Nucleic Acids Research, Vol. 20, No. 19 5011–5016

A comprehensive classification of nucleic acid structural families based on strand direction and base pairing

Richard Lavery*, Krystyna Zakrzewska, Jian-Sheng Sun[1] and Stephen C.Harvey[2]
Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, Paris 75005, [1]Laboratoire de Biophysique, Muséum National d'Histoire Naturelle, 43 rue Cuvier, Paris 75005, France and [2]Department of Biochemistry, University of Alabama at Birmingham, Birmingham AL 35294, USA

Received July 30, 1992; Revised and Accepted September 3, 1992

ABSTRACT

We propose a classification of DNA structures formed from 1 to 4 strands, based only on relative strand directions, base to strand orientation and base pairing geometries. This classification and its associated notation enable all nucleic acids to be grouped into structural families and bring to light possible structures which have not yet been observed experimentally. It also helps in understanding transitions between families and can assist in the design of multistrand structures.

INTRODUCTION

In recent years, many unusual DNA structures have come to light involving unusual base pairing, parallel strands, modified nucleotides, loops, cruciforms and multistrand complexes (1). Although these structures often initially appeared to be experimental curiosities many have turned out to play biological roles or to be adaptable for the purposes of artificial genetic control. One clear example of this involves triple helices which are encountered *in vivo* in so-called H-DNA complexes (1–3) where an opened duplex loops back to form a triple helix with an upstream sequence. Triple helices have also become important in the search for anti-gene probes (4) where normal or modified oligonucleotides can be used to precisely target a chosen duplex sequence.

Because of this rapid increase in the complexity of known nucleic acid architectures, it seems to be of interest to attempt to classify all such structures in a clear and homogeneous way. The simple creation of such a classification should help in understanding DNA structure, in defining which new structures may be created in the future, in looking at transitions between structural forms and, at least partially, in characterizing their conformations—without the necessity of any detailed structural studies. In the present article we will thus introduce a simplified diagrammatic representation and an associated notation for possible nucleic acid structures having from 1 to 4 strands. This classification ignores conformational detail and will be based exclusively on the type of base pairing within the structure and the phosphodiester strand directions.

RESULTS AND DISCUSSION

The starting point for our classification is the observation that, due to their molecular asymmetry, each nucleic acid base has two unique faces (5,6). The definition of one such face can be made using the vector construction shown in figure 1 for an adenine-thymine Watson–Crick base pair. For each base, two vectors are drawn starting from the centre of mass, respectively towards the glycosidic atom (a) and towards the centre of the Watson–Crick base pairing face (b). The cross product of these vectors (a×b) gives rise to a vector normal to the base plane, pointing in a defined direction. From now on we will colour the face of the base corresponding to this vectorial direction in white (face I) and the opposing face in black (face II).

Rose and his colleagues (5) have already proposed a definition of the faces of ring molecules based on the sense of numbering the first ring (clockwise numbering corresponding to the 'α-face' and anti-clockwise to the 'β-face'). In the case of DNA, this convention has the advantage of giving unique names to the faces of Watson–Crick base pairs. This result has already been used in discussing the B–Z transition which involves turning over the base pairs (6). The Rose notation cannot however be used easily in a general notation for nucleic acids since, firstly, it leads to purine and pyrimidine bases in a single strand of DNA having different faces pointing upwards with respect to the 5′→3′ direction and, secondly, it does not respect the pseudo-dyad symmetry within a normal B-DNA duplex.

With our definition, we can see, in figure 1, that a Watson–Crick base pair will have one white face and one black face (I+II) visible from whichever side it is viewed. Figure 2 shows four known forms of base pairing, which we will describe hereafter by a single-letter code: Watson–Crick (W), reversed Watson–Crick (C), Hoogsteen (H) and reversed Hoogsteen (R). (Although this code is not conventional it is advantageous for the notation we will derive to describe pairing by a single letter). Since reversed Hoogsteen pairs can be obtained from Watson–Crick pairs by sliding the second base position without turning it over, they also have one white and one black face visible (I+II). In contrast, reversed Watson–Crick and Hoogsteen pairs (which are again mutually interchangeable by sliding one base) both have either two white faces (I+I) or two black faces (II+II) visible. It should be remarked that, in these terms, wobble pairs

* To whom correspondence should be addressed

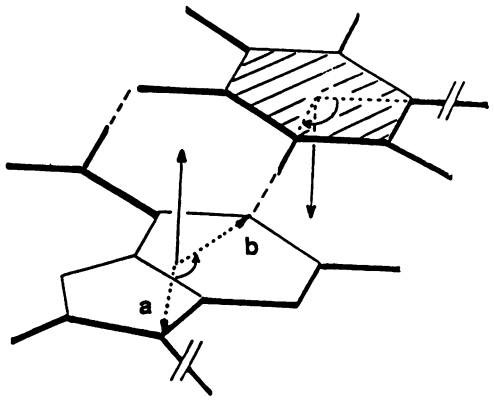

Figure 1. Adenine-Thymine Watson–Crick base pair showing the vectorial definition of the base faces (I: white, II: black).

