

Even nucleic acids with 2',5'-linkages facilitate duplexes and structural polymorphism: Prospects of 2',5'-oligonucleotides as antigene/antisense tool in gene regulation

V. Lalitha and N. Yathindra

Department of Crystallography and Biophysics, University of Madras, Madras 600 025, India

Nature of sugar-phosphate linkage (3',5' or 2',5') in DNA is found to influence differently the sugar-pucker dependent intrachain phosphate ... phosphate separations. A new concept of 'compact' and 'extended' repeat nucleotide is introduced to unify DNA duplex generation of both linkages. Adaption of this concept results in stereochemically favourable canonical 2',5'-DNA duplexes (A, B & Z) and triplexes. Resemblance of mixed strand duplexes and triplexes to regular 3',5'-isomers together with the high resistance of 2',5'-linkages to nucleases advocate their potential utility as antisense agents. Topological similarities of 2',5'-duplexes as well as exhibition of structural polymorphism akin to 3',5'-DNA, therefore, leaves open the question of Nature's choice of 3',5'-linkages for information storage and transfer, to reasons other than the helix-forming ability *per se*.

THE sugar-phosphate linkage in nucleic acids can be of either 2',5' or 3',5' type. The former occurs rarely in nature, during intron splicing¹ and also in interferon-treated cells^{2,3}. Although it is the most abundantly formed linkage under simulated prebiotic conditions⁴, the preference of 3',5'-linkages in nucleic acids is attributed to the inability of 2',5'-linkages to promote duplex structures^{5,6}. A critical examination of the stereochemistry of 2',5'- and 3',5'-linked nucleoside diphosphates reveals a hitherto unrecognized inverse relationship between nucleotide geometry and nature of phosphodiester linkage. A C3' *endo* sugar leads to an extended nucleotide backbone with 2',5'-linkage, in sharp contrast to a compact backbone with 3',5'-linkage. A similar contrasting behaviour is found with respect to the other commonly occurring C2' *endo* sugar pucker. Recognition and adaptation of these distinguishing stereochemical features readily facilitate Watson-Crick paired duplexes of the types feasible with 3',5'-linkages as also with 2',5'-linkages. They exhibit topological features and conformational polymorphism similar to their 3',5'-linked helices and most importantly also

facilitate drug binding in the minor groove and protein ligands, as well as triplex formation along the major groove in the extended B-type helix. The equivalent RNA duplex (3'-oxy) favours only the compact A-type duplex just like its 3',5'-linked RNA helix. Therefore, factors other than the presumed inability to form duplexes may be responsible for the absence of 2',5'-linkages in nucleic acids. Stereochemical permissibility of 2',5'-duplexes endowed with high resistance to hydrolysis by most of the nucleases⁷ brings out their prospective role as antigene and antisense tool.

While the propensity of 3',5'-linkages to readily promote duplexes is well documented, no comprehensive evidence exists in relation to 2',5'-linkages. A few reports based on experimental⁸⁻¹⁴ and theoretical investigation^{15,16} provide contrasting evidence. Early CD and NMR studies^{5,9,11} on di- and tri-nucleoside phosphates gave indications that 2',5'-phosphodiester linkages form compact structures similar to 3',5'-linkages. Several 2',5'-oligoribonucleotides are shown¹² to associate into Watson-Crick paired antiparallel duplex structures, although with a much lower stability compared to the corresponding 3',5'-linked duplexes. Recently, it has been concluded¹⁴ that 2',5'-oligodeoxynucleotides do not associate to form a duplex; instead, they associate to form triple helix at high salt concentrations. Crystallographic studies carried out on a couple of self-complementary 2',5'-dinucleotide monophosphates have been on acid forms of A⁺pU⁸ and GpC⁺¹³, which obviously do not facilitate Watson-Crick paired duplex. Interestingly, a parallel acid duplex is found to be formed by CpA¹⁷, similar to its 3',5' counterpart¹⁸. Theoretical studies, on the other hand, have suggested the possibility of mini but not extended polymeric duplexes¹⁵, and a hybrid possessing features of both A- and B-type is proposed¹⁶ instead. With a view to elucidate helix-forming ability, or the lack of it, of 2',5'-linked polynucleotides and to gain further understanding of nature's preference for 3',5'-linkages, we report here the results of model building efforts that incorporate certain stereochemical features intrinsic to

