A.G. JOHNSON C3a PAPER FEBRUARY 1981

# D.N.A. IS NOT ALWAYS A SIMPLE DOUBLE-HELIX

1. Introduction
2. D.N.A.:- Its functions
3. The classic structure
4. Problems with function and problems with detailed determination of the classic structure
5. An alternative structure
6. A brief discussion of some implications of this alternative structure
7. A consideration of still wider implications stemming from this work

## INTRODUCTION

The original paper outline no longer faithfully expresses the present intended content of the paper. It is now a layman's expression of the current state of play in the debate over the structure of D.N.A.

The debate is rather one-sided. While the existence of alternative conformations in different media is generally accepted these forms, i.e. P, D, B & A are minor variants of the major established Watson-Crick model. This model of the structure of D.N.A. is so well established that proposed alternatives seem to be largely ignored.

The first papers setting out radically different alternatives that I am aware of appeared in 1976 from two separate sources almost simultaneously. (2&3) The New Zealand group have played no further part and the Bangalore group, dropping the first of their two alternative structures (4), have continued to support their Type 2 structure in what appears to me to be a very comprehensive fashion; their latest paper appearing early last year, 1980.

Another alternative structure of substantial deviance was proposed last year by a group headed by Wang in Massachusets, the "Z" structure (5). Many of the features of the Type 2 structure can be accommodated in Wang's model (6) and this independent agreement would seem to demand the joining of debate. Very little is forthcoming and at present the Bangalore group are the only group doing detailed analysis of the fine structure and stereochemistry of D.N.A./R.N.A.(7) components. In fact base-stacking analysis must be one of the few fields in modern research in which the most recent previous work was published over 10 years ago (8).

D.N.A.:- ITS FUNCTIONS

Living organisms, generally, (9) are cellular. Cells are differentiated and are identified by their products. These products consist of waste-products of metabolism and protein-based materials synthesised within the cell. The proteins are all composed of sequences of amino acids of which there are 21. The different shapes and chemical properties of proteins are determined by the precise sequence of their amino acids.

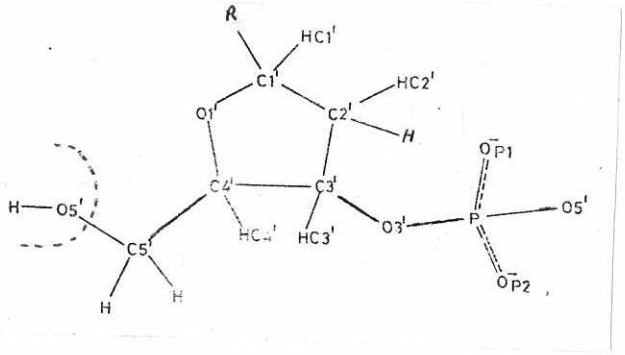
D.N.A. is the material which at root controls protein production. D.N.A. is composed of sugar-phosphates and 4 bases which lie in sequence within the molecule. The base sequences are not random but are arranged in triplets. The 4 bases can be arranged in threes 64 different ways and each "code" triplet relates, via a series of intermediaries, to an amino acid or a stop-end. Some of the amino acids have more than one code but each code is specific. Hence the original D.N.A. base sequence determines the all-important amino acid sequence and thus the specific proteins produced by the cell.

D.N.A. is at the same time the hereditary material of almost all living organisms (7). The passage of identical triplet sequences to subsequent generations ensures continuity of function for the cell line and for the macro-organism via sex-cells. Each cell's D.N.A. carries all the codes necessary for the correct functioning of all the cells in the mature organism as well as for their functioning during its development.

Mutation and mixing notwithstanding, in both major functional roles, protein-synthesis and inheritance, D.N.A. must maintain its sequences precisely. For this reason its structure must be stable. At the same time it must be able to split apart to provide templates for R.N.A. and to replicate for cell reproduction. Its stability must be limited in the normal conditions of pH, temperature, and hydration pertaining in most living organisms (10 & 11). The precise balance between these two requirements must be inherent in its structure, its physico-chemical make-up, or its stereochemistry. If this dualism is not inherent and is merely "imposable" by the action of various reagents then I would doubt that D.N.A. could perform the many functions it does with the subtlety and flexibility which living systems require. A model for D.N.A. structure which shows the molecule to be too stable and to have great internal regularity fails to faithfully represent the in vivo situation. This is because it demands of the living cell energy expenditure for strand separation via overall changes in the D.N.A.'s chemical medium which must be at odds with the fine and subtle responses which D.N.A. obviously displays.

## THE CLASSIC STRUCTURE

D.N.A. is composed of two major types of chemical compound, the sugar-phosphate complex and the 4 bases, 2 purines and 2 pyrimidines. The sugar-phosphates are joined in sequence giving a repeating chain. In R.N.A. there is one only and in D.N.A. there are two which are arranged in anti-parallel, i.e. one chain points one way w.r.t. the orientation of the sugar, the other being reversed (see Figs. 1 & 4b)




In this diagram we can see the component unit of the D.N.A. backbone. The phosphate group is attached at the C3' and the next group will link via the C5'. The base is represented by R at the C1'

**Figure 1.**

The chain is bonded covalently and thus has high tensile strength. The bases are covalently joined to each sugar at the C1 to make the glycosidic bond (see Fig. 2).

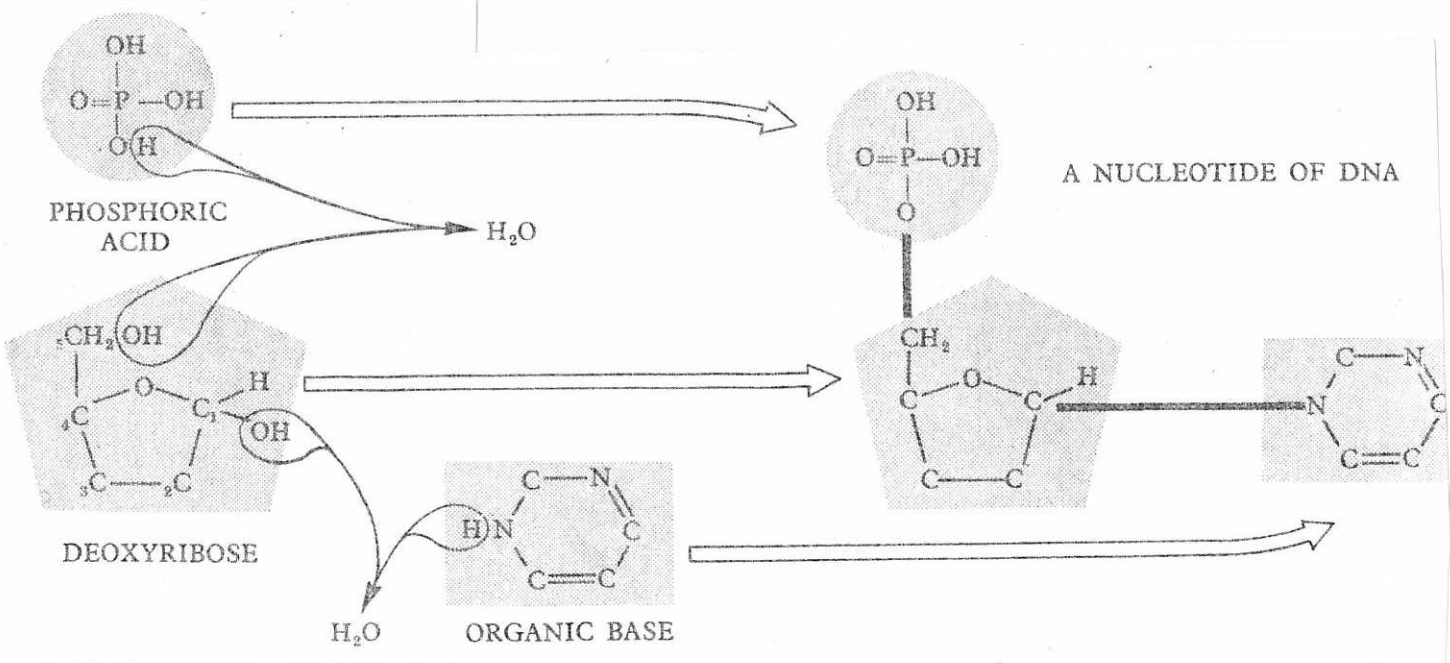
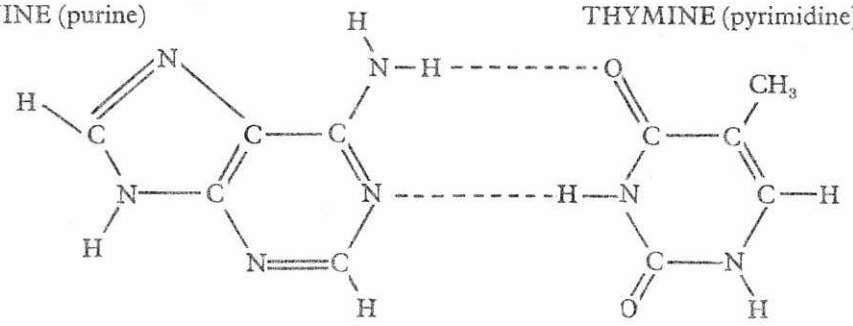



**Figure 2**

The bases align as shown in Fig 3 with specific partners and by in-plane hydrogen-bonding, a much weaker type of force than covalent bonds, hold the two chains together. Analysis combining X-ray diffraction data and bond length determination for hydrogen-bonding give a picture of D.N.A. shown schematically in Fig 4 and more comprehensively in the space-filling version version - Fig 5.

ADENINE (purine) THYMINE (pyrimidine)



GUANINE (purine) CYTOSINE (pyrimidine)

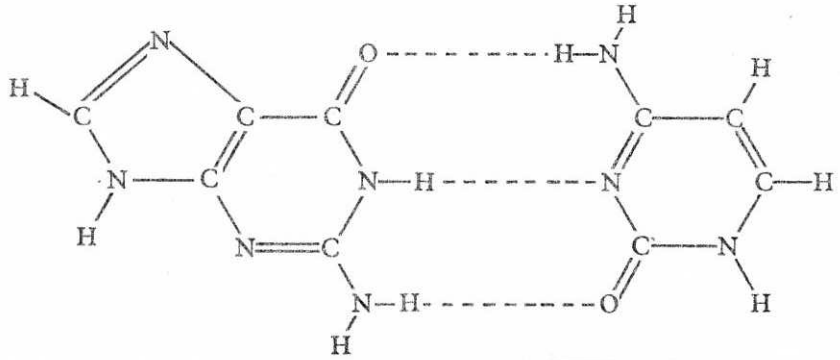
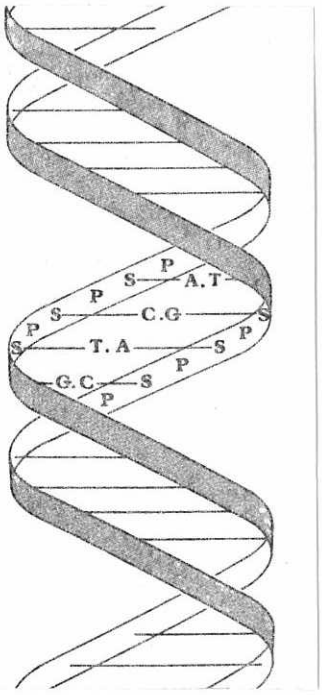


