Energy Band Structure of Models of Periodic Deoxyribonucleic Acid (DNA)

by

Lily Y. JOESOEF

BEDFORD COLLEGE

UNIVERSITY OF LONDON

Dissertation undertaken as part of the
M.Sc. in Solid State Physics at Bedford
College, University of London (1975)

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ACKNOWLEDGMENTS

I would like to thank Dr. N.W.Ramsey from Charing Cross Hospital who gave me a tremendous moral support and understanding while I was doing this course as a part time student.

I would like to record my sincere thanks to my supervisor Professor L. Pincherle from Bedford College for all his help, patience, kindness and encouragement throughout the course.

During the course of my studies I have been continually helped and encouraged by Dr. N.E.Hill from Bedford College and I would like to express my deep gratitude to her for her kindness and valuable advice.

I should like to thank Dr.T.Lake for the computer proggraming and Miss Janey Varley who has so carefully typed this dissertation.

ABSTRACT

This work originated from studies of the damage caused to DNA by radiation, damage which has been considered as one of the causes of induced cancer. It was suggested, mainly by Ladik and co-workers, that in order to understand the nature of this damage, it was first necessary to have a picture of the electronic structure of DNA, one of the problems being the degree of localization of the electrons on chains. Several attempts were made by this group to calculate the electronic structure, by considering more and more complicated periodic chains approximating more and more the true structure of DNA. These calculation are critically reviewed in the present Dissertation. The early work had been carried out on the simple Hückel model; later the Pariser-Parr-Pople approach was employed, yieding considerably different results. The inadequacies of the one-dimensional model and of other approximations are pointed out.

These calculations yielded an energy gap of over 3 eV, rather higher than experimental value. They gave fairly narrow bands, indicating that electrons cannot be de-localized over the whole chain.

One result was that the bands were narrower in infinite chains made up of pairs of bases than in chains of the same nucleotide. Clearly this must be due to a smaller overlap of similar atomic orbitals because of the interposal of a different structure, and this explaination is checked in the present Dissertation by an exact calculation of the energy bands of chains of square potential wells.

ENERGY BAND STRUCTURE OF MODELS OF PERIODIC DEOXYRIBONUCLEIC ACID (DNA)

CHAPTER 1

Introduction

It is commonly accepted that DNA was discovered by F.H.C. Crick and J.D. Watson in 1953[1]. Strictly speaking what they discovered was not DNA itself but the three dimensional molecular arrangement of this macromolecule, which is now known as DNA, a short form for deoxyribonucleicacid.

The substance itself which later on was called DNA was first discovered by Friederik Mischer[2] when he isolated the pus cell obtained from discarded surgical bandages in 1869. He called the substance he discovered, which contained an unusual phosphorus compound, "nuclein". Later on it became known as nucleoprotein.

In 1870, he began his investigation on salmon spermatozoa, which when isolated contained an acidic compound, now recognized as nucleicacid.

Various cells and tissues were then examined and were found to contain nucleic acid.

Kossel in 1884 pioneered the work to find out the detailed chemistry of nucleic acid and the proteins associated with them in the cell nucleus.

In 1889, Richard Altman gave Mischer's "nuclein" a proper name "nucleic acid".

James Watson and Francis Crick became interested in work on DNA and eventually in 1953 suggested a stereometric model aided by the X-ray photographs of Maurice Wilkins, who had been working for some time to find the structure of DNA .

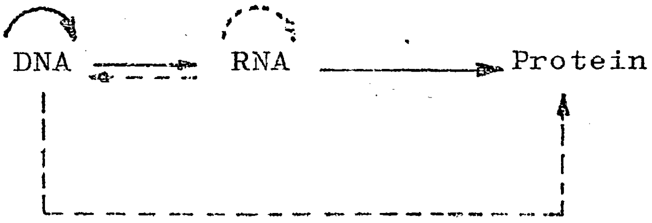
In the next decade researchers focussed their interest on trying to find out more about DNA; its properties, confirmation of Watson-Crick stereometric model, the relationship between DNA and genetic material, in other words "what is the role of DNA in our life?". In recent years attention has been focussed on the damage caused to DNA by radiation, and thus an attempt has been made to find a direct relation between DNA and induced cancer, and even natural cancer . Some work has also been done on the effect on DNA of drugs, again aiming to find a more effective way of curing cancer, some even hoping that they can eventually discover the cause of cancer.