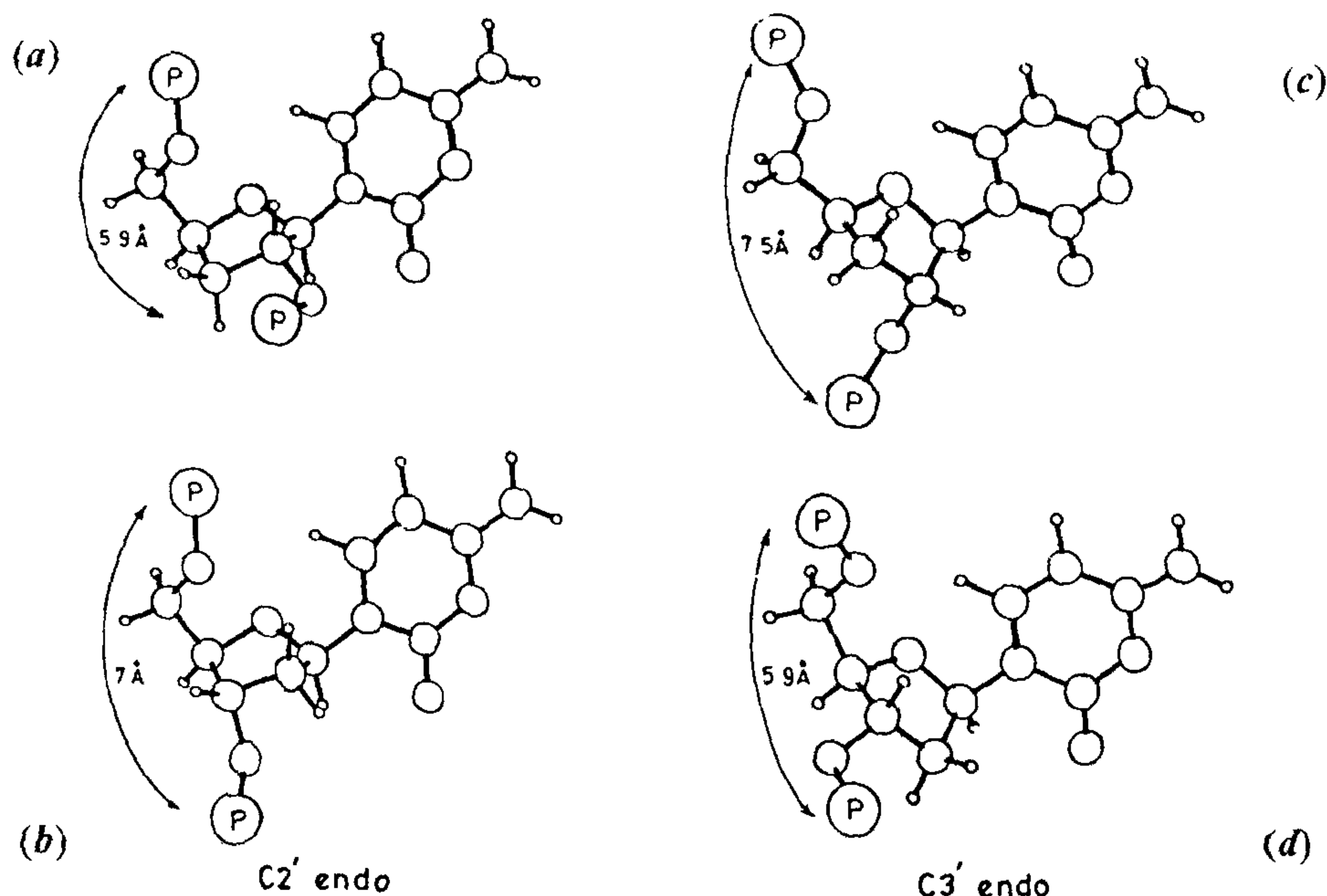


Figure 1. Structure of nucleoside 2',5'-diphosphate (*a*, *c*) and nucleoside 3',5'-diphosphate (*b*, *d*) for C3'endo and C2'endo sugar pucker, respectively. Note the inverse relationship between the sugar pucker dependent P...P separation (compact vs extended) and the phosphodiester linkage (2',5' vs 3',5'). All the C-O bond torsions correspond to 180°.

their repeating nucleotides that are different from the naturally occurring 3',5'-linked nucleic acids.

Figure 1 shows the molecular structure of 2',5'- and 3',5'-nucleoside diphosphates for the two predominantly occurring sugar ring conformations, viz. C3'endo (³E) and C2'endo (²E). A comparison readily reveals that there are fundamental differences in the shape of nucleotides, manifested in the adjacent P...P separations between the two systems. With 2',5'-linkage, a C2'endo (Figure 1*a*) sugar produces a compact nucleotide (P...P = 5.9 Å), while the same sugar generates an extended nucleotide with a P...P separation of 7 Å with 3',5'-linkage (Figure 1*b*). Likewise, a C3'endo sugar (Figure 1*c*) develops into an extended nucleotide (P...P = 7.5 Å) with 2',5'-linkage and a compact one (P...P = 5.9 Å) with 3',5'-linkage (Figure 1*d*). These conspicuously contrasting behaviours, although appearing to be trivial, have not been recognized and are expected to influence profoundly the polymer structures generated.

It is well known that in nucleic acids, extended B-type helix (pitch = 34 Å) is obtained by the repeat of extended C2'endo nucleotide (Figure 1*b*), while compact A-type helix (pitch = 28 Å) is formed by the repeat of compact C3'endo nucleotide (Figure 1*d*). In analogy with this, dimer duplexes of both A- and B-type possessing 2',5'-phosphodiester linkages have been

modelled. Initially, dimer base pairs as found in extended (B-DNA) and compact (A-DNA) helices¹⁹ are generated. To these dimer base pairs, appropriate repeat nucleotides, either C2'endo or C3'endo nucleotide are appended, depending upon the nature of the helix (extended or compact) to be generated. The dimer duplexes are subjected to constrained energy minimization using the program AMBER²⁰ (which has been suitably modified to handle 2',5'-linkages), for obtaining optimal geometry at the 2',5'-phosphodiester linkages. It is seen from Figure 2*a* that a compact C2'endo nucleotide repeat (Figure 1*a*) readily fits in between the base pairs and enables a natural course for the 2',5'-linked sugar-phosphate backbone. The base pairs in the dimer duplex have a slide of the same magnitude as in the compact A-type duplex¹⁹ and are related by a helical twist of 32.7° and a rise of 2.56 Å. Hence, the stacking pattern and other base pair properties are identical to it. Attempts to fit in extended C3'endo nucleotide (Figure 1*c*) result in serious steric hindrance as well as drastic positional variations in the base pairs. While this will not be surprising if the inherent stereochemistry of 2',5'-linked nucleotides described above is borne in mind, it might appear as interchange of sugar pucker if semantics of sugar-pucker-based definition of A- and B-type helices alone are strictly adhered to.