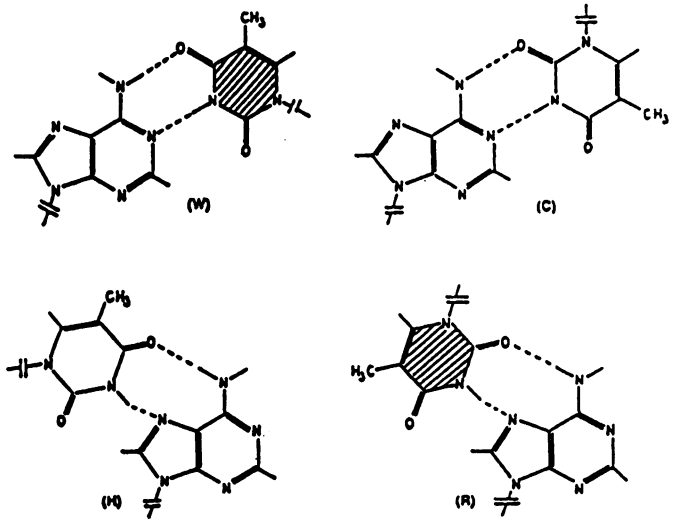

Figure 2. Four types of base pairing. W: Watson–Crick, C: Reversed Watson–Crick, H: Hoogsteen, R: Reversed Hoogsteen.

are the same as Watson–Crick pairs and do not require a special classification, since they only involve a lateral sliding of the bases along the base pair pseudo-dyad axis.

(i) Nucleotides

We will now introduce a simplified diagram to represent any given nucleotide within a nucleic acid structure. Firstly, the base is represented by a square with a diagonal line indicating the glycosidic bond and a jagged edge representing the Watson–Crick base pairing face. The squares can then be coloured white (I) or black (II) to show the upward pointing face. To complete our shorthand diagrams, we add a circle at the end of the glycosidic bond line to indicate the strand direction: an open circle corresponding to the direction 5′→3′ pointing upwards and a filled circle corresponding to the direction 5′→3′ pointing downwards. Figure 3 (bottom) shows that, for a single nucleotide, only two possibilities exist. Either the white face of the base is associated with a 5′→3′ strand direction—termed form 'a'—or the white face is associated with the 3′→5′ direction—termed form 's'. These letters were chosen to recall the anti and syn conformations around the glycosidic bond, but *do not generally imply that such conformations must be associated with the corresponding nucleotides*. We will return to this point shortly in connection with Z-DNA. In the upper part of figure 3 the base pairs presented in figure 2 are redrawn using the shorthand notation (note all nucleotides are shown in the a-form.

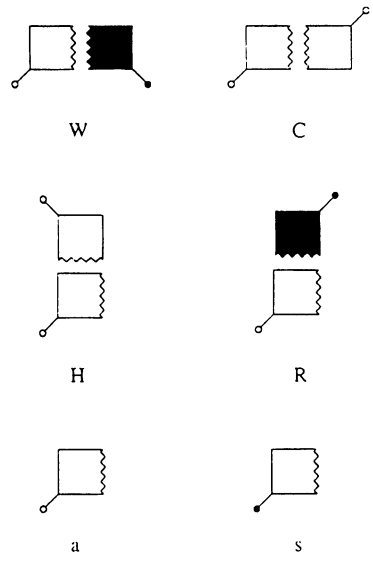

Figure 3. Diagrammatic representation of nucleotides (lower 2 diagrams) and of the four types of base pairing (upper 4 diagrams).

(ii) Double stranded structures

From this starting point, it is now possible to pass to double stranded helices and ask how many distinct structural families may be generated. In figure 4, we have listed the four possible combinations of strand directions for each of the four types of base pairing. Note that, for simplicity, each base pair is oriented so that the left hand base (or the lower base in the case of Hoogsteen or reversed Hoogsteen pairs) shows its white face. Each of these choices leads to a structural family which can be defined by a concise notation consisting of a letter to specify the base pairing (W, C, H or R), a prefix indicating whether the strand directions are parallel (+) or anti-parallel (−) and a suffix specifying whether the left-hand (or lower) nucleotide is of type 'a' or type 's'. Note that from this notation it is possible to deduce the type of the second nucleotide as shown in table 1—for W and R base pairs, anti-parallel strands imply that both nucleotides are of the same type, while parallel strands lead to nucleotides of different types. For C and H base pairs the opposite rules apply.