Accepting the concept that DNA is a carrier of genetic information, it has to serve two main functions.

1. Duplication or replication, that is being able to make an exact copy of itself.
2. Transcription, that is being able to pass on the information coded within it to the messenger RNA (mRNA).

The messenger RNA may in its turn translate the information in the four letter language of the DNA into the twenty letter language of the amino acids and proteins. This concept was developed by Watson and Crick and called "The Central Dogma". This dogma stated that the information should always be carried from either DNA ⟶ DNA, or DNA ⟶ RNA ⟶ Protein[3],and cannot be reversed. However, its general validity has recently been questioned as the result of the discovery of the RNA-dependant DNA polymerase in certain RNA viruses[4] in 1970, which directly opposed this dogma and which brought about an important break through in cancer research, since the RNA viruses were oncogenic, that is capable of bringing about malignant change.

However, Crick[5] defended his dogma by modifying it as shown below.



The solid arrows indicate general transfers, and the dotted arrows refer to special cases, which may occur.

The study of the electronic configuration of giant biomolecules was initiated by Szent-György as early as 1941. Coulson[5] in 1952 suggested that there was some evidence that some organic substances with benzene configuration could induce cancer. Thus Szent-György, Linus Pauling, Coulson, etc. introduced Quantum Theory into Biochemistry.

Before DNA was discovered scientists focussed their attention on finding the electronic configuration of proteins. However, the discovery of the DNA structure attracted many scientists to find the electronic structure of DNA. Many scientists since have done a great deal of calculation on energy levels of purine bases, pyramidine bases and their individual components. A great deal of this work has been done by Albert and Bernard Pullman, who were credited as the pioneers in finding a quantum mechanical interpretation of chemical carcinogenesis and of antitumor activity in cancer and leukaemia.

The possible correlation between some carcinogenic agents and the electronic structure of DNA was first suggested by Hoffmann and Ladik[7]. Their hypothesis was based on the results of their theoretical calculations on energy bands of DNA. However, as DNA is a complicated macromolecule, all theoretical calculation have to make very drastic assumptions. One of them which is vital in order to perform a calculation, is that it is a periodic structure, while in reality the DNA molecule is almost certain not periodic.

Without assuming that it is a periodic structure, no calculation can be made based on the available Solid State Theory.

The two methods normally used in Organic Chemistry when dealing with π-electrons are either the Valence Bond Method or the Molecular Orbital Method.

The Valence Bond method, abbreviated VB, considers each electron to be bound to a certain nucleus, and the atoms are bound by covalent bonding. However, for a biomolecular structure, this method is inadequate, due partly to the complications in the mathematical treatment. For example treatment of the naphtalene problem involves finding the solution of a determinant of order 42^8, and the order of determinant involved increases rapidly with the increase in molecular dimension. It is also inadequate, because in a biomolecular structure there exist other types of bonding besides covalent bonding, for example ionic bonding. Benzene, which is the simplest form of the heterocyclic-hydrocarbon family, has itself 170 ionic structures! So it seems that the VB method is not so efficient when dealing with biomolecular structures.

The Molecular Orbital Method, abbreviated MO, which starts from several different approaches, for example the Semiempirical Linear Combination of Atomic Orbital (LCAO-method), the Self Consistent Field-Molecular appoach (SCF-MO), and so on, assumes that only the nuclei and the inner electron shell are bound together, and that the valence electrons move in polycentric orbits, which extend though the whole molecule. The object of the MO-method for π-electrons is to deduce the form of the orbital for each π-electron, and to calculate each energy. Therefore the order of the determinant to be solved in a resonating system is equal to either the number of π-electrons in the system or the number of atoms carrying the π-electrons (for atoms with "lone pair electrons"). Thus the order of the determinant is small. Naphtalene has a determinant of order 42 in the VB method, while in the MO method it has a determinant of order 10. This low order of the determinant makes it possible to use the same approximation for a series of widely different compounds. Another advantage of this method is that the elements of the determinant itself expressed in terms of molecular orbitals of each one of the π-electrons.

The aim of this dissertation is to present a historical survey of the work that has been done by different methods in attempting to find the electronic structure of DNA and whereever possible to compare theoretical results with the available experimental data.