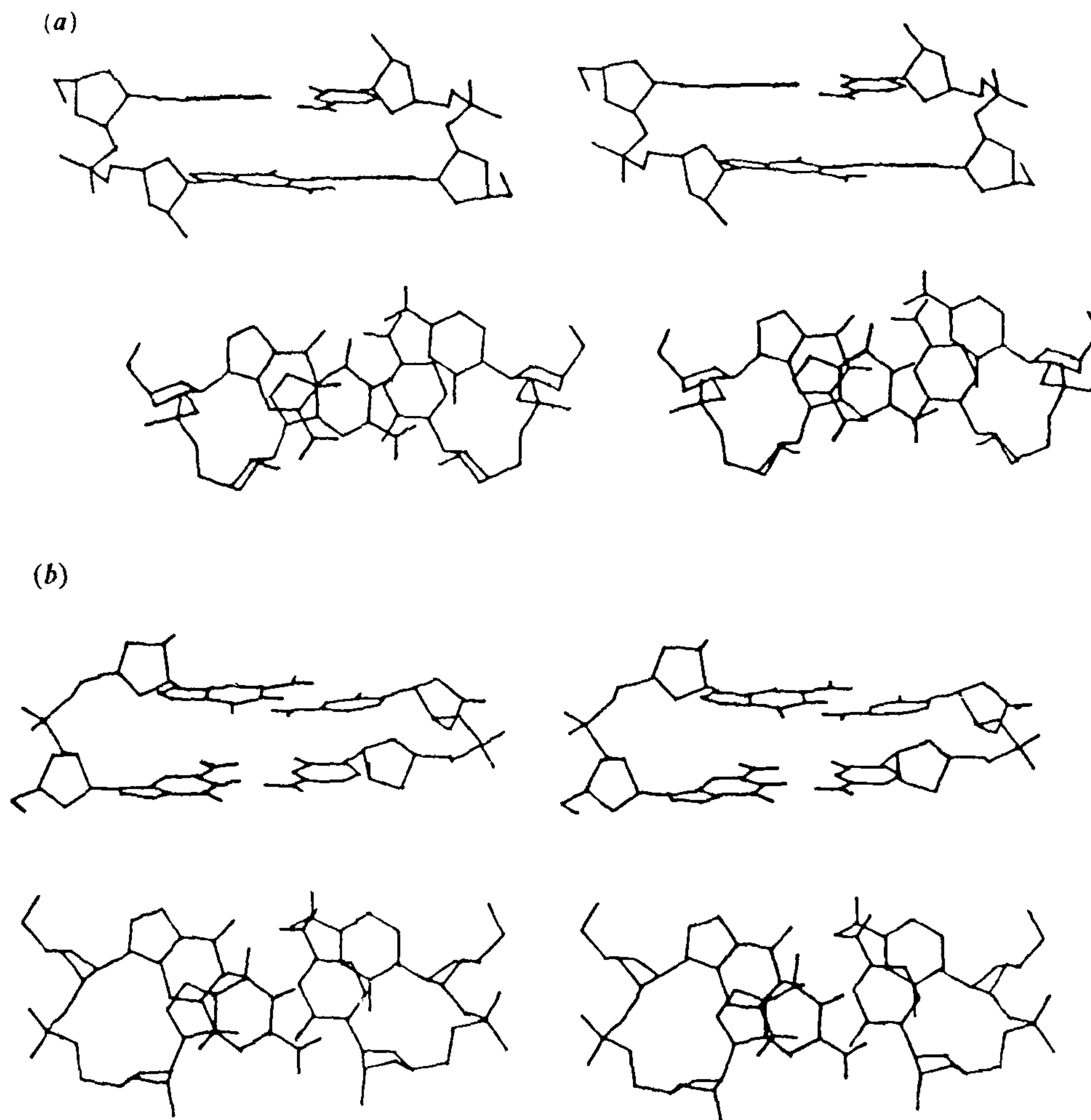


Figure 2. Stereo diagrams of 2',5'-linked CG dimer duplex (a) Perpendicular and parallel views of the B-type dimer duplex generated using 'extended' C3'*endo* nucleotide related by a twist of 36° and a rise of 3.34 \AA (b) Perpendicular and parallel views of the A-type dimer duplex obtained using 'compact' C2'*endo* nucleotide related by a twist 32.7° and a rise of 2.56 \AA