Figure 3



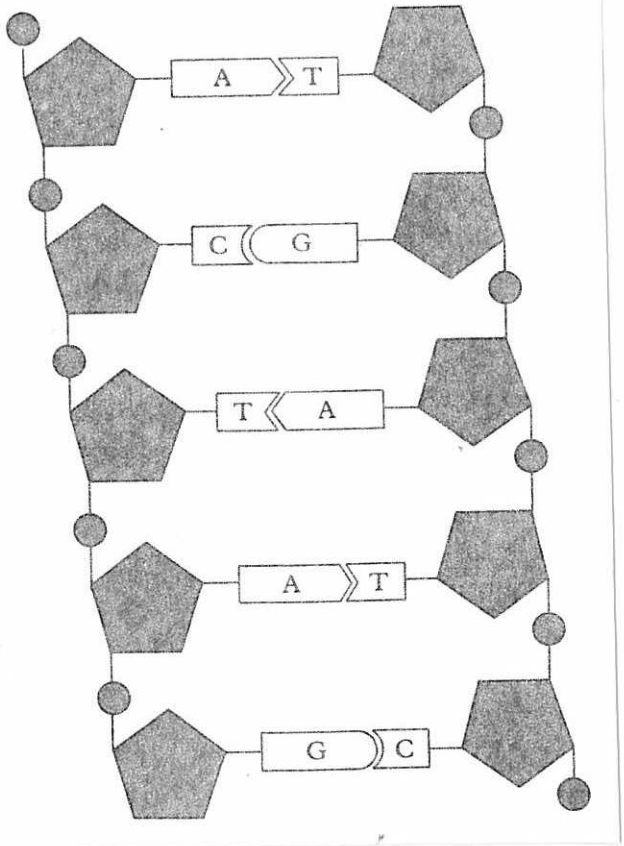




Figure 4

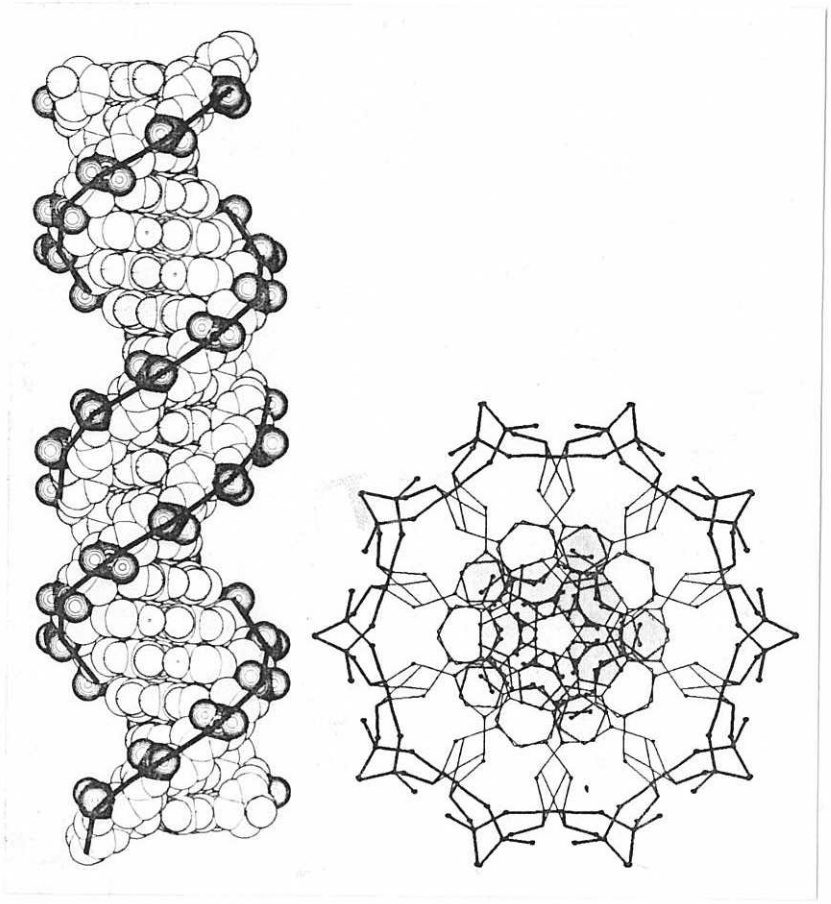


Figure 5

PROBLEMS WITH FUNCTION AND PROBLEMS WITH DETAILED DETERMINATION OF THE CLASSIC STRUCTURE

There are a number of problems with the double-helix structure normally accepted. Very generally, as mentioned above the structure must have inherent "dualism", being stable to maintain the integrity of its base sequences while being able to make these readily and selectively available for template formation and replication. Conventional explanations point to the role of the histones and other D.N.A.-associated proteins in controlling and regulating D.N.A. cleavage. I don't find this model totally satisfying. Their role has been demonstrated (17) but there are life forms, phages and viruses, with little or no associated material. Indeed this lack is part of the explanation put forward for their ability to survive in adverse conditions. Whatever position one takes over what is or what is not a life form and however one views the development of living systems it seems unlikely for a system to develop which requires complex associations for the performance of its most elemental synthetic functions. In "immediate" structural terms the most stable, regular D.N.A. might appear most energetically feasible while in a more whole sense its need for associated compounds makes it less attractive.

In more precise terms there are still problems among which there are some which give a lie to the Watson-Crick model being in any sense the most energetically feasible model for the structure of D.N.A. These criticisms stem from work done on the stereochemistry of stacked bases (11) and (8) with which we shall deal in the next section.

I find it difficult to see how a structure coiled around itself with two "backbones" would accommodate the normal movements and resonance effects which must be expected at up to 40 degrees C. That is to say that bends in the molecule will effectively alter the alignment of the two sides of the chain putting the weak hydrogen-bonding under stress and continually breaking the bonds. These are easily remade in the centre but at the ends, in a hydrated medium this may not be so easy due to the bonding of water molecules. There is the associated question of tight spiralling in which the coil tightens beyond the normal and so cannot release its bases for template-forming or replication at all. (This is not supercoiling).

Supercoiling is another observed phenomenon which the normal model does not fully elucidate. Circular D.N.A. supercoiling as seen in plasmids is not such a problem (12). Linear D.N.A. supercoiling as seen in conjunction with histones i.e. as chromatin has not been dealt with very convincingly (13). This is a very important area. The conventional model is not flexible enough to take up such conformations without tight-spiralling (14). The Bangalore group Type 2 structure is the only structure that I know of from the literature that can accommodate the observed supercoiling (15).

This is the only group that is doing any detailed work on the fine stereochemistry of the nucleic acids, R.N.A. as well as D.N.A. It is to the work of this group to which we now turn.

## AN ALTERNATIVE STRUCTURE

The Bangalore group is a small group consisting , V. Sasisekharan, Goutam Gupta, and N. Pattabiraman. The order in which I deal with their work is not the order in which it was published. It is the best order for making the features clear, I hope.

The essence of the Type 2 structure can best be grasped by comparison to the double-helix. Imagine two cylinders. On the first trace a spiral, followed by another, close to it. This is the double-helix. On the second, trace a sinusoidal wave. Then, on the other side, trace its mirror wave. These are the lines followed by the sugar-phosphates in the alternative structure.

Except for work by Pullman (8) crystallographic data on crystal structures of nucleic acid constituents had been gathered only from mononucleosides and mononucleotides. By using information from di-, tri-, and tetranucleosides and nucleotides a much more realistic picture can be evinced. This is because the conformations then produced come from crystals in which base-stacking has played a part. The base-stacking is held by this group to play the major part in determining the secondary structure of D.N.A. and R.N.A.

Briefly their argument runs as follows. There is great flexibility in the orientation of the sugar-phosphate chains in view of the large degrees of rotational freedom around the phospho-diester bonds (16). Therefore these chains resemble 'everyday' chains having high tensile strength due to the strong covalent bonding but with great lateral freedom. Imagine two chains hanging free. The bases bond covalently to the sugars and then bring the two chains together, as in the conventional model. The difference here though is that while the hydrogen bonds hold the two chains together as in the conventional structure the great variety in stacking between consecutive bases greatly effects the conformation of the molecule as a whole. Evidence from both the energetics of certain base stacking and certain parameter conjunctures in sugar phosphate conformations dictate that the chain kinks at various points thus altering the spiral sense and giving rise to left-handed sections, along which the bases and the sugars are inverted in relation to those in the right-handed section. To agree with the established chain parameters for the helix (n=the number of nucleotide residues per turn = 10+1, h=height per residue = 3.4Å approx), the sections alternate giving a sinusoidal/near-sinusoidal chain, roughly 5 bases per section.

The Bangalore Type 2 structure is referred to henceforth as the Right-Left (RL) structure. (see Fig 6 below).

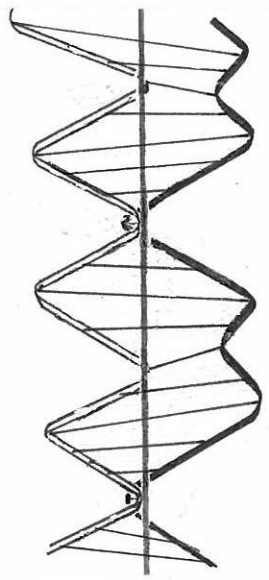


Figure 6

The Alternative 'Backbone'

We now turn to the evidence, starting with the studies of the backbone, working mainly from a paper published in Nature in 1978 (volume 275 pp 159-162). "Structure of D.N.A. predicted from stereochemistry of nucleoside derivatives". (16)

This work was done on di, tri, and tetranucleoside phosphates. Torsion angles about the different bonds were measured. (See Table I & refer to Fig 1) In the table Sets I, II, and III, refer to the parameters of three main conformations. Set IV serves to compare the parameter values proposed by groups whose structural types are shown inColumn 1.

The sugar and nucleoside components are observed to be similar to those seen in mononucleosides and mononucleotides. Sugar puckering was C3-endo or C2-endo. That is to say that the furanose ring has these two major conformations, analogous to the 'chair' and 'boat' forms in cyclohexane rings. In furanose the 'sharp' angle comes at the C3 or C2 respectively. $\zeta = 90^\circ$ denotes C2-endo, $\zeta = 140^\circ$ denotes C3-endo. ($\zeta$ = the C4-C3 bond angle but is responsible for this shift about C3 & C2.)

The C4-C5 and C5-O5 angles are of constant type gg & trans respectively although precise values varied a little, (columns 7 & 6 respectively). It should be remembered later that C4-C5 can form gt & tg as well, these will be discussed a little and they appear in Fig 7.