In the final chapter a very simple one dimensional model is treated exactly. The philosophy behind this physically not very meaningful calculation was to ascertain whether, by a suitable choice of the parameters of this model, it was possible to reproduce some of the theoretical results obtained by calculations based on more plausible physical assumptions, but involving drastic computational approximations. In particular, it was shown that the fact that the bands in a heteronucleotide structure are narrower than in homonucleotide structures is due simply to the smaller overlap of the wavefunctions caused by the interpolation of different molecules.

References:

1. The Double Helix, James D. Watson (Penguin Book)

2. The Discovery of DNA, Alfred E. Mirsky (Scientific American, June 1968).

3. The Biochemistry of the Nucleic Acid, J.M. Davidson (7th ed. Chapman & Hall)

4. Baltimore, D., Nature, 226, 1209 (1970)
Temin, H.M. and Mizutani, S., Nature 226, 1211 (1970).

5. Crick, F.H.R., Nature, 227, 561 (1970)

6. Advances in Cancer Research, I, 1-56, (New York, Academic Press, 1953)

7. Advances in Chemical Physics, Vol. V11, Ch.2, Hoffmann, T.A. and Ladik, J.

8. Horizon in Biochemistry, 1961, A. Pullman & B. Pullman.

CHAPTER 2

2.1 General Structure and Properties of DNA

As described by Watson and Crick in 1953[1], a DNA molecule is a double-stranded helix. Each backbone strand contains sugar, bases and phosphate. The two strands are linked to each other through the bases by two or three hydrogen bonds. However, the complete structure of the molecule is very complicated as can be seen in figure 1. For the sake of simplicity, it is assumed that the DNA molecule has only four different bases, which can be divided into two groups, one consisting of Adenine and Guanine called Purine bases, and the other consisting of Cytocine and Thymine, called Pyramidine bases, both aromatic. Adenine pairs with Thymine and Guanine with Cytosine. The pair Adenine-Thymine is normally abbreviated AT and the pair Guanine-Cytosine abbreviated GC. The fact that a purine base on one strand is always paired with a pyramidine base on the other, allows the helix to have a constant diameter. The chemical structure of this molecule is represented in figure 2, while figure 3 shows the pairing of the bases by two or three hydrogen bonds, linking the two strands together. In order to be able to form a picture clearly when discussing the possibility of a periodic model of the DNA molecule, a ribbon model is represented in figure 4. Here it is clearly seen that the DNA molecule looks like a sort of spiralled ladder, with both sides of the ladder winding around an imaginary vertical axis. The back bone are then the sides of the ladder, which are phosphate and sugar, while the steps represent organic bases interconnected by hydrogen bonds. These bases are arranged up and down the ladder and constitute a sort of four letter code, as represented in figure 3. This letter code is understood to spell out long words in a very large vocabulary, and carries information used in a biological reaction. The sequence of bases on a single chain does not appear to be restricted in any way, but it determined the sequence of bases in the other chain. So far, nobody has been able to read out this coded information. It is also believed that the genetic code in the DNA molecule is established by the particular sequence of the bases. There are three conformations of DNA, the A form, the B form and the C form. All of them are right handed double helices and cannot be separated without unwinding, but they have slight differences in their back bones and bases.

The A form has a rather tilted base, not horizontal as in the B form. It also has a slight difference in the structure of the back bone compared with the B form. The pitch of the helix and the number of bases