A 2',5'-dimer duplex of the B-type (Figure 2b) obtained likewise by adapting an extended C3'*endo* nucleotide (Figure 1c) (as opposed to extended C2'*endo* nucleotide used in 3',5'-linked case) is shown in Figure 2b. The base pairs here have identical relationship as in standard B-DNA¹⁹ (3',5'-linked) and are related by a twist of 36° and a rise of 3.4 \AA . It is to be noted that sugar-phosphate chain conformation on either strand is identical in both the dimer duplexes. Torsion angles around the backbone and glycosyl bonds essentially remain in the preferred range in both the

dimer duplexes. The phosphodiester torsion angle, C2'-O2'-P-O5' is nearly *trans* in A-duplex, and *gauche*⁻ in B-duplex, while the other phosphodiester torsion O2'-P-O5'-C5' is *gauche*⁻ in both the duplex forms. Despite the fact that the 2'-phosphate is located a little inside compared to 3'-phosphate, it does not preclude the formation of duplex dimers of both A- and B-type.

Convinced about the stereochemical feasibility of 2',5'-dimer duplexes of both B- and A-types, their polymeric forms comprising at least one full turn have been generated in a similar manner to compare their

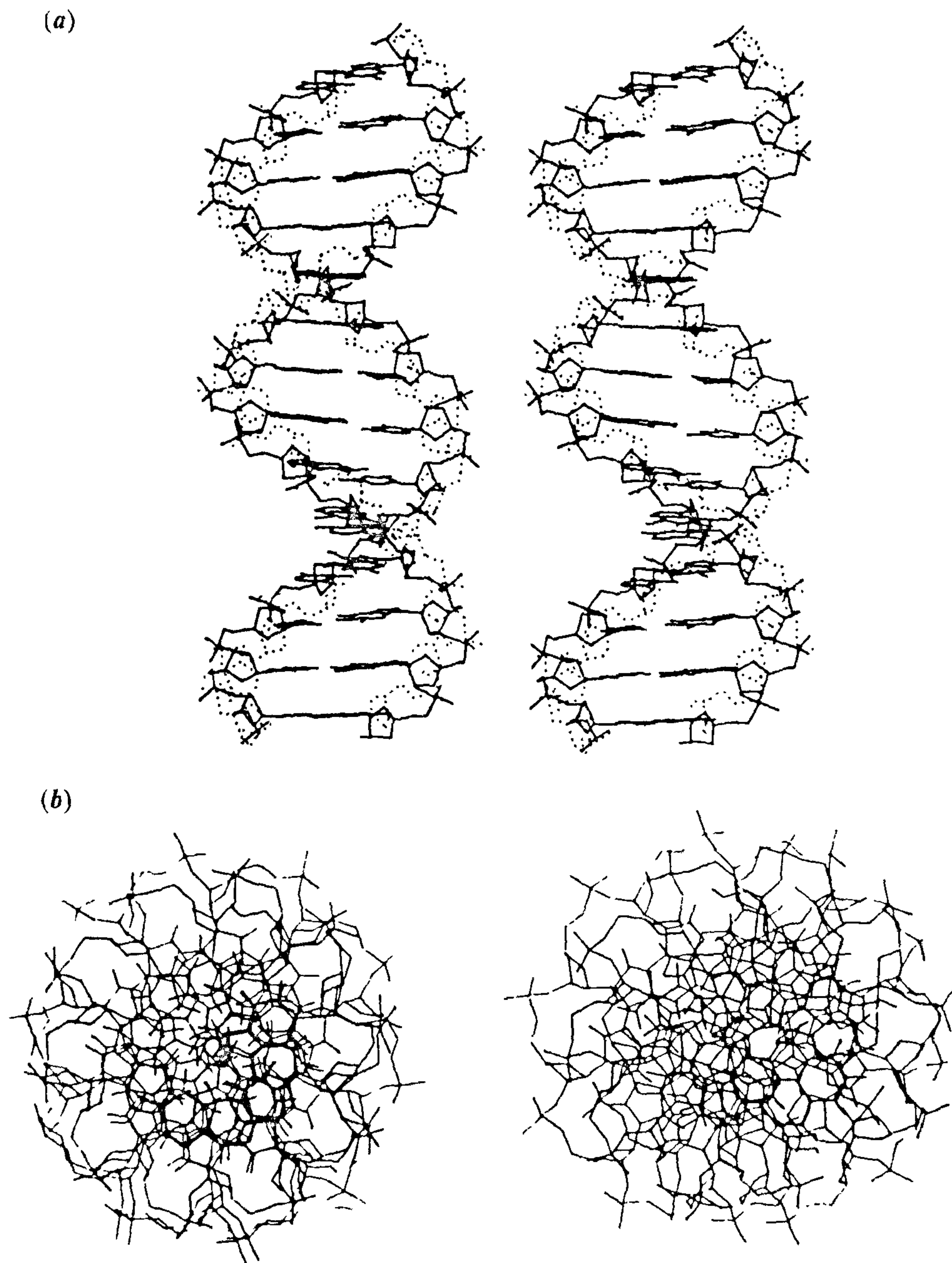


Figure 3. Stereo view of more than one full turn of 2',5'-linked B-DNA viewed (a) perpendicular and (b) parallel to helix axis. Equivalent 3',5' B-duplex (shown in dotted lines) is superposed to highlight similar base pair arrangements

topological properties with their 3',5' counterparts. The 2',5'-linked B-type and A-type duplex structures viewed parallel and perpendicular to the helical axis are shown in Figures 3 and 4. In one of the views the 3',5'-duplex

is superposed to highlight the similarity between 2',5'- and 3',5'-linked duplexes. The above polymeric 14-mer B- and A-type duplexes when subjected to unconstrained minimization show that they are indeed