TABLE I

Backbone torsion angles for the fragments from dinucleoside monosphates and related molecules

| | | $\zeta(5')$ C4'-C3' | $\alpha$ C3'-O3' | $\beta$ O3'-P | $\gamma$ P-O5' | $\delta$ O5'-C5' | $\varepsilon$ C5'-C4' | $\zeta(3')$ C4'-C3' | Ref. |
|---|---|---|---|---|---|---|---|---|---|
| Set I | GpC | 89 | 211 | 292 | 285 | 184 | 50 | 77 | 1 |
| | GpC1 | 79 | 222 | 294 | 291 | 181 | 47 | 79 | 2 |
| | GpC2 | 73 | 217 | 291 | 293 | 172 | 57 | 80 | 2 |
| | GpC3 | 96 | 224 | 290 | 286 | 167 | 63 | 74 | 2 |
| | GpC4 | 88 | 216 | 288 | 283 | 181 | 52 | 87 | 2 |
| | ApU1 | 84 | 213 | 293 | 288 | 177 | 57 | 74 | 3 |
| | ApU2 | 78 | 221 | 284 | 295 | 168 | 58 | 77 | 3 |
| | $ApA^+$ | 82 | 223 | 283 | 297 | 160 | 53 | 81 | 6 |
| | ApT*(1) | 90 | 213 | 294 | 293 | 176 | 68 | 134 | † |
| | ApT*(2) | 83 | 212 | 284 | 302 | 171 | 64 | 139 | † |
| Set II | UpA(2) | 77 | 224 | 164 | 271 | 192 | 54 | 93 | 4 |
| | TpT* | 157 | 252 | 163 | 288 | 187 | 41 | 158 | 5 |
| | TpA* | 134 | 204 | 168 | 286 | 186 | 49 | 83 | † |
| Set III | UpA(1) | 86 | 206 | 81 | 82 | 203 | 55 | 85 | 4 |
| | $A^+pA^+$ | 81 | 207 | 76 | 92 | 186 | 56 | 79 | 6 |
| Set IV | A RNA | 95 | 202 | 294 | 294 | 186 | 49 | 95 | 14 |
| | A DNA | 83 | 178 | 313 | 285 | 208 | 45 | 83 | 15 |
| | B DNA | 156 | 155 | 264 | 314 | 214 | 36 | 156 | 16 |
| | C DNA | 141 | 211 | 212 | 315 | 143 | 48 | 141 | 17 |
| | D DNA | 156 | 141 | 260 | 298 | 208 | 69 | 156 | 18 |
| | B DNA (left) | 140 | 200 | 212 | 282 | 165 | 40 | 140 | ‡ |

* Deoxy sugar.
† M. A. Viswamitra, personal communication.
‡ Present work.

The major degrees of freedom are about the phosphodiester bonds O3-P & P-O5, and the forms taken up by these depend on the precise conjunction of forms taken up by the other bonds, i.e. $\zeta$ being 90° or 140° determines $\beta$ as g- or t i.e. in the range around 290° or 160°; and C4-C5 as gg, gt, or tg makes P-O5, or $\gamma$, $g^-$, t, or $g^+$ (See Fig 7)

It will be noticed that the exception to C4-C5 gg giving P-O5 as $g^-$ comes in Set III where it gives P-O5 at 82° & 92°. This represents a bend or fold conformation of previously mentioned importance.

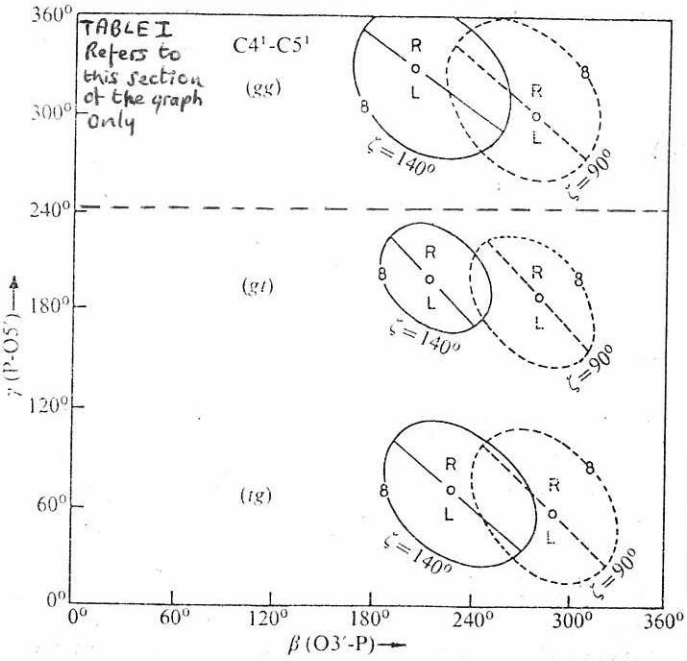
Fig 7 shows the graphical representation of the data relating to the effect of the different forms of C4-C5 on the phosphodiester groups. It shows helical domains, sets demonstrating the conjunction of values for $\beta$ & $\gamma$ which give a helical structure. These domains are 'false' since while n=8 is possible, h-0 is obviously not. However, what is shown is that right and left helical senses are taken up with very small changes in values of $\beta$ & $\gamma$. These values are mostly determined by changes in $\zeta$ the C4-C5 bond as is also shown.

Figs 8 & 9 use a similar principle of domains, expanding the scope of application to give a graph capable of comparing the helical viability of a number of D.N.A. types and nucleoside derivatives. In these $\alpha$, C3-O3 was plotted against $\beta$, O3-P, for the observed values from the data. $\gamma$, $\delta$, & $\varepsilon$ were kept constant and $\zeta$ was shown for its two values by using the graph in Fig 8 for C2-endo and the graph in Fig 9 for C3-endo.

Since h-0 divides the domains into Left and Right helical sections and an iso-h contour subdivides this, roughly at right angles, 4 sections were produced shown as RI & LI, RII & LII.

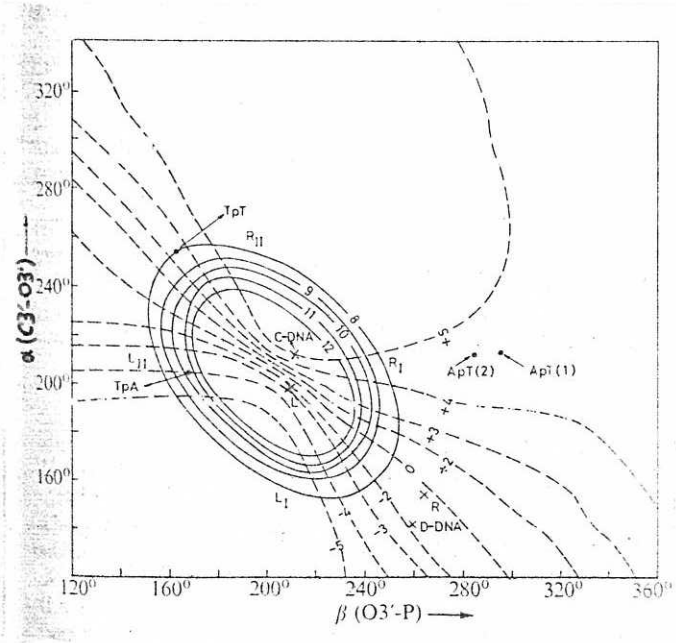
Only some of the structures in these domains give double-helical structures with viable base-pairing. Hence in the C2-endo case, ApT(1) & ApT(2) cannot form double-helixes at all. Their n-values are too small. (N.B. They can in C3-endo). In RII & LII the bases are too far from the axis and so cannot form double-helixes. This notwithstanding these conformations may well have some bearing on the structure of other nucleic acids i.e. m&t R.N.A.

In the C3-endo case, Fig 9, RI & LI give structures with n & h-values very similar to B D.N.A. This is not to say that they refer to this specifically but to forms with the same parameters and which would be hard to separate from B D.N.A. in X-ray diffraction analysis. See Figure 22.



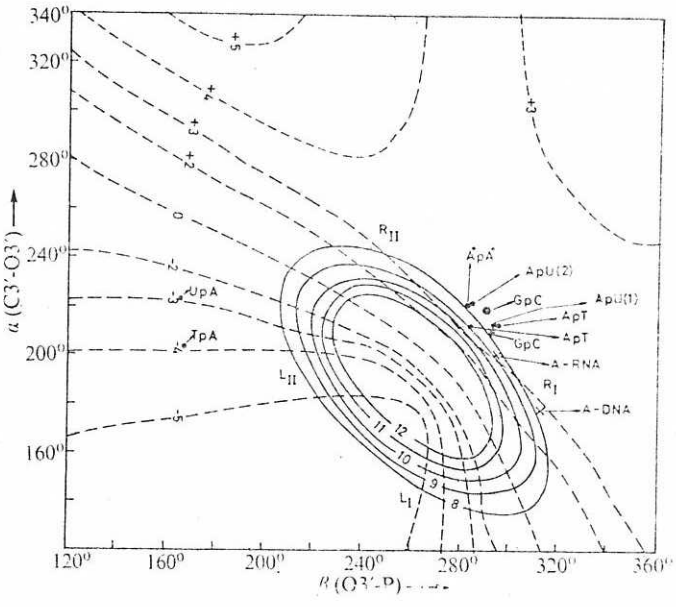

Helical domains in the $\beta$–$\gamma$ space corresponding to the conformations *gg*, *gt* and *tg* about the C4'–C5' bond, for C2' *endo* (solid line) and C3' *endo* (dashed line) sugar puckering. Only $n = 8$ and $h = 0$ are shown. Right and left helical sections are denoted by R and L.

**Figure 7**




Helical domains in the $\alpha$–$\beta$ space for C2' *endo* puckering ($\zeta = 140°$). $n$ (solid line) and $h$ (dashed line) contours and observed torsion angles $\alpha$ and $\beta$ of the fragments with C2' *endo* puckering for the sugar are shown. R and L denote right- and left-handed B DNA.

**Figure 8**




Helical domains in the $\alpha$–$\beta$ space for C3' *endo* puckering ($\zeta = 90°$). $n$ (solid line) and $h$ (dashed line) contours and observed torsion angles $\alpha$ and $\beta$ of the fragments with C3' *endo* puckering for the sugar are shown. For GpC (Ca) the plotted conformation refers to the average values of the four independent molecules in the crystal.

**Figure 9**

Base Stacking and Secondary Structure

Base stacking analysis is of predominant importance to this group and I demonstrate this by quoting part of their first paper (3)

> "As the Watson-Crick base pairing has been maintained in our models, the essential difference between our two structures and the Watson-Crick model lies in the mode of stacking of bases." (Page 780)

The details of the base stacking were investigated by the group after the work of the groups of Pullman, De Voe, and Langridge, and reported in 1978 (11). This material appeared in Nucleic Acids Research vol. 5(2) in consecutive articles on pages 1639 & 1655. The first deals with stacking in free bases, the second with stacking in polynucleotides.

The stacking energies are component and are the sum of the following energies.

1. Monopole-monopole interaction

   Some bases have polar substituents and asymetrically distributed fractional charges which react together. This energy is Emm.

2. Monopole-induced dipole interaction

   The nucleic acid bases have polarisable rings and the point fractional charges (above) in one base can set up polar charges in another base. This dipole can then react with the point charge to which it is due. This is Emd.

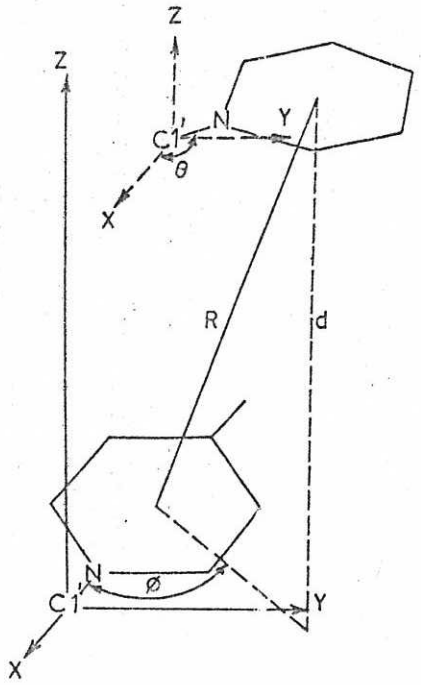
3. London-dispersion interaction

   This comes about because the oscillating electrons in the base ring, even if it is without charge can at any time be displaced giving one side of the cloud a slightly positive charge (electrons are negative), with the opposite occuring below the plane of the ring. This can induce interaction with a similar effect on juxtaposed base rings. This Eld.

4. Repulsive interaction

   This limits how close the rings can approach, the other forces notwithstanding. This is Erp and is negative.