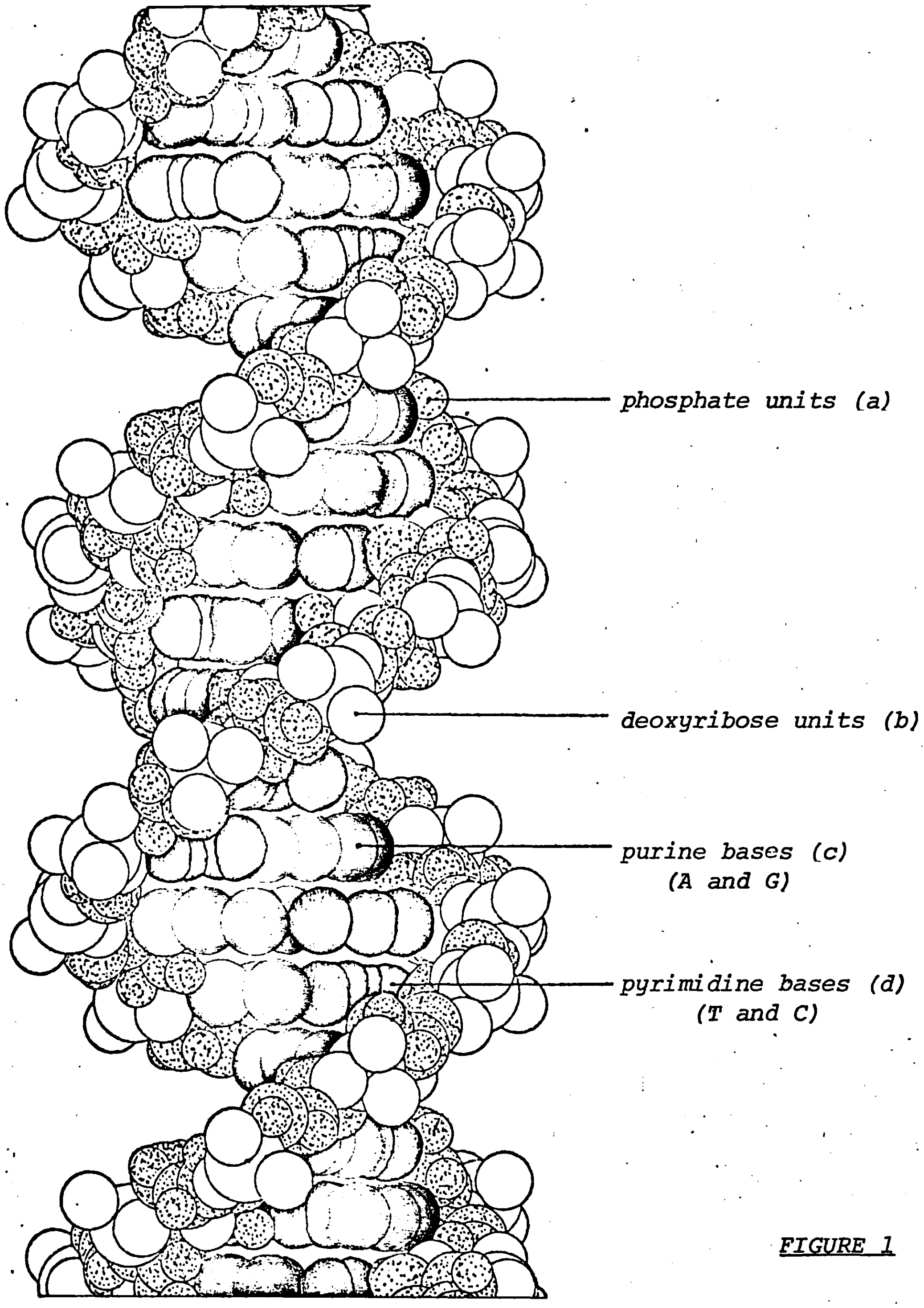


FIGURE 1

DNA molecule, as suggested by James D. Watson and F.H.C. Crick in 1953. Each strand consists of phosphate units (a) and deoxyribose units (b) in alternating sequence. The two strands are crosstied by bases, purine (c) and pyrimidine (d).

Adenine always paired with Thymine and Cytosine always paired with Guanine.