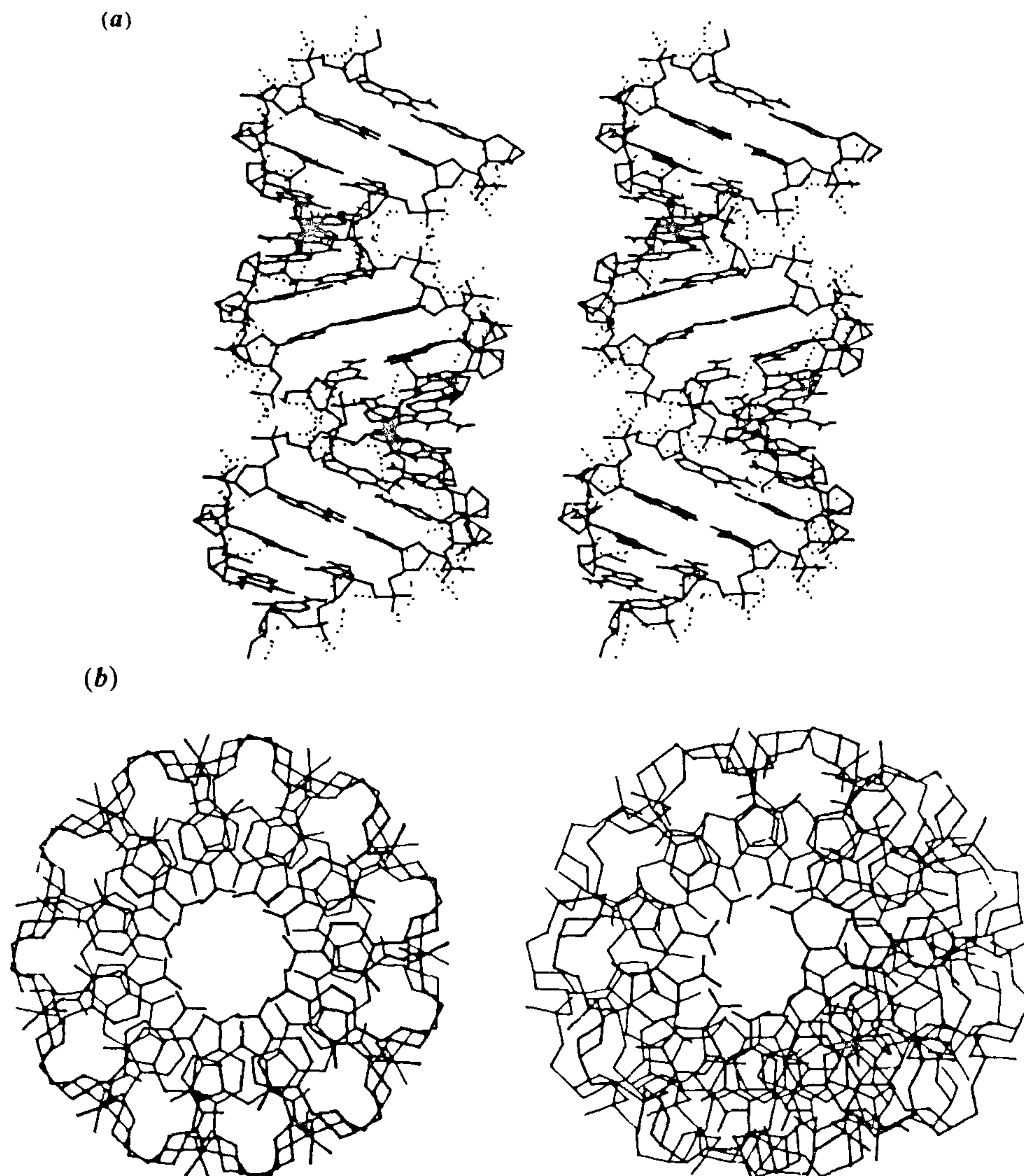


Figure 4. Stereo view of more than one full turn of 2',5'-linked A-DNA viewed (a) perpendicular and (b) parallel to the helix axis. Equivalent 3',5' A-duplex (shown in dotted lines) is superposed for comparison.

Table 1. Conformational parameters* of repeat nucleotides in 2',5'-linked A- and B-DNA duplexes

Type	α	β	γ	δ'	ϵ'	ζ	χ	Φ	T_m
B-DNA	286	170	48	280	168	272	216	24	41
A-DNA	296	148	57	206	183	178	225	153	40

$\delta' = C4'-C3'-C2'-O2'$ and $\epsilon' = C1'-C2'-O2'-P$

*IUPAC-IUB nomenclature is adopted

energetically favoured (-293.5 and -272.6 kcal/mol) and exhibit minor variations in the torsions. Conformational parameters that characterize the nucleotide repeat of the 2',5'-duplexes are listed in Table 1.