Therefore the total attractive base-base energy = E=Emm+Emd+Eld+Erp Emm and Emd are the asymmetric energies since they depend on the relative positions and orientations of the bases. Eld and Erp are Symmetric, depending on the geometric overlap. Position and orientation are determined by the two parameters $\theta$ & $\phi$. Geometric overlap is determined by d & R. What these 4 symbols relate to is shown in the diagram - Figure 10.




Four parameters varied in the calculation

d = Vertical distance between two bases stacked at parallel planes.

R = Centre to centre distance between the two bases.

θ = Angle in projection between two C1'-N bonds of the two bases, as shown.

ø = Angle in project, between R and C1'-N bond of the lower base, as shown.

**Figure 10**

**There are two main types of base stacking, normal and inverted. These types are shown in Figure 11.**

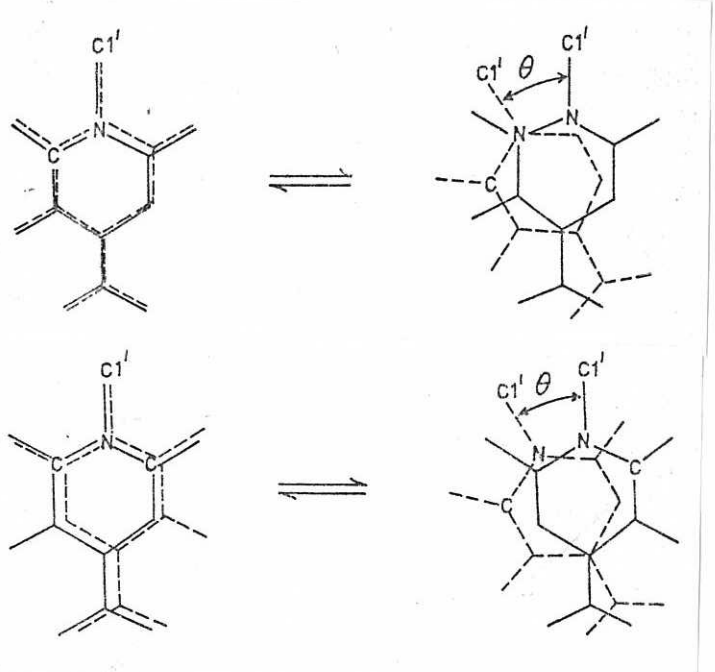



**Figure 11**

In (a) the normal stack is shown; when $\theta=0$ the two bases lie directly above one another, altering $\theta$ displaces them but the reference atoms C lie one the same side of the C1-N bond.

In (b) the inverted stack is shown; the reference atom C of each base lies on a different side of the C1-N bond as becomes apparent on the change in $\theta$ from $0^\circ$. One of the bases has been flipped over and lies on its back compared with the other.

The results were therefore determined for both normal and inverted stacking using values for d & R of 3.4Å and 3.7Å respectively. Values of $\theta$ & $\phi$ were altered by $30^\circ$ at a time. The interaction energy components were calculated using previous techniques (8), based on the observed parameters from crystal structures of methylated bases. Tables 2 and 3 show the results.

TABLE 2

Normal Type

| | θ \ φ | 0° | 90° | 180° | 270° |
|---|---|---|---|---|---|
| G-G Stacking | 0° | +2.0(+2.6) | -11.7(-12.0) | -14.0(-13.3) | -9.0(-11.0) |
| | 90° | +1.0(+1.8) | -13.0(-14.0) | -7.1(-9.6) | -11.1(-13.0) |
| | 180° | +0.0(+0.9) | -11.5(-13.2) | -9.0(-11.6) | -13.8(-14.7) |
| | 270° | +1.4(+2.0) | -9.9(-11.9) | -12.4(-12.7) | -10.4(-12.3) |
| A-A Stacking | 0° | -7.9(-7.8) | -11.0(-12.1) | -8.6(-10.4) | -9.3(-10.9) |
| | 90° | -6.6(-7.2) | -8.8(-10.3) | -8.4(-9.7) | -9.6(-10.8) |
| | 180° | -6.6(-7.2) | -8.3(-9.9) | -8.9(-10.4) | -8.8(-10.4) |
| | 270° | -9.0(-9.9) | -10.3(-11.4) | -12.5(-13.2) | -10.7(-11.5) |
| T-T Stacking | 0° | -1.5(-1.4) | -6.9(-7.4) | -7.7(-7.7) | -4.9(-5.8) |
| | 90° | -1.0(-1.2) | -5.0(-6.0) | -3.0(-4.6) | -7.5(-7.8) |
| | 180° | -1.4(-1.4) | -1.9(-3.7) | -3.7(-4.8) | -4.4(-5.6) |
| | 270° | -0.2(-0.6) | -5.6(-4.6) | -5.1(-6.3) | -3.6(-1.7) |
| C-C Stacking | 0° | +4.8(+5.2) | -7.7(-8.8) | -10.5(-12.0) | -8.0(-8.8) |
| | 90° | +3.5(+4.8) | -5.8(-7.2) | -15.9(-16.4) | -7.2(-8.3) |
| | 180° | +5.3(+4.6) | -5.4(-6.9) | -14.0(-15.8) | -5.9(-7.3) |
| | 270° | +2.7(+4.1) | -8.2(-9.0) | -7.2(-9.8) | -6.2(-7.5) |

TABLE 3

Inverted Type

| | θ \ φ | 0° | 90° | 180° | 270° |
|---|---|---|---|---|---|
| G-G Stacking | 0° | -10.0(-11.1) | -2.7(-2.2) | -3.3(-5.1) | -13.4(-14.0) |
| | 90° | -8.1(-10.1) | -2.7(-6.7) | -5.9(-6.2) | -14.6(-15.0) |
| | 180° | -12.0(-12.8) | -3.5(-3.8) | -5.3(-6.5) | -11.8(-13.1) |
| | 270° | -12.0(-12.7) | -3.6(-4.0) | -5.6(-6.0) | -10.6(-12.1) |
| A-A Stacking | 0° | -9.9(-11.1) | -10.0(-11.7) | -7.9(-9.3) | -9.3(-8.8) |
| | 90° | -8.5(-9.8) | -8.9(-10.3) | -8.3(-9.5) | -7.9(-8.7) |
| | 180° | -8.7(-10.0) | -7.9(-10.3) | -8.3(-9.6) | -6.6(-7.7) |
| | 270° | -11.3(-12.1) | -11.0(-11.2) | -10.8(-11.2) | -9.0(-9.3) |
| T-T Stacking | 0° | -4.1(-4.6) | -3.2(-3.8) | -8.0(-8.1) | -5.9(-6.8) |
| | 90° | -4.6(-5.1) | -1.1(-2.1) | -4.4(-5.9) | -6.2(-6.6) |
| | 180° | -4.0(-4.5) | -0.5(-1.6) | -2.5(-4.4) | -2.8(-6.4) |
| | 270° | -1.5(-3.0) | -3.0(-3.6) | -3.7(-5.3) | -2.2(-6.3) |
| C-C Stacking | 0° | -12.8(-13.4) | -6.2(-7.0) | -2.7(+.27) | -6.7(-6.1) |
| | 90° | -16.5(-16.0) | -6.5(-7.0) | +0.9(+2.6) | -4.4(-6.2) |
| | 180° | -12.1(-13.0) | -6.6(-7.0) | +5.1(+4.1) | -6.6(-7.7) |
| | 270° | -6.4(-8.9) | -6.4(-7.1) | +1.1(+1.9) | -7.4(-8.4) |

The stacking energies are in K calories per two moles of bases. The figures in the brackets refer to d=3.4Å and R=3.54Å.

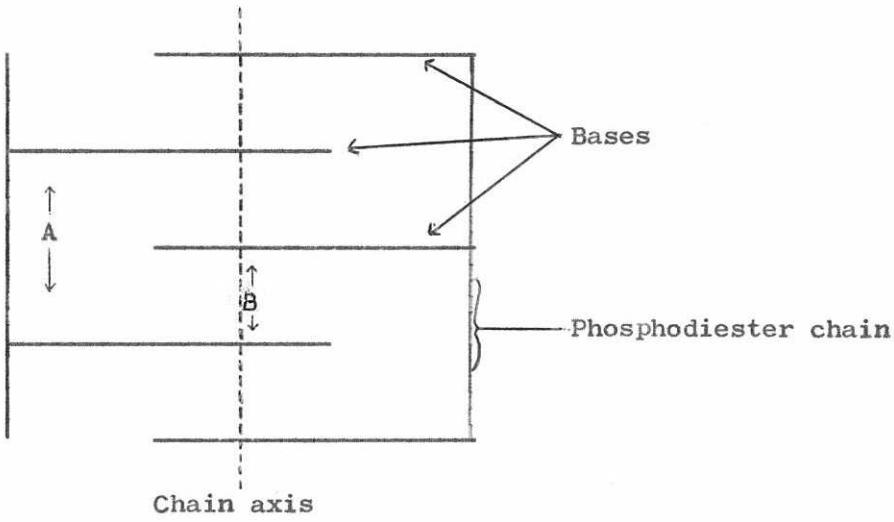
Features here are: 1) that generally, inverted stacking is as feasible as normal stacking is.

2) that some values of $\theta$ clearly prohibit same-chain stacking, i.e. $\theta = 180^{o}$ but that this data can be used to calculate inter-strand stacking, another element of importance for double polynucleotide structures. - Figure 12 below.

A=intra strand stacking

B=inter strand stacking (the overlap here is exaggerated)

(A is dominated mainly by Emm or Eld while B is dictated by Emm and Emd)




3) that for some values of $\theta$, C-C and to some extent G-G stacking is highly unlikely in the normal conformation and the inverted stack shows itself to be over 10 K cal/mole more favourable for C-C. This is of significance for the 'fold' points of the RL structure.

4) that the hydrogen bonding must modify the stack effects, generally increasing stack stability, possibly even stabilising the otherwise unstable stacks.

Stacking in free bases must be investigated in crystals of the same type e.g. pure C-C, G-G, A-A, & T-T. In the second paper the investigation was extended by the use of dinucleoside units to study interactions of different bases.

First the study concentrated on nearest-neighbour stacking on a single chain polynucleotide. Then the study extended to double strand polynucleotides . In this section both left and right hand conformations were considered, by alteration of $\theta$ values. The other parameters in chains are much more constrained due to the backbone. Both inverted and normal stacking for these was also reported.

It must be pointed out that previous fine determination of these interactions has been largely hamstrung by consideration only of such interactions in right-handed B D.N.A. (8)

The single chain interactions for homo-dinucleosides are shown in graph form. Figure 13. From this one can see the clear unfavourability of C-C stacking in the normal conformation (hydrogen bonding not being considered). For a conformation to be favourable the energy in K Kcal/2moles must be as negative as possible. Both A and P geometry are shown, denoting in A a $6^{\circ}$-$7^{\circ}$ tilt of the bases w.r.t. the helical axis, while P is horizontal.