Figure 4 indicates that, for Watson–Crick base pairs, there are in fact only 3 possible structural families, the third and fourth diagrams $^{+}W_a$ and $^{+}W_s$ being degenerate (by a rotation around the base pair pseudo-dyad axis). These two duplexes can thus be defined by the simplified notation ^{+}W. A similar degeneracy occurs for reversed Watson–Crick base pairs, where the $^{-}C_a$ and $^{-}C_s$ diagrams can be inter-converted by a rotation around the base pair normal. These duplexes can thus be described by the notation ^{-}C. For Hoogsteen and reversed Hoogsteen base pairs no such degeneracies occur due to the absence of a pseudo-

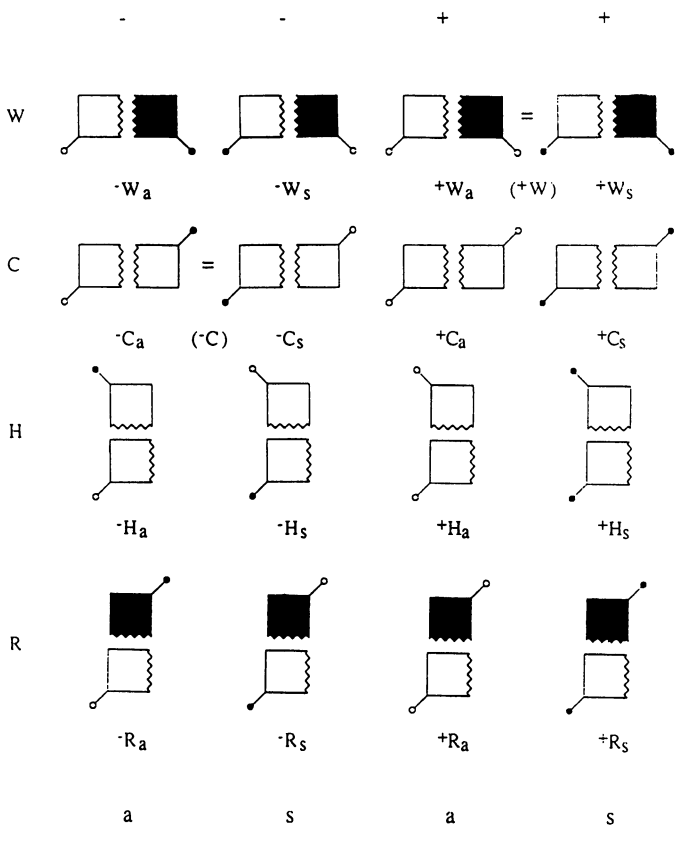

Figure 4. Diagrammatic representation of the families of duplex structures.

Table 1. Second nucleotide types as a function of strand direction and base pairing. Bracketed letters correspond to duplexes containing mixed nucleotide types.

Strands:	−	−	+	+
1st nucleotide:	a	s	a	s
Pairing		2nd nucleotide		
W	a	s	(s)	(a)
C	(s)	(a)	a	s
H	(s)	(a)	a	s
R	a	s	(s)	(a)

Table 2. Structural families of duplex DNA

Strands:		−	−	+	+	
1st nucleotide:		a	s	a	s	
Pairing		Notation				
W		$^{-}W_a$	$^{-}W_s$	$^{+}W_a$ =	$^{+}W_s$	($^{+}$W)
C	($^{-}$C)	$^{-}C_a$ =	$^{-}C_s$	$^{+}C_a$	$^{+}C_s$	
H		$^{-}H_a$	$^{-}H_s$	$^{+}H_a$	$^{+}H_s$	
R		$^{-}R_a$	$^{-}R_s$	$^{+}R_a$	$^{+}R_s$	

dyad axis. There are thus a total of 14 distinct families of duplex DNA (see table 2).

The first family in figure 4, described by the code '$^{-}W_a$', corresponds to the most common form of duplex DNA in the B or A conformations with anti-parallel strands, Watson–Crick hydrogen bonding and a-type nucleotides which are indeed in anti conformations. The second family $^{-}W_s$ actually corresponds to Z-DNA. This representation of the Z form makes it clear that base pairs have to be turned over in passing from the B to the Z conformation (6,7), since if we wish to align the strand directions between the first two families in figure 4 it is necessary to invert the $^{-}W_s$ diagram around a horizontal axis leading to a base pair with a black face on the left and a white face on the right. If we interpret the nucleotide types 'a' and 's' as leading to anti and syn conformations, our notation would imply that Z-DNA should contain only syn nucleotides. In reality, the difficulty of forcing pyrimidines into the syn form leads Z-compatible purine-pyrimidine base sequences to adopt alternating anti and syn conformations. This however cannot change overall strand directions and thus the rotation of the pyrimidine sugar along with the base pair results in the contorted 'Z' backbone pathway characteristic of this type of duplex DNA. We must thus re-emphasize that the 'a' and 's' notation only relates base orientation to the strand direction and does not refer to the state of the glycosidic angle.