The 2',5'-linked B-type duplex (Figure 3a) bears a strong resemblance to the 3',5' B-DNA. The helical

twist and rise are approximately equal to 36° and 3.4 \AA , respectively. Similarity of base pair disposition here with the $3',5'$ -duplex is obvious from their complete overlap. The sugar-phosphate backbone alone deviates due to switch in the linkage and sugar puckering mode. It is clear, however, that the backbone helical course follows nearly the same path as the duplex with $3',5'$ -linkage. The major and minor groove widths of $2',5'$ -duplex correspond to 18.1 \AA and 10.2 \AA , which are comparable to the values of 17.5 \AA and 11.5 \AA found for fibre B-DNA¹⁹. The major groove has slightly expanded by about 0.5 \AA concomitant with a constriction of about 1 \AA in the minor groove. This is obviously due to slight positional variation of the phosphate group consequent to $2',5'$ -phosphodiester linkage. The narrower minor groove here should serve as an even better site for binding of typical minor groove ligands of DNA. In a similar way, the slightly expanded major groove should facilitate the triple helix formation²¹ rather easily, as in the case of $3',5'$ B-DNA, as well as binding of protein ligands.

The $2',5'$ -linked A-type duplex (Figure 4) has 11 base pairs per turn with a helical twist angle of 32.7° and a rise of 2.56 \AA . The base pairs here are related by a slide of the same magnitude as in the A-type duplex¹⁹ and as such the base pairs in both $2',5'$ - and $3',5'$ -linked duplexes exhibit total overlap. The $2',5'$ -duplex also exhibits a characteristic deep major groove and a shallow minor groove similar to $3',5'$ -linked A-DNA fibre duplex¹⁹. It should be realized that unlike B-duplex, the inside of the major groove of A-DNA is very wide and hollow; in fact, $C1' \dots C1'$ separation across is more than 17 \AA , compared to a width of about 8.2 \AA at the mouth of the major groove. Switch from $3',5'$ -linkage to $2',5'$ places phosphate groups towards the inside of the major groove. This results in expansion of the width of the major groove at its mouth from 8.2 \AA to 11.2 \AA , concomitant with narrowing, by a nearly similar magnitude, of minor groove from 17 \AA to 14.3 \AA compared to $3',5'$ -duplex. Nevertheless, these values are well within the observed range for A-DNA duplexes²². Groove widths corresponding to A-DNA fibre model can be obtained by effecting minor modifications along the sugar-phosphate backbone chain. The central hole arising from the shift of the base pairs with respect to the helical axis in A-DNA duplex is also seen here (Figure 4b). Visual inspection also brings out that the groove depths are nearly equivalent in the two helices. Other polymorphs of $3',5'$ -duplexes, within A- and B-family, obtainable by small variations in torsion angles and helical parameters can be easily realized here also by appropriate modifications along the backbone of $2',5'$ -duplexes. Our further modelling efforts reveal that proper utilization of the concept of extended and compact nucleotides (instead of adapting sugar-pucker-based definition) in the 'dinucleotide repeats' enable

even a $2',5'$ -linked left-handed Z-type structure (authors' unpublished results) similar to $3',5'$ -linked Z-DNA structures²³.

In RNAs, compressed A-type duplex alone is favoured with ribo(3-oxy) $C3'endo$ nucleotide repeat, whereas $2'$ -deoxysugars favour both B- and A-type duplexes. In view of the evidence provided above for the inverse relationship of the repeating nucleotide structure with altered phosphodiester linkages ($3',5'$ vs $2',5'$), it is important to ascertain whether similar implications extend also in relation to equivalent RNA ($3'$ -oxy) helices with $2',5'$ -phosphodiester linkage. Our analysis reveals that duplex of the compact A-type, i.e. with 'compact' $C2'endo$ repeating nucleotide (Figure 1a), alone is stereochemically favourable, whereas helices generated with 'extended' $C3'endo$ ribonucleotides (Figure 1c) as repeats result in extensive steric clashes between $3'$ -hydroxyl and phosphate groups. Thus, it is evident that helices with $2',5'$ -linkage bear an overall similarity in all its structural properties to regular $3',5'$ -linkages, especially in relation to its helical preferences.

In view of the fact that $2',5'$ -linked nucleotides form topologically equivalent helical structures analogous to their $3',5'$ -complements and that they are alien to nucleic acids and also possess higher stability for nuclease digestion⁷, equip them adequately to be used as tools in the antisense approach to gene regulation, it is, therefore, relevant to visualize duplexes and triplexes with mixed strand linkage. A duplex structure formed by complexing a $2',5'$ -strand with a complementary $3',5'$ -strand is shown in Figure 5. Similarly, formation of a triplex by complexing a $2',5'$ -strand to a regular $3',5'$ -duplex is also viable (Figure 6). In this context, serious experimental studies involving such mixed complexes should be attempted, with additional stabilizing factors, if needed, to realize the utility of $2',5'$ -oligoribonucleotide in gene regulation.

