984, **12**(12) 5061
113. M F Summers, R A Byrd, K A Gallo, et al. *Nucleic Acids Res.*, 1985, **13**(17) 6375
114. C A G Haasnoot, J H J Hartog, J F M de Rooij, et al. *Nucleic Acids Res.*, 1980, **8**(1) 169
115. C W Hilbers, C A G Haasnoot, S H de Bruin, et al. *Biochimie*, 1985, **67** 685

116. S Ikuta, R Chattopadhyaya, H Ito, et al. *Biochemistry*, 1986, **26**(17) 4840
117. L E Xodo, G Manzini, F Quadrifoglio, et al. *Nucleic Acids Res.*, 1986, **14**(13) 5389
118. F H Martin, O C Uhlenbeck, P Doty *J. Mol. Biol.*, 1971, **57**(2) 201
119. O C Uhlenbeck, P N Borer, B Dengler, I Tinoco *J. Mol. Biol.*, 1973, **73**(4) 483
120. R V Hosur, A Sheth, K V Chary, et al. *Biochem. Biophys. Res. Commun.*, 1986, **139**(3) 1224
121. D M Gray, T Cui, R L Ratcliff *Nucleic Acids Res.*, 1984, **12**(19) 7565
122. K A Hartman, A Rich *J. Am. Chem. Soc.*, 1965, **87**(9) 2033
123. R B Inman *J. Mol. Biol.*, 1964, **9**(3) 624
124. D M Gray, M Vaughan, R L Ratcliff, F N Hayes *Nucleic Acids Res.*, 1980, **8**(16) 3695
125. J Ninio *Biochimie*, 1979, **61** 1133
126. S Ikuta, K Takagi, R B Wallace, K Itakura *Nucleic Acids Res.*, 1987, **15**(2)
127. H Werntges, G Steger, D Riesner, H-J Fritz *Nucleic Acids Res.*, 1986, **14**(9) 3773
128. B Kramer, W Kramer, H-J Fritz, *Cell*, 1984, **38**(3) 879
129. D J Patel, S A Koslowski, L A Marky, et al. *Biochemistry*, 1982, **21**(3)
130. D Hare, L Shapiro, N J Patel *Biochemistry*, 1986, **25**(23) 7456
131. S Roy, B Sklenar, E Appella, J S Cohen *Biopolymers*, 1987, **26**(12)
132. L Joshua-Tor, D Rabinovich, H Hope, et al. *Nature*, 1988, **334**(6177) 82
133. M Miller, R Wlodawer, E Appella, J L Sussman *J. Mol. Biol.*, 1987, **195**(4) 967
134. M Miller, R W Harrison, A Wlodawer, et al. *Nature*, 1988, **334** 85
135. M Miller, W Kirchhoff, F Schwarz, et al. *Nucleic Acids Res.*, 1987, **15**(9)
136. S Roy, S Weinstein, B Borah, et al. *Biochemistry*, 1986, **25**(23) 7417
137. S A Woodson, D M Crothers *Biochemistry*, 1988, **27**(9) 3130
138. K M Morden, Y G Chu, F H Martin, I Tinoco *Biochemistry*, 1983, **22**(24) 5557
139. M W Kalnik, D G Norman, M G Zagorski, et al. *Biochemistry*, 1989, **28**(1) 294
140. Y T van den Hoogen, A A van Beuzekom, H van den Elst, et al. *Nucleic Acids Res.*, 1988, **16**(7) 2971
141. Y T van den Hoogen, A A van Beuzekom, E de Vroom, et al. *Nucleic Acids Res.*, 1988, **16**(11) 5013
142. D H Evans, A R Morgan *J. Mol. Biol.*, 1982, **160**(1) 117
143. D H Evans, A R Morgan *Nucleic Acids Res.*, 1986, **14**(14) 4267
144. S A Woodson, D M Crothers *Biochemistry*, 1987, **26**(3) 904
145. S A Woodson, D M Crothers *Biochemistry*, 1988, **27**(1) 436
146. T E England, T Neilson *Can. J. Biochem.*, 1975, **55** 365
147. D Alkema, R A Bell, P A Hader, T Neilson *J. Am. Chem. Soc.*, 1981, **103**(10) 2866
148. P L D'Andrea, D Alkema, R A Bell, et al. *J. Am. Chem. Soc.*, 1983, **105**(3) 636
149. P J Romaniuk, D W Hughes, R J Gregoire, et al. *J. Am. Chem. Soc.*, 1978, **100**(12)
150. D Alkema, R A Bell, P A Hader, T Neilson *FEBS Lett.*, 1981, **136**(1) 70
151. T Neilson, P J Romaniuk, D Alkema, et al. *Nucleic Acids. Res., Symp. Ser.*, 1980(7) 293
152. M Petersheim, D H Turner *Biochemistry*, 1983, **22**(2) 256
153. M Petersheim, D H Turner *Biochemistry*, 1983, **22**(2) 269
154. S M Freier, B J Burger, D Alkema, et al. *Biochemistry*, 1983, **22**(26) 6198
155. D J Patel, S A Kozlowski, L A Marky, et al. *Biochemistry*, 1982, **21**(3) 451
156. J W Nelson, F H Martin, L Tinoco *Biopolymers*, 1981, **20**(12) 2509
157. S M Freier, N Sugimoto, A Sinclair, et al. *Biochemistry*, 1986, **25**(11)
158. S M Freier, D Alkema, A Sinclair, et al. *Biochemistry*, 1985, **24**(17) 4533
159. M Petersheim, D H Turner *Biochemistry*, 1983, **22**(2) 264
160. J M Benevides, A H-J Wang, G A van der Marel, et al. *Biochemistry*, 1989, **28**(1)
161. D Rabinovich, T Haran, M Eisenstein, Z Shakked *J. Mol. Biol.*, 1988, **200**
162. M Kouchakdjian, B F L Li, P F Swann, D J Patel *J. Mol. Biol.*, 1988, **202** 139
163. E L Edwards, R L Ratcliff, D M Gray *Biochemistry*, 1988, **27**(14) 5166
164. J Petruska, M F Goodman, M S Boosalis, et al. *Proc. Nat. Acad. Sci. USA*, 1988, **85** 6252
165. M-D Lai, K L Beattie *Biochemistry*, 1988, **27**(5) 1722
166. S V Shteinberg, V I Poltev *Biofizika* 1988, **33**(6) 922
167. D R Hickey, D H Turner *Biochemistry*, 1985, **24**(15) 3987
168. J M Benevides, G J Thomas *Biochemistry*, 1988, **27**(10) 3868
169. R T Walker *Nucleic Acids Res.*, 1988, **16**(7) 3091
170. M Senior, R A Jones, K J Breslauer *Biochemistry*, 1988, **27**(10) 3879
171. D R Groebe, O C Uhlenbeck *Biochemistry*, 1989, **28**(2) 742