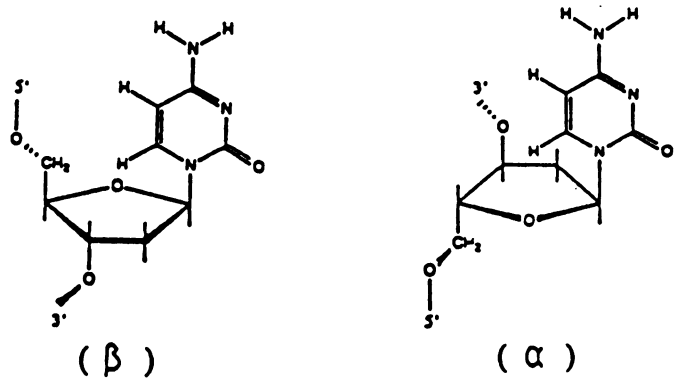

Figure 5. Comparison of a β-nucleotide and an α-nucleotide.

An alternative way to solve the difficulty of forming syn nucleotides is to change the stereochemistry at C1′. This takes us from usual β-nucleotides to α-nucleotides and consequently diminishes steric hindrance with C5′ in the case of the syn conformation (figure 5). It is worth pointing out that an all α-nucleotide duplex belonging to the family $^{-}W_s$ has indeed been observed (8).

The final family which can be made with Watson–Crick base pairs ($^{+}$W) has also been observed recently in parallel stranded duplexes where the second strand is again composed of α-nucleotides which easily accept the syn conformation (9).

Of the reversed Watson–Crick duplexes, only one family, $^{+}C_a$, is currently known. This structure appears to form with certain salts in poly(dC).poly(dC^{+}) duplexes under acidic conditions (10). It is also found in the parallel stranded structure proposed by Pattabiraman on the basis of molecular modelling (11) and subsequently observed in the parallel stranded AT sequences by Ramsing and Jovin (12).

Since normal duplex DNA generally prefers W or C type base pairing, no Hoogsteen or reversed Hoogsteen pairings are seen unless the former possibilities are excluded. This is the case in triple helices where the Watson–Crick hydrogen bonding possibilities are absorbed in the underlying duplex. In these cases, a third strand can be added to form duplexes of the families $^{+}H_a$, $^{-}R_a$ and (with α-nucleotides) $^{+}R_a$. These cases will be discussed in the following section. It is also possible to block the Watson–Crick hydrogen bonding face of chosen bases by chemical modification. This was done in the case of adenine

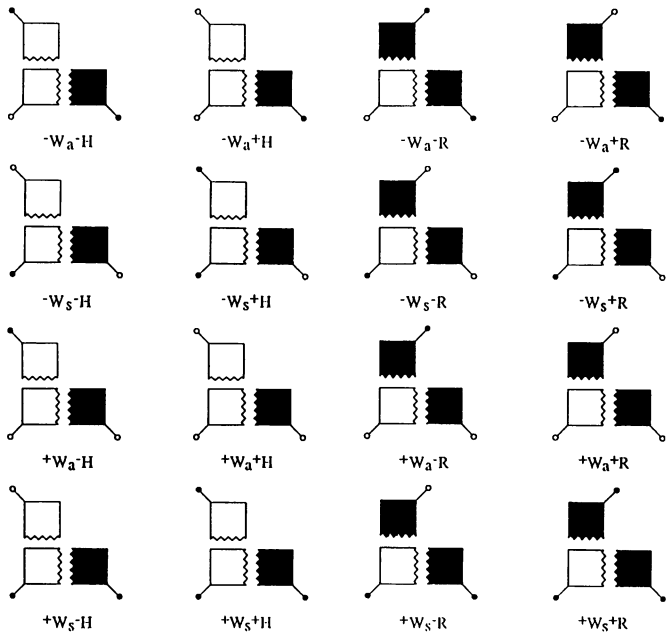

Figure 6. Triple helix families based on Watson–Crick duplexes.

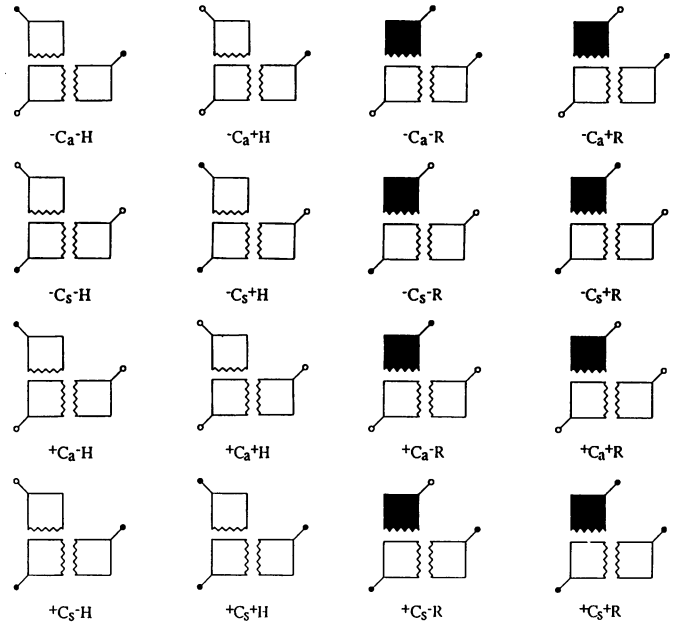

Figure 7. Triple helix families based on reversed Watson–Crick duplexes.