Both $2',5'$ - and $3',5'$ -linked polynucleotides can facilitate the formation of nearly equivalent Watson-Crick paired duplex structures of all known types once it is recognized that an inverse relationship exists between their repeating nucleotide moieties and the nature of the phosphodiester linkages. Essentially, interchange of sugar puckers from $C3'endo$ to $C2'endo$ and vice versa is sufficient to manifest the common features found in A, B- and Z-helices. The duplexes, although non-equivalent in relation to sugar ring conformation, are equivalent when gross stereochemical effects are considered. The results on $2',5'$ -linked duplexes suggest that duplex structures need to be described in terms of cumulative structural properties rather than semantics of sugar-pucker-based nomenclature. The observation that stereochemically feasible duplexes can be formed irrespective of whether the sugar-phosphate linkage is of $2',5'$ or $3',5'$ type raises the question why nature favours exclusively $3',5'$ -linkages and does not use



Figure 5. Stereo view of a Watson and Crick paired duplex formed by complexing a naturally occurring 3',5'-linked nucleic acid strand (yellow) with a 2',5'-linked polynucleotide strand.

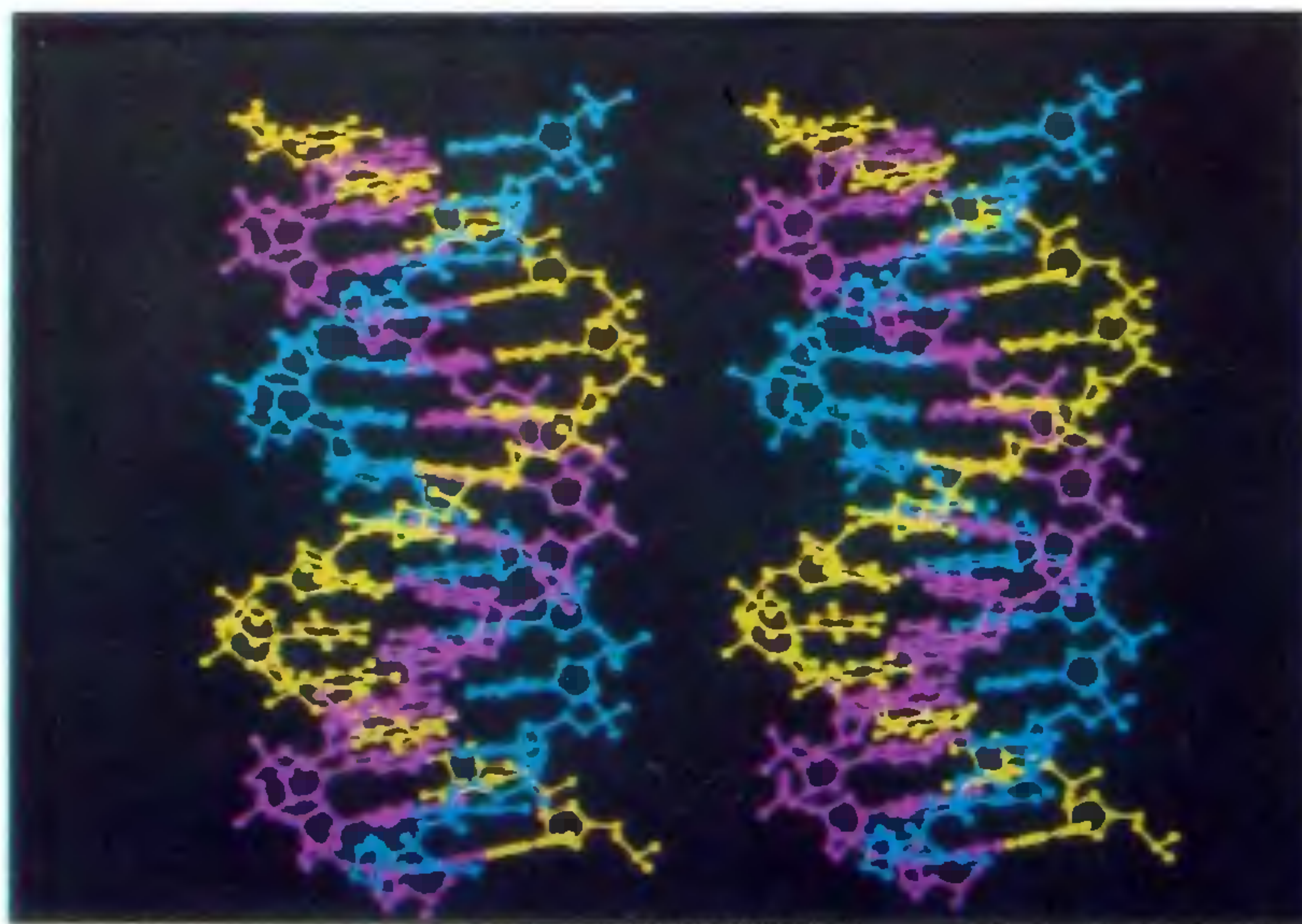


Figure 6. Stereo view of a triplex formed by complexing a 2',5'-linked thymine third strand (yellow) to a 3',5' Watson-Crick (A...T) paired duplex through Hoogsteen base pair.

2',5'-linkages. Whether the difference in duplex stability between 3',5'- and 2',5'-isomers or the greater resistance of 2',5'-linkages for most of the nuclease digestion⁷ compared to 3',5'-linkages is responsible for the latter's

selection remains to be explored. The stereochemical rationale deduced here for the formation of 2',5'-duplexes shows promise for its use in regulation of gene expression both at the level of translation (antisense

strategy) and transcription (antigene strategy) by forming duplex and triplex with the complementary isomeric 3',5'-strand.