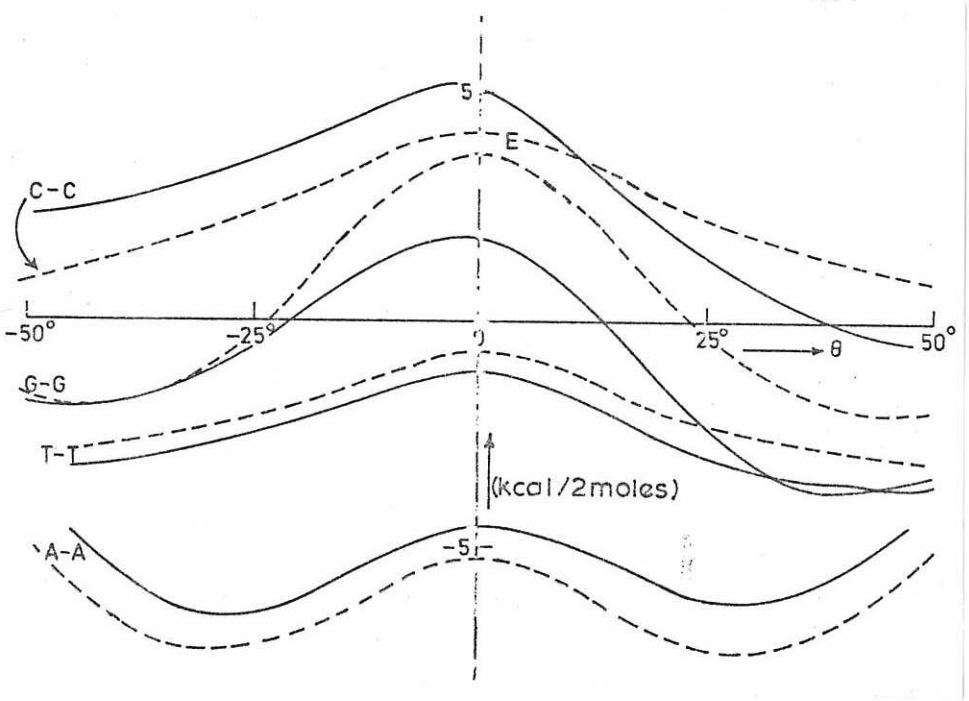



Figure 13

From this graph definite energy minima at +/- 35° are observed These correspond to the Watson-Crick helical twist (approx.36°) but these are in both directions, denoting both left and right hand helical senses.

Complementary base dinucleosides show themselves to be generally more stable in stacking. Figure 14. Minima are observed in T-A & A-T at +/- 35°. The different values of the minima on each side for G-C & C-G indicate that they prefer right and left handed stacking respectively.

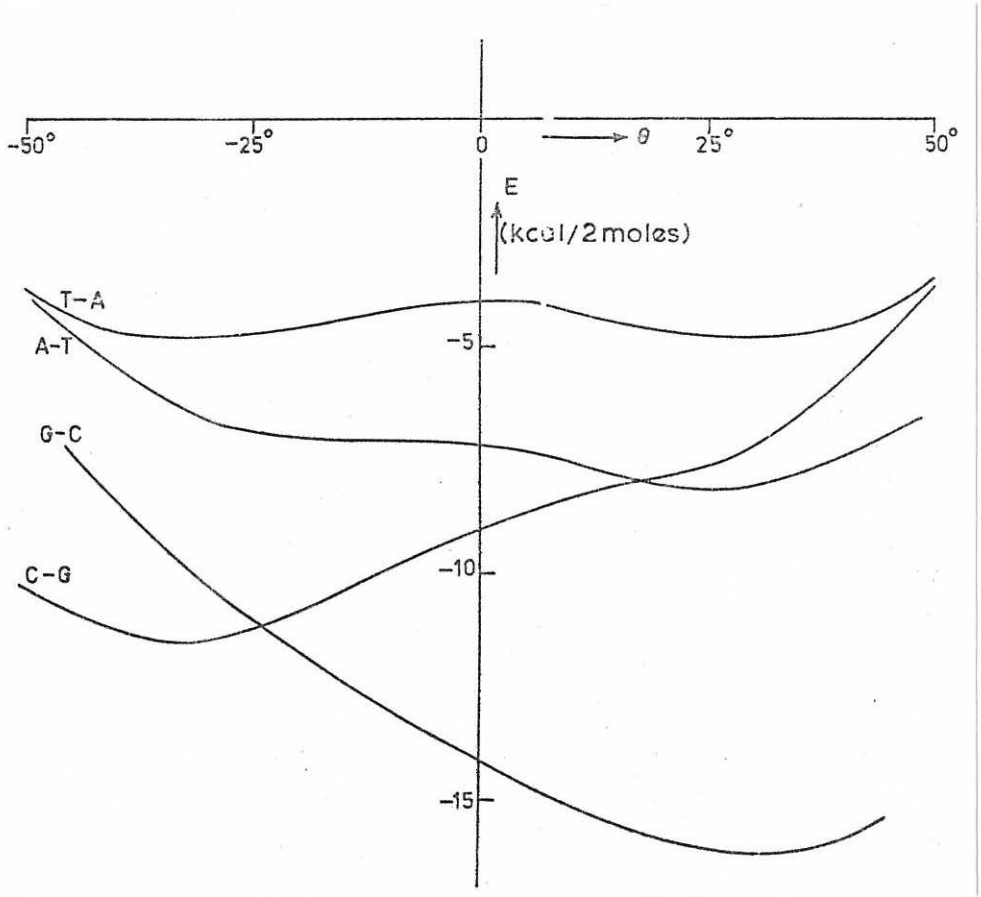



Figure 14

Non-complimentary base dinucleosides show T-G & C-A to have +/- 35° minima but T-C & G-A prefer left and right handed stacking respectively.

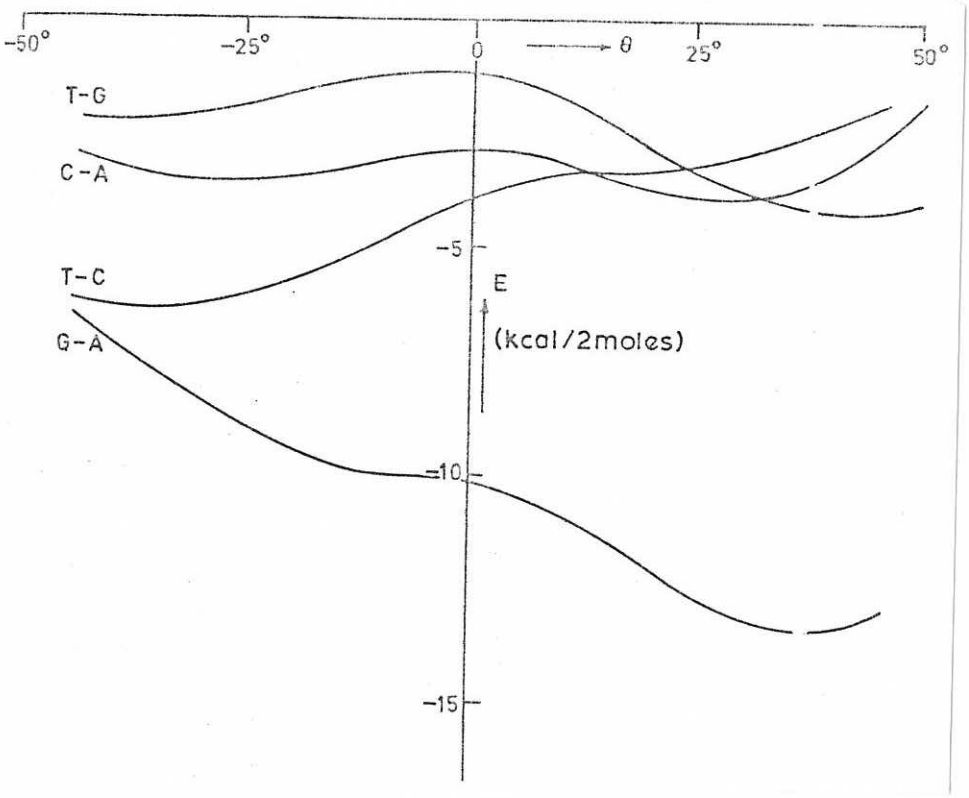
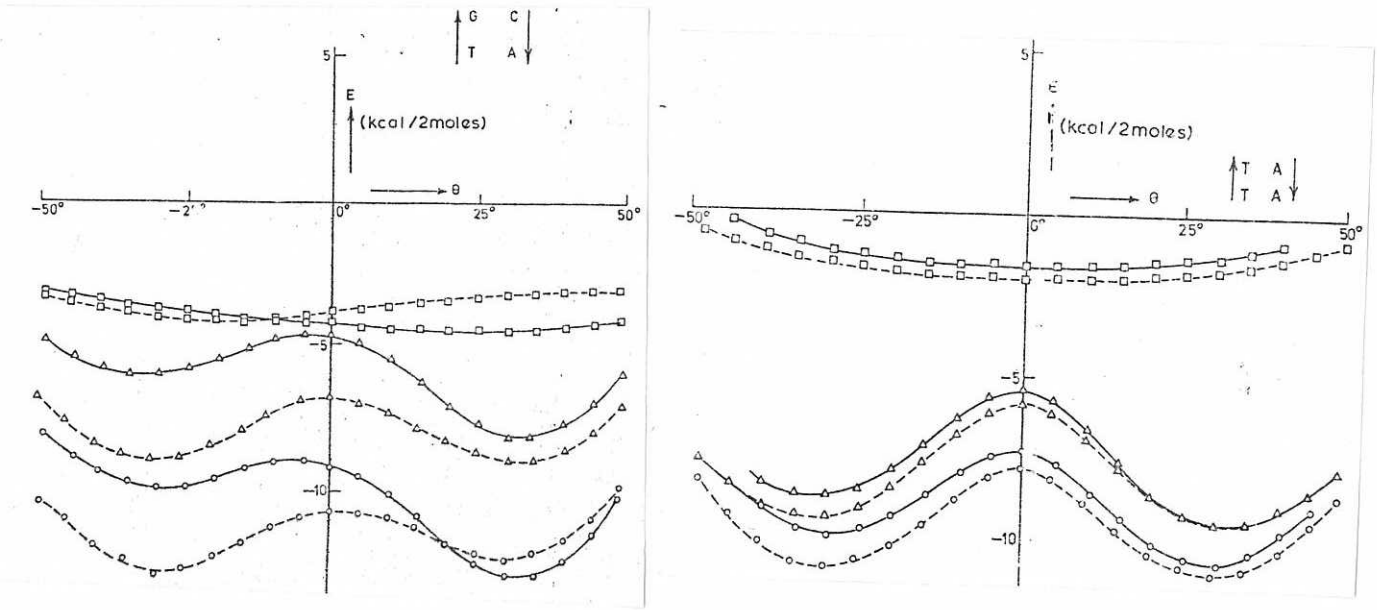



Figure 15

Inreractions between adjacent base pairs in a duplex were calculated using θ values from + to - 50° at 5° intervals. The full results are best seen graphically, below.