substituted at position 2 by a methyl group or by other more bulky substituents (13). The consequence was the formation of poly(dU).poly(dA*) parallel stranded $^{+}H_a$ duplexes. The same family is formed when protonation of adenine at N1 under acidic conditions (14) leads to the possibility of building parallel stranded poly(dA+).poly(dA+). No such double stranded structures using reversed Hoogsteen pairing are currently known.

Note that for Hoogsteen and reversed Hoogsteen base pairs, the letter 'a' or 's' in the notation we have proposed refers to the base using its major groove face for hydrogen bonding (as implied by the orientation of the H and R diagrams in figure 4). This choice is preferable since, as we shall see below, it enables us to build up triple helices very easily by combining two duplexes with a common nucleotide strand. As one referee of this article has pointed out, the notations $^{-}H_a/^{-}H_s$ and $^{+}R_a/^{+}R_s$ would be interchanged if the definition was made with respect to the base using its Watson–Crick bonding face. The same freedom does not exist for W or C pairing due to the degeneracy of the $^{+}W_a/^{+}W_s$ and $^{-}C_a/^{-}C_s$ states discussed above.

(iii) Triple stranded structures

In order to form triple helices (or 'triplexes') it is necessary to combine two double helix families from figure 4. Due to base pairing possibilities this requires the presence of a purine which can bind to two other bases via Watson–Crick (or reversed Watson–Crick) pairing at the same time as Hoogsteen (or reversed Hoogsteen) pairing. If we assume that the left hand base of each Watson–Crick family is a purine we can then build a total of 16 possible triple helices (figure 6) using in turn each of the Hoogsteen and reversed Hoogsteen families compatible with a given Watson–Crick family. It should be noted that there are no other possibilities with Watson–Crick base pairs, since if we had assumed the right hand base to be a purine we would have simply obtained the same families in a different orientation. It should also be noted that, once a triple helix has been formed, the degenerate ($^{+}W_a$, $^{+}W_s$) and ($^{-}C_a$, $^{-}C_s$) families become distinct since only one of their constituent bases carries the third strand.

Table 3. Double stranded and triple stranded nucleotide types.

Double strand		Triple strand (W)		Triple strand (C)	
$^{-}W_a$	<u>aa</u>	$^{-}W_a{-}H$	aas	$^{-}C_a{-}H$	ass
$^{-}W_s$	<u>ss</u>	$^{-}W_a{+}H$	<u>aaa</u>	$^{-}C_a{+}H$	asa
^{+}W	<u>as</u>	$^{-}W_a{-}R$	<u>aaa</u>	$^{-}C_a{-}R$	asa
		$^{-}W_a{+}R$	<u>aas</u>	$^{-}C_a{+}R$	ass
^{-}C	as	$^{-}W_s{-}H$	ssa	$^{-}C_s{-}H$	saa
		$^{-}W_s{+}H$	sss	$^{-}C_s{+}H$	sas
^{+}Ca	aa	$^{-}W_s{-}R$	sss	$^{-}Cs{-}R$	sas
$^{+}C_s$	ss	$^{-}W_s{+}R$	ssa	$^{-}Cs{+}R$	saa
$^{-}H_a$	as	$^{+}W_a{-}H$	ass	$^{+}C_a{-}H$	aas
$^{-}H_s$	sa	$^{+}W_a{+}H$	asa	$^{+}C_a{+}H$	aaa
$^{+}H_a$	<u>aa</u>	$^{+}W_a{-}R$	asa	$^{+}C_a{-}R$	aaa
$^{+}H_s$	ss	$^{+}W_a{+}R$	ass	$^{+}C_a{+}R$	aas
$^{-}R_a$	<u>*aa</u>	$^{+}W_s{-}H$	saa	$^{+}C_s{-}H$	ssa
$^{-}R_s$	ss	$^{+}W_s{+}H$	sas	$^{+}C_s{+}H$	sss
$^{+}R_a$	<u>*as</u>	$^{+}W_s{-}R$	sas	$^{+}C_s{-}R$	sss
$^{+}R_s$	sa	$^{+}W_s{+}R$	saa	$^{+}C_s{+}R$	ssa

Known structural families are underlined. (*duplexes found within triple helices)

As shown in figure 6, it is possible to give each of these triple helices a unique notation corresponding to the two constituent double helices. The first family, built from a $^{-}W_a$ Watson–Crick duplex and a $^{-}H_a$ Hoogsteen duplex, thus becomes $^{-}W_a{-}H_a$. In fact, the nucleotide type indicated for the second base pair can be dropped since it must be identical to that of the first pair – remember that the nucleotide type refers to the left-hand or lower nucleotides for the duplexes in figure 4 and

is thus necessarily the same for any pair of duplexes used to form a triplex. The first triplex family can thus be uniquely defined as $^{-}W_a{}^{-}H$.