- 1 Padgett, R. A., Kornarska, M. M., Grabowski, P. J., Hardy, S. H. and Sharp, P. A., *Science*, 1984, **225**, 898-903
- 2 Kerr, I. M. and Brown, R. E., *Proc Natl Acad Sci USA*, 1978, **75**, 256-260
- 3 Lesiak, K., Imai, J., Smith, G. F. and Torrence, P. F., *J Biol Chem*, 1983, **258**, 13082-13088
- 4 Lohrman, R. and Orgel, L. E., *Tetrahedron*, 1978, **34**, 853-855
- 5 Dhirga, M. M. and Sarma, R. H., *Nature*, 1978, **272**, 798-801
- 6 Parthasarathy, R., Malik, M. and Fridley, S. M., *Proc Natl Acad Sci USA*, 1982, **79**, 7292-7296
- 7 Ball, A. L., in *Enzymes* (ed. Boyer, P. D.), Academic Press, New York, 1982, Chap. 11, pp. 281-313
- 8 Shefter, E., Barlow, M., Sparks, R. A. and Trueblood, K. N., *Acta Cryst B*, 1969, **25**, 895-909
- 9 Kondo, N. S., Holmes, H. M., Stempel, L. M. and Ts'o, P. O. P., *Biochemistry*, 1970, **9**, 3479-3498
- 10 Doornbos, J., Charubala, R., Pfeleiderer, W. and Altona, C., *Nucleic Acids Res*, 1983, **11**, 4569-4581
- 11 Gopalakrishna, V., Ghadage, R. S. and Ganesh, K. N., *Biochem Biophys Res. Commun.*, 1991, **180**, 1251-1257
- 12 Kierzek, R., He, L. and Turner, D. H., *Nucleic Acids Res*, 1992, **20**, 1685-1690

- 13 Krishnan, R. and Seshadri, T. P., *J Biomol Struct Dyn*, 1993, **10**, 727-745
- 14 Jin, R., Chapman, Jr. W. H., Srinivasan, A. R., Olson, W. K., Breslow, R. and Breslauer, K. J., *Proc Natl Acad Sci USA*, 1993, **90**, 10568-10572
- 15 Srinivasan, A. R. and Olson, W. K., *Nucleic Acids Res*, 1986, **19**, 379-384
- 16 Anukanth, A. and Ponnuswamy, P. K., *Biopolymers*, 1986, **25**, 729-752
- 17 Krishnan, R., Seshadri, T. P. and Viswamitra, M. A., *Nucleic Acids Res*, 1992, **19**, 379-384
- 18 Westhof, E. and Sundaralingam, M., *Proc Natl Acad Sci USA*, 1980, **77**, 1852-1856
- 19 Arnott, S. and Hukins, D. W. J., *Biochem Biophys Res Commun*, 1972, **47**, 1504-1510
- 20 Weiner, S. J., Singh, U. C., Kollman, P. A. and Case, D. A., 'Molecular Mechanics and Dynamics Program - AMBER, Ver 3.0', University of California, San Francisco, USA
- 21 Sekharudu, C. Y., Yathindra, N. and Sundaralingam, M., *J Biol Mol Struct Dyn*, 1993, **11**, 225-244
- 22 Kennard, O. and Hunter, W. N., *Q R Biophys*, 1989, **22**, 327-379
- 23 Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., Van Boom, J. H., Van der Marel, G. and Rich, A., *Nature*, 1979, **283**, 680-686

ACKNOWLEDGEMENTS. V. L. thanks DAE for the award of K. S. Krishnan fellowship. We thank Bioinformatics Group, MK University Madurai, for the colour pictures.

Received 17 October 1994, revised accepted 15 November 1994

Graphical analysis of DNA sequence structure: II. Relative abundances of nucleotides in DNAs, gene evolution and duplication

A. Nandy* and P. Nandy†

*Computer Division, Indian Institute of Chemical Biology, Calcutta 700 032, India

†Department of Physics, Jadavpur University, Jadavpur, Calcutta 700 032, India

A technique for graphical representation of gene sequences on a two-dimensional cartesian coordinate system is shown to highlight visually the relative abundances of nucleotides along a DNA sequence on a global scale. In some sequences such as the rat myosin heavy-chain gene this reveals a rich structure in the DNA map. In several cases the gene sequences are shown to map to an almost uniform linear structure. The possibility that this may be due to gene evolution by gene duplication, as hypothesized by Ohno, or due to extensive repetitive segments in the sequence is discussed with reference to myosin heavy-chain genes, where the rod-encoding part is known to have a large number of repeats, and the kinetoplast genes of e.g. *L. tarentolae*, where almost the entire gene fragment is composed of six repeating sequences. This feature of the graphical repre-

sentation provides an easy analytical tool to identify the parts of a gene sequence with large repetitive segments. A comparison is made with the chaos generator diagrams of Jeffrey¹ and the two methods are shown to complement each other in the analysis of gene sequences.

ONE of the interesting problems in molecular biology concerns the interpretation of base composition and distribution in long DNA sequences. While several approaches have led to identification of small segment motifs^{2,3} such as signal sequences, the TATA box, repeats and hairpin loops, techniques for analysis of the total span of a DNA sequence still remain elusive. There has been a renewal of interest in this problem in recent times, brought about by the chaos generator technique of