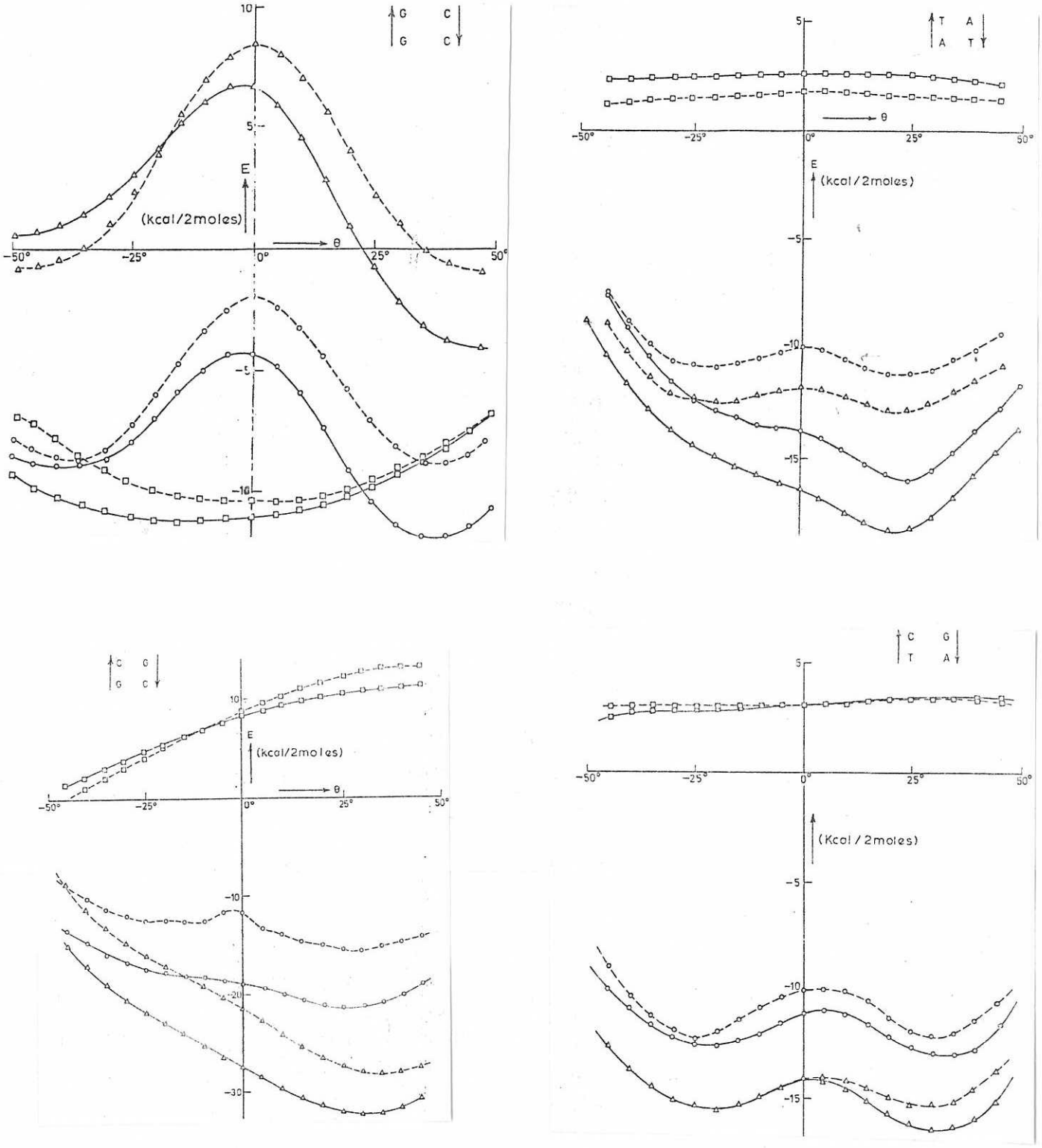



Figure 16

The solid line curves represent energy profiles in A geometry.

The broken line curves represent energy profiles in P geometry.

Square graph plot points denote interstrand energy and depend on Emm and Emd mainly.

Triangular graph plot points denote intrastrand energy which depend on Emm and Eld.

Circular graph plot points denote the sum or total interaction energy.

In most of the doublets the total energy shows roughly +/- 35° minima A and P curves are usually similar. In more detail:-

1) For A geometry both right and left handed stacking arrangements are nearly equally favourable in

G C, C G, A T, T A, T A
C G, T A, T A, T A, G C

However C G, G C, G C, T A, T A all favour right helical stacking.
G C, G C, T A, A T, C T

2) In P geometry (reckoned to be the naturally occurring form) all the doublets show equally favourable left and right stacking except C G which favours right stacking and G C its 'opposite'
G C C G
which favours left stacking by about the same amount.

Inverted stacking was investigated for doublets with reference to the bend region of the RL structure. See table 4. The base pairs were taken in P geometry, the upper base pair is inverted w.r.t. the lower one.

TABLE 4

| Sequence of Base pairs | Intrastrand interaction energy (Kcal/4 moles of bases) | Interstrand Interaction Energy (Kcal/4 moles of bases) | | | Total Interaction energy (Kcal/4 moles of bases) |
|---|---|---|---|---|---|
| | | Interstrand stacking energy | In-plane base-pairing interaction energy | Total interstrand Interaction energy | |
| I II<br>C G<br>G C | - 1.8 | -13.4 | -46.0 | -59.4 | -61.2 |
| G C<br>G C | - 22.1 | + 9.5 | -46.0 | -36.5 | -58.6 |
| C G<br>A T | -13.9 | + 1.4 | -31.1 | -29.7 | -43.6 |
| G C<br>A T | - 9.0 | - 2.1 | -31.1 | -33.2 | -42.2 |
| T A<br>T A | -14.0 | + 1.0 | -16.2 | -15.2 | -29.2 |
| T A<br>A T | -11.6 | + 0.1 | -16.2 | -16.1 | -27.7 |

The results show that for all of these pairs the inverted stacking is as feasible as normal left and right handed arrangements.

In some, i.e. G C the inverted stacking is more stable than either left normal
G C
or right, with 58.6 Kcal/2 moles cf 55.1 Kcal/2 moles.

Therefore inverted stacking is also possible for B D.N.A. helical parameters and with Watson-Crick base pairing. As we have already seen, the backbone conformations allow for a 'fold' region, which is essential to the RL structure. The above findings for inverted stacking now make the 'fold' region increasingly feasible. (The 'fold' region as reported above - p 8, The Alternative 'Backbone'.)

To further investigate the fold region and the requirements it imposes, some triplet interactions were investigated. In the fold region triplet stacking in the top base pair needs must be different to that of the lower pair. 4 representative triplets were taken in P geometry. The results in Table 5 are startling.

TABLE 5

| Sequence of Base-pairs in Triplet | Nature of stacking | | Nature of Bend, if any | Energy of Interaction (Kcal/mole of triplet) | | |
|---|---|---|---|---|---|---|
| | Bottom Doublet | Upper Doublet | | Intra-strand inter-action energy | Interstrand interaction energy | Total energy of Interaction |
| I II<br>G C<br>C G<br>G C | Left | Left | No | -41.4 | -54.1 | -95.5 |
| | Right | Right | No | -41.4 | -54.1 | -95.5 |
| | Left | Right | Type I | -26.2 | -65.0 | -91.2 |
| | Right | Left | Type I | -56.2 | -43.2 | -99.4 |
| | Left | Right | Type II | -14.9 | -80.4 | -95.3 |
| | Right | Left | Type II | -29.9 | -69.5 | -99.4 |
| I II<br>A T<br>T A<br>A T | Left | Left | No | -23.1 | -21.3 | -44.4 |
| | Right | Right | No | -23.1 | -21.3 | -44.4 |
| | Left | Right | Type I | -22.4 | -21.3 | -43.7 |
| | Right | Left | Type I | -23.8 | -21.5 | -45.3 |
| | Left | Right | Type II | -22.8 | -22.7 | -45.5 |
| | Right | Left | Type II | -23.5 | -22.8 | -46.3 |
| I II<br>G C<br>G C<br>G C | Left | Left | No | + 0.2 | -87.4 | -87.2 |
| | Right | Right | No | + 0.2 | -87.4 | -87.2 |
| | Left | Right | Type I | + 0.2 | -87.4 | -87.2 |
| | Right | Left | Type I | + 0.2 | -87.4 | -87.2 |
| | Left | Right | Type II | -22.0 | -68.8 | -90.8 |
| | Right | Left | Type II | -22.0 | -68.8 | -90.8 |
| I II<br>T A<br>T A<br>T A | Left | Left | No | -18.8 | -27.3 | -46.1 |
| | Right | Right | No | -18.8 | -27.3 | -46.1 |
| | Left | Right | Type I | -18.8 | -27.3 | -46.1 |
| | Right | Left | Type I | -18.8 | -27.3 | -46.1 |
| | Left | Right | Type II | -23.4 | -24.8 | -48.2 |
| | Right | Left | Type II | -23.4 | -24.8 | -48.2 |

In all 4 triplets the necessity of folding actually gives them higher interaction than they show in the Type 1 structure (The papers are not presented in sequence here) and even in the conventional B D.N.A. conformation.

To round the description of the alternative structure I will use the work reported in the periodical Current Science in 1980 'On the alternative structure of D.N.A.: role of syn conformation of the bases'.(6) Which shows how this group's research, while refining their model has met with independent corroboration from an adjacent area.

The syn conformation of bases is not possible for pyrimidines in helical domains yet it is possible for purines in left-handed duplexes for certain phosphodiester conformations and has been observed in certain crystal structures.

Investigating with trinucleoside diphosphate which gives an alternating pyrimidine/purine sequence (18) two structures arose with helical domains similar to B D.N.A. One was uniform, the other was zig-zag.

In the uniform structure alternate right and left sections could be constructed with all the bases in the right hand structure in anti conformation with the purines in the left hand section in the syn conformation.

In the zig-zag, helixes were created by marrying a helical domain dinucleoside to a non-helical domain dinucleoside. This results in a zig-zag relationship between the phosphates of the backbone. Again due to stacking and inversion both left and right-handed versions of this structure are possible and can join together. This is essentially what Wang's group came to but in their model the left handed zig-zag joined to B D.N.A. (19) Figure 23.

## BRIEF DISCUSSION OF SOME IMPLICATIONS OF THIS ALTERNATIVE STRUCTURE

Having established the principles of this structure I would now like to look at some of its implications and functional features.

Functionally the structure has many advantages. Base-stacking studies show not only that it is not inherently weak due to 1) its openness nor 2) its bends but that it is in fact probably stronger than the established model. I refer you to Proceedings of the National Academy of Science vol 75 (9) pp 4092-6 'Some implications of an alternative structure for D.N.A.' (15)

Here are some of its properties:-

1) It is able to open its bases simply by the breaking of hydrogen bonds without as well having to unspiral its chain as B D.N.A. is supposed to for replication and template forming.

2) It is able, similarly, to undergo 'semi-conservative' replication. See Figure 17.

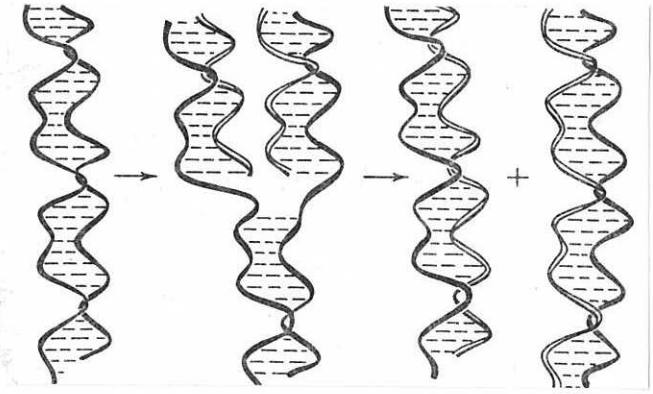


Figure 17

3) It is able to undergo linear supercoiling without extreme disruption of its structure. The article explains this very well, much more convincingly than Crick and Klug (13) See Figure 18.