For completeness, we also note that it is possible to construct triple helices starting from reversed Watson–Crick base pairs, leading again to 16 unique possibilities, shown in figure 7. Table 3 lists the type of each nucleotide in all the double stranded and triple stranded structures we have defined above and may be of use to the reader, although it is recalled that these types can also be deduced using the rules illustrated by table 1.

The best known triple helix made by adding a Hoogsteen bonded thymidine strand to an poly(dA).poly(dT) double helix (15) corresponds to the family $^{-}W_a{}^{+}H$ since all nucleotides are of a-type and the Hoogsteen bound poly(dT) strand is parallel to the adenosine strand of the duplex. An identical triple helix family CGC^{+} can also be formed under acidic conditions by adding a protonated cytosine strand to a poly(dG).poly(dC) duplex, again using Hoogsteen hydrogen bonding.

The only other way to form an all a-type triple helix starting from a Watson–Crick duplex is the family $^{-}W_a{}^{-}R$ (see table 3) which has indeed been experimentally observed for AAT and CGG triple helices (16,17) where the two purine strands form an anti-parallel reversed Hoogsteen duplex. A related family containing s-type nucleotides in the third strand, $^{-}W_a{}^{+}R$, is also known to exist when α-thymidine nucleotides, which can easily adopt the syn conformation, are built into the third strand of an TAT triplex (18).

Forming the triple helices shown in the second row of figure 6 seems unlikely since the only known form of the $^{-}W_s$ duplex is Z-DNA in which the major groove face of the base pairs is sterically hindered. In contrast, starting from a parallel strand Watson–Crick duplex ^{+}W (formed using an a-nucleotide pyrimidine strand) it may be possible to form triple helices belonging to the families $^{+}W_a{}^{+}H$ or $^{+}W_a{}^{-}R$ which only require syn conformations in the Watson–Crick bound pyrimidine strand (see table 3). Without using a-nucleotides, but starting from a parallel stranded reversed Watson–Crick duplex $^{+}C_a$ of the type described by Ramsing and Jovin (12), it may also be possible to build triple helices of the families $^{+}C_a{}^{+}H$ or $^{+}C_a{}^{-}R$ where all strands contains only a-type nucleotides. Forming any of these 4 latter triplexes however depends on the ability of the underlying duplex to adopt a conformation which sufficiently exposes the major groove face to accept a third strand. Results from the Jovin group suggest that an internal H-DNA triple helix formed under negative supercoiling stress by a plasmid containing a parallel-stranded $(dA)_{15}.(dT)_{15}$ insert in fact undergoes rearrangement to form an A.A.T triplex belonging to the family $^{-}W_a{}^{-}R$ (19).

(iv) Quadruple stranded structures

We do not attempt to define here all the possible four stranded complexes which may be formed. It is clear however that it is only with Hoogsteen base pairing that two duplexes can be put together to form a base quadruplet with 4 base pairing interfaces. If we limit ourselves to a-type nucleotides, this structure would be of the $(^{+}H_a)(^{+}H_a)$ family (which could also be written $(^{+}H_a)^2$) (see figure 8 left). This structure is indeed observed in poly(dG) and poly(dI) gels (20,21).

If only 3 base pairing interfaces are acceptable, then it is possible to combine two Watson–Crick duplexes which bind by Hoogsteen pairing between their major groove faces, $(^{-}W_a)^2$, see figure 8 centre. Such a structure appears to be formed by two AT duplexes linked by A-A Hoogsteen hydrogen bonds or

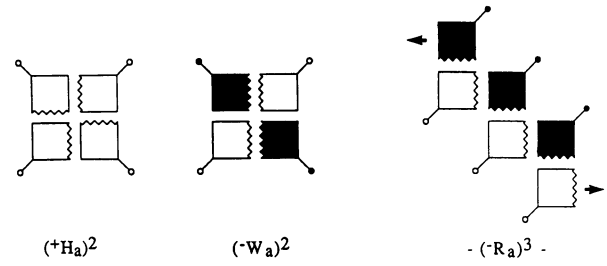

Figure 8. Two examples of four-stranded structures and a hypothetical multistrand structure.

central hydrogen bonds involving A⟷T interactions as reported by Borisova et al (22).

Finally, it is interesting to speculate that if a reversed Hoogsteen duplex is taken as the starting point, then it is possible to imagine creating an extended structure of the family $^{-}R_a$ (containing only a-type nucleotides) which could be denoted by $(^{-}R_a)^n$ (see figure 8 right) and might be considered the topological equivalent for nucleic acids of the well-known β-sheet structures formed by polypeptides.