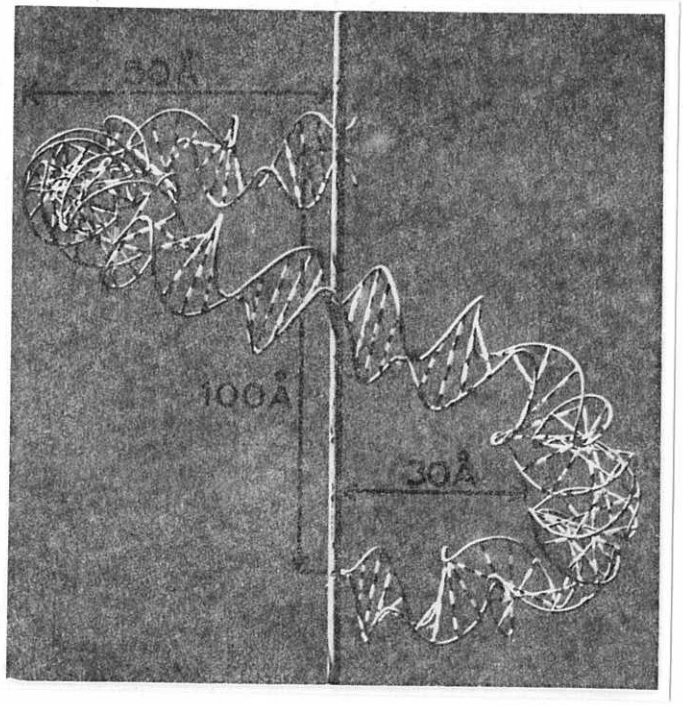
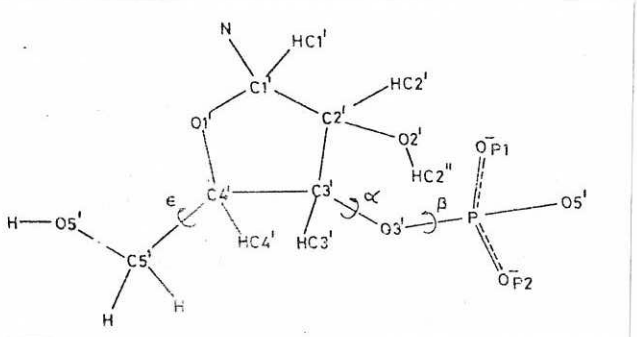



Figure 18

4) It provides an alternative for histone binding which allows for a much less rigorous intwining attachment, the histone merely bonding along one side of the open chain.

5) It is a structure which explains D.N.A.'s low optical activity because, with opposite sense chains in the same molecule light will be absorbed in two planes, at least, as it passes through the lattice. (With the conventional model the light may even be plane-polarised but still have its other vectors un-absorbed).

As I elaborate in the next section, part (2), this group's approach from the fine structure is potentially very valuable. This is well demonstrated by their latest paper investigating the flexibility of the C3 bond - see Figure 19, in both ribose and deoxyribose. "Is 3'-nucleotide rigid?" in Nature vol 284 pp187-8 (21)




In this diagram ribose-phosphate is shown. For comparison see figure 1 for deoxyribose phosphate. NB The C2' groups.

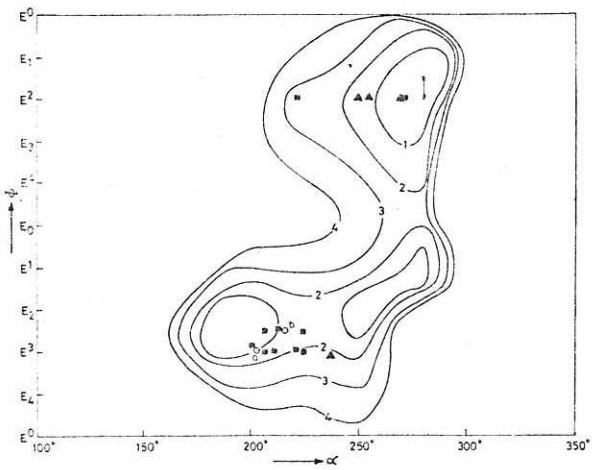
Figure 19

The geometric parameters of the sugar component were selected from minimum energy regions of the pseudo-rotation space ($E^0$,$E^1$,$E^2$,$E^3$,$E^4$ in graphs) That is to say $\Phi$ represents the sugar's minimal energy conformationa w.r.t. $\alpha$, the 3' nucleotide angle, the independent variable.

$\alpha$ was changed between $120^o$ and $300^o$ at $10^o$ intervals with $1^o$ intervals at the minima. Above $300^o$ bonds are formed between C5' groups and the phosphates. Below $120^o$ bonds are formed between C2' groups and the phosphates.

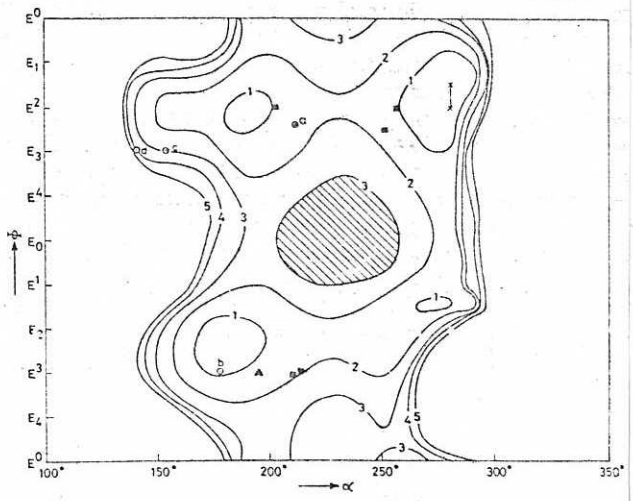
$\beta$ & $\epsilon$ were fixed at average values $290^o$and $50^o$while the pendant oxygens were fixed tetrahedrally. (Note that again only the gg values for C4-C5, E, are presented here.

$\Phi$ was plotted against $\alpha$ and energy contours of 1Kcal/mole were drawn to show minima. At the same time the positions of known/proposed structures were included. The minima in the graphs are shown as a straight line along which energy is very stable. For both R.N.A. and D.N.A. this minimum occurs at $\alpha$ =280 in C2 endo. See Figures 20 & 21.



Φ–α plot for the ribose–phosphate fragment. Energy contours of 1.5 kcal $mol^{-1}$ in the C-3′ *endo* region are shown. ▲ Denotes 3′-nucleotides, ▣ denotes dinucleoside monophosphates and higher oligomers, ⊙ denotes proposed structures of polyribonucleotides. $\odot^b$, RNA-11 (ref. 11); $\odot^a$, A-RNA (ref. 10).

Figure 20



Φ–α plot for the deoxyribose–phosphate fragment. ▲ Denotes 3′-nucleotides, ▣ denotes dinucleoside monophosphates and higher oligomers, ⊙ denotes proposed structures for polydeoxyribonucleotides. $\odot^a$, C-DNA[18]; $\odot^b$, A-DNA[17]; $\odot^c$, B-DNA[12,13]; $\odot^d$, D-DNA[14].

Figure 21

The graphs show the different energetically feasible areas. Notably the ribose has only two such areas, one C2 endo and one C3 endo while the deoxyribose has two for each of these pucker formations, 4 in all. This is due to the bonds formed by the extra OH group on C2′ in ribose. The graphs also show that there are transitions between forms. For ribose C2 endo changes to C3 endo along a feasible energy gradient when $\alpha$ is in the $270^{\circ}$ region. For deoxyribone there are two feasible energy gradients, when $\alpha$ is in the $280^{\circ}$ region as well as when $\alpha$ is about $180^{\circ}$.

The deoxyribose fragment's larger conformational flexibility is perhaps surprising. In the context of the Watson-Crick double-helical model this variability should not be the case. In the context of the base-stacking-determined secondary structure of the RL model however, this variability is expected.

Interestingly as well, the proferred models for B D.N.A. are shown to be very unfavourable with $\alpha$ at low values around $150^{\circ}$. With adjustment the B D.N.A. can be made favourable i.e. as the RL structure.

## A CONSIDERATION OF STILL WIDER IMPLICATIONS STEMMING FROM THIS WORK

New techniques have almost certainly aided the enquiry into D.N.A. structure based on base-stacking pursued by the Bangalore group. The difficulty of interpreting X-ray analysis etc. (Figure 22 & comment) has probably been a factor in preventing detailed questioning of D.N.A. structure for so long.

However, more important has been the reification of the structure evidenced 1) by Pullman's study of base-stacking which totally ignored the left-handed possibility and 2) by the sluggishness of the research community to consider this group's work. (Crick, in his paper on plasmid supercoiling (20) dismisses all left-handed alternatives with a 'topographical' model and argument which seems more than derisory).

If the Bangalore group were misinterpreting generally accepted facts, for instance on base stacking, this would be in order. However, their work, as far as I have found, is the only work done on base-stacking since 1968. (8)

Heuristic principles seem to be lacking at this level. Let us now go further and, using the basic principle of base-stacking-determined secondary structures, consider what else requires fresh consideration.

The work of the Bangalore group, while producing a radical alternative to the established structure, conforms to the basic parameters of N=10 and h=3.4Å. I am led to question these parameters.

1) D.N.A. in vivo has not and probably never will be measured. Therefore the figures for N and h refer to a non-typical, severely treated, naked, low energy form of D.N.A. (I confess to being ignorant of the processes and conditions under which X-ray analysis et al are carried out). I feel that researchers must be much more flexible before rejecting models for D.N.A./R.N.A. with many functionally valuable features because these models do not edactly meet these two requirements. For instance, above p 21 para 1 the zig-zag structure requires the alternation of non-helical with helical components (6)

This is a suggestion made to the Bangalore group as much as to any other. Their stress on the role of base-interactions in determining structure indicates a far from regular structure; structure being pragmatic to base sequence and relative base orientation. N=10, as evidenced by figure 22 (of <u>processed</u> D.N.A.) is not a very convincing rigid figure and should be elastic. At the same time h=3.4Å is bound to vary as a function of the attractive minus the repulsive stacking forces; determined by syn/anti base conformations and precise orbital overlap/interaction.

From this reasoning, an RL structure of 5 bases per sensical section is no more to be primarily expected than the Watson-Crick model while each would remain possible forms, depending on specific parameter conjunctures.

2) One of the major areas remaining unclear in the elucidation of D.N.A. behaviour concerns the mechanisms controlling precisely which sections of D.N.A. 'open' to present bases for template formation etc.

It is possible to postulate levels of increasingly specific control. A first level might be the provision of a chemical medium to 'open' the molecule. This might be the action of a general reagent like a reducing agent to break hydrogen bonds. Secondly there could be blocking action of histones and other D.N.A.-associated proteins. Thirdly this action could be mediated by blocking effects of the gross structure of the D.N.A. e.g. tight spiralled 'Z' D.N.A. would physically withstand the action of the reducer more so than a more open form. Fourthly, above p 14 para 2(4)), the reagent would be effective according to the importance of hydrogen bonds in stabilising the base stacks of sequences of specific base pairs.

Using the new model we have 'open' structures of D.N.A. whose bases hold themselves more availably to free nucleotides and whose flexible backbones would allow base rotation out of the normal molecular axis.

Consideration of both hypothetical control levels and the behavioural potential of different models may lead to quite comprehensive and testable ideas of precise mechanisms for template formation.

3) A complicating feature is that of mobility. Work done on crystals reveals a series of essentially stable forms. Each stable crystal derivative is only one of the possible conformations of its parent units. Following basic thermodynamic principles these different conformations can be taken up, at least some of the time, by the bases in vivo. How long, proportionally, the bases hold any form is a function of that form's energetic favourability and also of the energy barriers or entropies between it and the alternatives. - see above p 23 para 3. Since the system most often operates at circa 35°C we must expect spontaneous base-stacking changes, for instance, as follows:

As the work done shows, most of the parameters are closely inter-related. C4-C3 angle determines C2 or C3 endo in the sugar phosphate backbone.(16) and above, The Alternative Backbone).

Also in (6) p43 column 2 para 1 it is shown that changes in pucker, in combination with changes in C4-C5 from gg to gt give a range of different glycosyl torsion angles. These can in turn effect base conformations, stabilising syn or anti-conformation depending on helical sense and base type i.e. purine or pyrimidine. This effects base-stacking and therefore overall secondary structure.