(v) RNA structures

It should be stressed that the classification described above applies equally well to deoxyribo and to ribo nucleic acids. Standard duplex RNA, like A-DNA, falls in the $^{-}W_a$ family, however more complex forms of RNA exhibit, at least locally, other structural families. This is notably the case for transfer RNA's. $tRNA^{Phe}$ (23) contains both unusual base pairing leading to double stranded zones of the family $^{+}C_a$ (G15-C48) and $^{-}R_a$ (m^1A58-T54), while base triplets form zones belonging to the classes $^{-}W_a{}^{-}R$ (G22-C13-m^7G46) and $^{-}W_a{}^{+}H$ (A23-U12-A9). It is interesting to note that, although these unusual families occur only locally within the RNA tertiary structure, they generally involve conserved bases and are thus likely to be important for the overall folding (24). It seems probable that further studies of RNA's with complex tertiary folds will bring to light other unusual structural zones.

(vi) Comments on 'Westhof's rule'

In a recent letter (24), Eric Westhof has proposed a rule for determining the strand direction within polynucleotide complexes on the basis of base pairing and the syn/anti state of the glycosidic bonds. Base pairing is split into two features: whether common (as in W/C pairs) or different (as in H/R pairs) base pairing faces are used and whether the glycosidic bonds lie cis (W/H) or trans (C/R) with respect to the base pair. The rule then states that when common pairing faces are combined with cis glycosidic bonds and either anti/anti or syn/syn nucleotides, then antiparallel strands result. Changing an odd number of these characteristics will result in parallel strands, while changing an even number will have no effect.

Our remark in section (ii) '*...for W and R base pairs, anti-parallel strands imply that both nucleotides are of the same type, while parallel strands lead to nucleotides of different types. For C and H base pairs the opposite rules apply*' appears to state the same rule in other words. However, this is not in fact the case since we have also stressed that our nucleotide types define the strand direction with respect to the upward pointing base face

and that this need not necessarily correspond to the state of the glycosidic bonds as in Westhof's rule. The reason for our precaution in this case is that strand direction is an overall property of a nucleotide complex, whereas the anti/syn conformation of the glycosidic bond is a local conformational feature. This distinction is clear in the case of Z-DNA which has antiparallel strands although each of its Watson–Crick base pairs combines a syn-guanosine with an anti-cytidine. Westhof's rule predicts Z-DNA to have parallel strands.

We must also recall that our aim is a complete classification and not a predictive rule. Westhof's rule means adding to our classification the assumption that the 'a' and 's' subscripts imply anti or syn nucleotides, which is not always true. On the other hand using cardboard cut-outs of the nucleotide types shown in figures 3 and assembling them as in figures 4 and 6–8 will enable relative strand directions to be determined and will also give the strand direction with respect to any chosen base, a feature not contained in the rule discussed.

A final remark can be made concerning the definition of pairing by a combination of common or different pairing faces and cis/trans glycosidic bonds. These two features yield the W, C, H and R pairs we have discussed, but also allow for pairing involving Hoogsteen-Hoogsteen faces as in the unusual A-A interaction with trans glycosidic bonds discussed by Westhof. Such arrangements are topologically equivalent to the corresponding Watson–Crick/Watson–Crick paired family ($^{+}C_a$ in the case cited for A9-A23 in $tRNA^{Phe}$, 26). If the cis and trans forms of such pairs were described, for example, by the letters M and N (M recalling major groove face interactions) they would give rise to the classes: $^{-}M_a$, $^{-}M_s$, ^{+}M, ^{-}N, $^{+}N_a$, $^{+}N_s$ equivalent to the rows W and C in figure 4.

CONCLUSIONS

We have presented a classification and an associated notation for nucleic acid structures having from 1 to 4 strands. This classification helps in understanding the relationships between the increasingly complex structures which have been discovered for nucleic acids. These include structures involving modified nucleotides (notably, α-nucleotides which enrich conformational possibilities due to their favouring of syn glycosidic conformations). Such a classification should also be useful in studying transitions between different forms and in pointing out new structural possibilities that remain to be investigated.

Our characterization of nucleic acid structure is voluntarily limited to describing base pairing, base-to-strand and inter-strand orientations. It thus classifies nucleic acids at a level roughly equivalent to the secondary structures of proteins. Nucleic acids are however seen to have a much richer variety of forms at this level. As pointed out, the base-to-strand orientation used here is not synonymous with a syn-anti classification of the glycosidic bond (although the two are often related). This choice means that Z-DNA, whose strands each contain alternating syn- and anti-nucleotides, still belongs to a simple structural family. However, this classification does not rule out local changes of family within DNA or RNA duplexes, for example, due to mispairing.

Overall this classification shows that there are a total of 14 families of double stranded structures (of which 7 have been observed, see table 3), 16 families of triple stranded structures based on Watson–Crick duplexes (of which 3 have been observed) and a further 16 triple helices based on reversed Watson–Crick duplexes (which have not yet been observed).

ACKNOWLEDGEMENTS

We would like to thank Richard Dickerson for helpful remarks concerning this classification. R.L. and K.Z. wish to thank the Association for International Cancer Research (St Andrews University, U.K.) for their support. S.H. thanks the National Institutes of Health (GM-31015) and the National Science Foundation (DMB-00-05767). S.H. and R.L. were also funded by a joint NSF/CNRS research program.

REFERENCES

1. Wells,R.D. and Harvey,S.C. (eds.) Unusual DNA Structures (1988) Springer-Verlag New York.
2. Hanvey,J.C., Shimizu,M. and Wells,R.D. (1988) *Proc. Natl. Acad. Sci.* (USA) **85**, 6292–6296.
3. Htun,H. and Dahlberg,J.E. (1989) *Science* **243**, 1571–1576.
4. Hélène,C. and Toulmé,J.J. (1990) *Biochem. Biophys. Acta* **1049**, 99–125.
5. Rose,I.A., Hanson,K.R., Wilkinson,K.D. and Wimmer,M.J. (1980) *Proc. Natl. Acad. Sci. (USA)* **77**, 2439–2441.
6. Harvey,S.C. (1983) *Nucleic Acids Res.* **11**, 4867–4878.
7. Wang,A.H.J., Quigley,G.J., Kolpak,F.J., Crawford,J.L., van Boom,J.H., van der Marel,G. and Rich,A. (1979) *Nature* **282**, 680–686.
8. Morvan,F., Rayner,B., Imbach,J.L., Chang,D.K. and Lown,J.W. (1987) *Nucleic Acids Res.* **15**, 4241–4255.
9. Sun,J.S., François,J-C., Lavery,R., Saison-Behmoaras,T., Montenay-Garestier,T., Thuong,N.T. and Hélène,C. (1988) *A.C.S. Biochemistry* **27**, 6039–6045.
10. Arnott,S., Chandrasekharan,R. and Leslie,A.G.W. (1976) *J. Mol. Biol.* **106**, 735–738.
11. Pattabiraman,N. (1986) *Biopolymers* **25**, 1603–1606.
12. Ramsing,N.B. and Jovin,T.M. (1988) *Nucleic Acids Res.* **16**, 6659–6676.
13. Fukui,T. and Ikehara,M. (1979) *Biochem. Biophys. Acta.* **562**, 527–533.
14. Rich,A., Davies,D.R., Crick,F.H.C. and Watson,J.D. (1961) *J. Mol. Biol.* **3**, 71–86.
15. Arnott,S. and Selsing,E. (1974) *J. Mol. Biol.* **88**, 509–521.
16. Broitman,S.L., Im,D.D. and Fresco,J.R. (1987) *Proc. Natl. Acad. Sci. (USA)* **84**, 5120–5124.
17. Pilch,D.S., Levensen,C. and Shafer,R.H. (1991) *Biochemistry* **30**, 6081–6087.
18. Sun,J.S., Giovannangeli,C., François,J.C., Kurfurst,R., Montenay-Garestier,T., Saison-Behmoras,T., Thuong,N.T. and Hélène,C. (1991) *Proc. Natl. Acad. Sci. (USA)* **88**, 6023–6027.
19. Klysik,J., Rippe,K. and Jovin,T.M. (1991) *Nucleic Acids Res.* **19**, 7145–7154.
20. Thiele,D. and Guschlbauer,W. (1973) *Biophysik* **9**, 261–277.
21. Zimmerman,S.B., Cohen,G.H. and Davies,D.R. (1975) *J. Mol. Biol.* **92**, 181–192.
22. Borisova,O.F., Golova,Yu.B., Gottikh,B.P., Zibrov,A.S., Il'icheva,I.A., Lysov,Yu.P., Mamayeva,O.K., Chernov,B.K., Chernyi,A.A., Shchyolkina,A.K. and Florentiev,V.L. (1991) *J. Biomol. Struct. Dynam.* **8**, 1187–1210.
23. Sussman,J.L., Holbrook,S.R., Warrant,R.W., Church,G.M. and Kim,S.-H. (1978) *J. Mol. Biol.* **123**, 607–630.
24. Kim,S.-H., Sussman,J.L., Suddath,F.L., Quigley,G.J., McPherson,A., Wang,A.H.J., Seeman,N.C. and Rich,A. (1974) *Proc. Natl. Acad. Sci. (USA)* **71**, 4970–4974.
25. Westhof,E. (1992) *Nature* 358, Scientific Correspondance 459–460.
26. Saenger,W. Principles of Nucleic Acid Structure (1984) Springer-Verlag Berlin.