From (22) and above p 23 para 3 the backbone's energy barriers are not high and variability is to be expected. The conclusion must be that the molecule is inherently complex and dynamic. Previous work has tended to assume that the histones were the major organisers and regulators of the otherwise static D.N.A. molecule. This must now change and

thrust of enquiry must turn to the elucidation of D.N.A.'s auto regulatory potential.

I feel strongly that the work of this group is not receiving enough attention and, as I hope I have shown, this may be delaying the opening up of an enormous amount of very interesting and important fields of scientific enquiry.

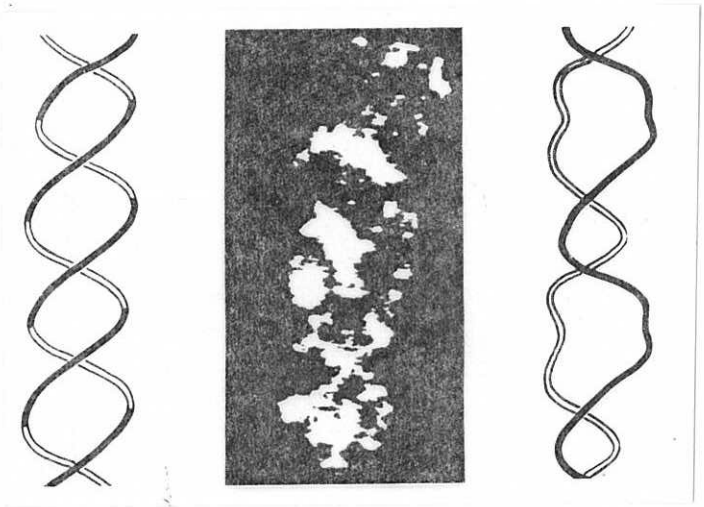
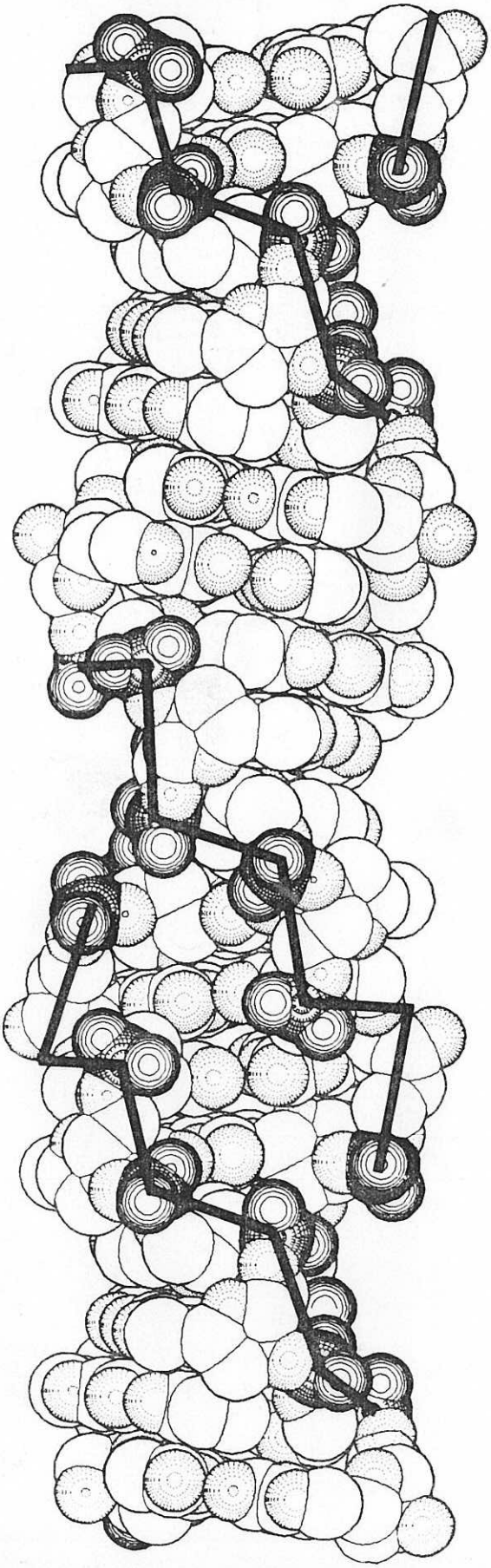


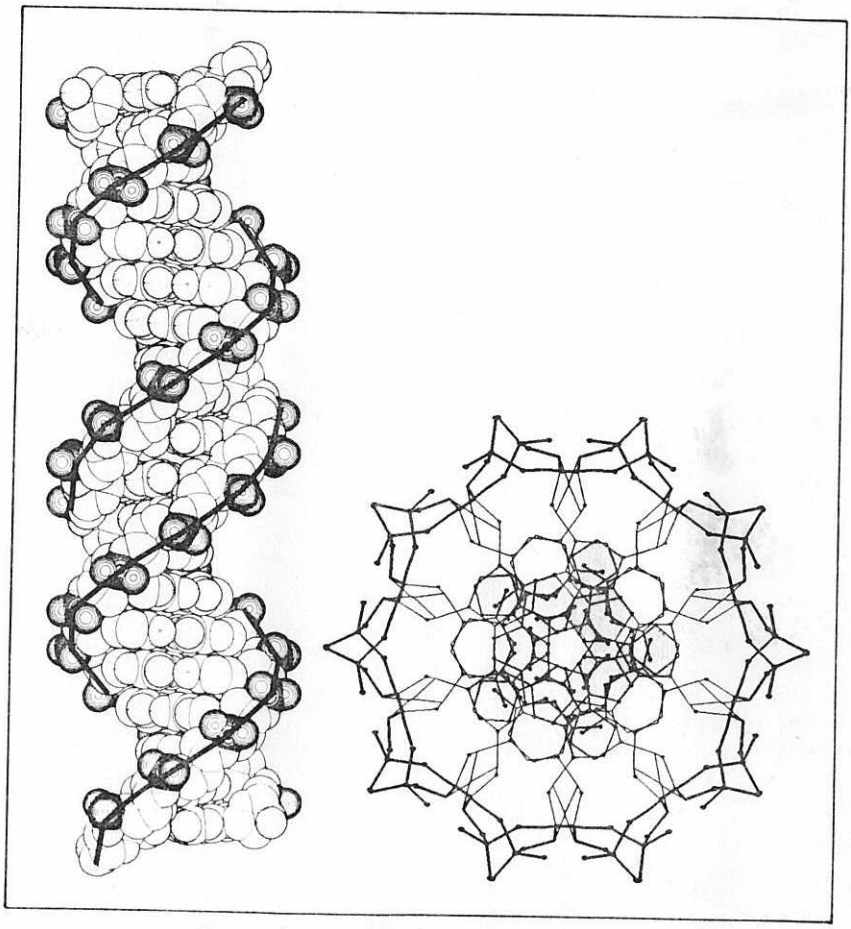
Figure 22

X-ray analysis etc.
This is an etc, not very clear.

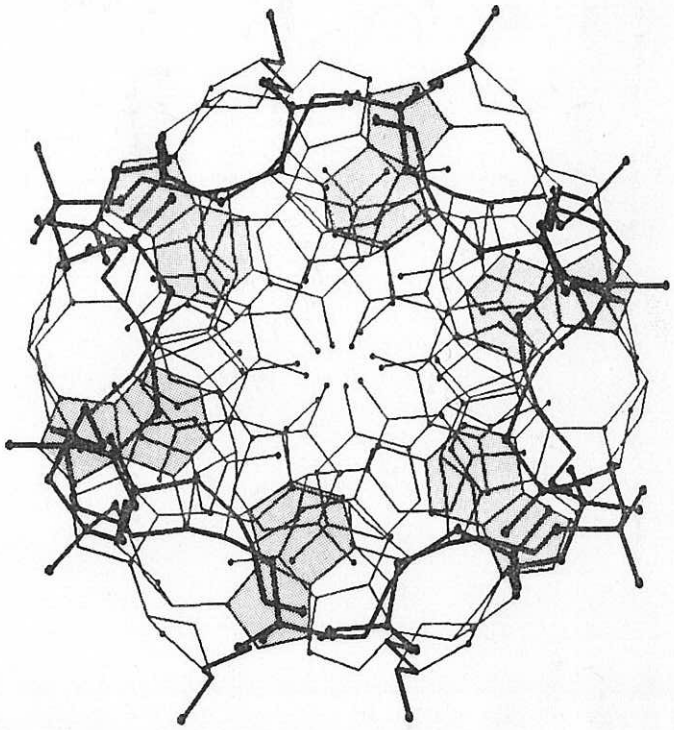
nature
Vol 282 No 5740 13 December 1979 £1.00 US $2.25

Figure 23





# Left-handed DNA



NOTES

1. These different forms usually are reckoned to exist in different media. I.E. P geometry D.N.A. is the hydrated form of A geometry D.N.A. - see text later.

2. Proceedings of the National Academy of Sciences (P.N.A.S.) Volume 73 p 2959 "An Alternative Model for the Structure of D.N.A." Rodley G.A., Scobie & Lewitt (1976)

3. V. Sasisekharan and N. Pattabiraman in Current Science Vol XLV No 22 pp 779-783 (1976)

4. V. Sasisekharan and N. Pattabiraman and Goutam Gupta in Current Science Vol XLVI No 22 pp 763-764 (1977)

5. Wang etal in Nature. Vol 282 p 680 (1979)

6. V. Sasisekharan, & Goutam Gupta in Current Science Vol XLIX (2) pp 43-48 (1980)

7. There are many implications for the structure and behaviour of R.N.A. implicit in the study that this group is doing. In the article in Nature 284 discussed at the end of this paper this is made clear.

8. A group reference.
   a) De Voe and Tinoco in Journal of Molecular Biology Vol 4 pp 500-1 (1968)
   b) Pullman and Pullman in Advances in Quantum Chemistry Vol 4 pp 267-321 (1962)
   c) Langridge, Marvin, Seeds, Wilson, Hooper, Wilkins and Hamilton in J. Mol. Biol. Vol 2 pp 38-64 (1960)

9. Phages and Viruses are not, strictly speaking, cellular, but then are they alive? D.N.A. equals life etc.

10. Again the existence of different D.N.A. conformations is usually reckoned to hinge on its medium. Ethyl bromide causes super-coiling, etc.

11. Goutam Gupta and V. Sasisekharan in Nucleic Acids Research Vol 5 (2) pp1639-1653 (1978) and pp 1655 -

12. Bauer, Crick and White. in Scientific American. Vol 243 (1) pp 100-13. (1980)

13. Crick and Klug in Nature . Vol 255 pp 530-33 (1977)

14. In Current Science - Note 3.

15. In Current Science - Note 4 and later.
V. Sasisekharan, N. Pattabiraman, Goutam Gupta in P.N.A.S. Vol 75 (9) pp 4092-6 (1978)

16. V. Sasisekharan & Goutam Gupta in Nature Vol 275 (5676) pp 159-62 (1978)

17. S.H. Cohen & J.A. Shapiro in Scientific American Vol 242 (2) (1980) see the note referring to a paper by Chamberlin in Journal of Cell Science on R.N.A. polymerase.

18. Adenosine-Thymine would be such a sequence. A-T long repeats may then have quite varied structure despite apparent homogeneity.

19. See 5. Wang approaches his structure from a different point of view. To my now very biased view his approach is relatively limited

20. See 12.

21. Is 3-nucleotide rigid? N. Pattabiraman, Shashidar N. Rao, V. Sasisekharan in Nature Vol 284 pp 187-8.