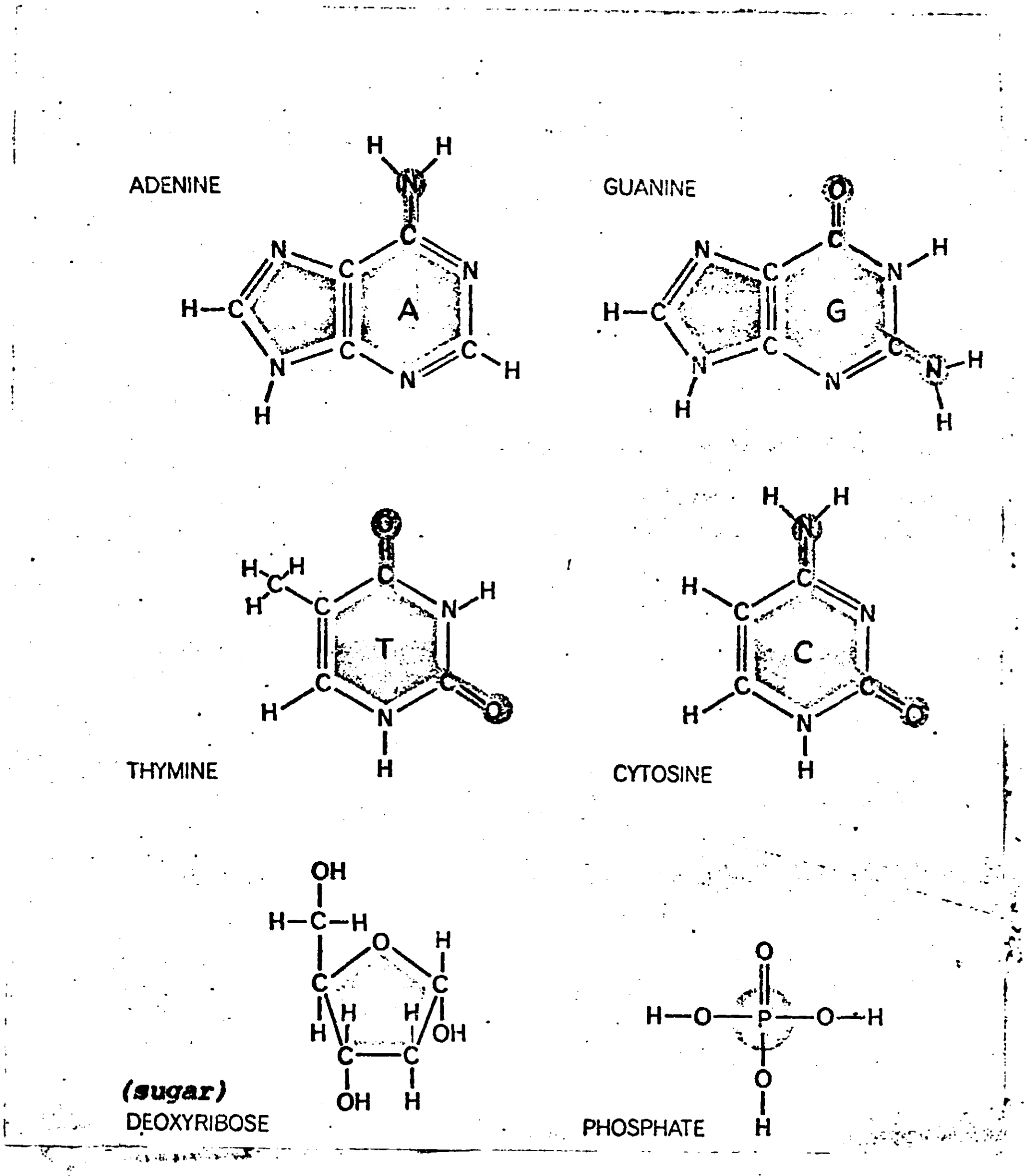


FIGURE 2

Chemical structure of each component of DNA macro-molecule. The shadowed area will be used in figure 3 to indicate how these components build up the DNA molecule.

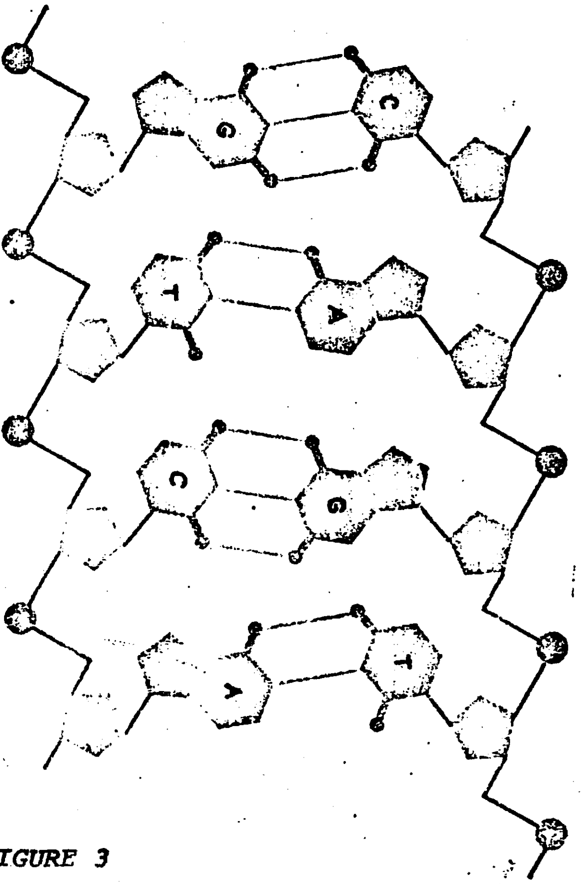
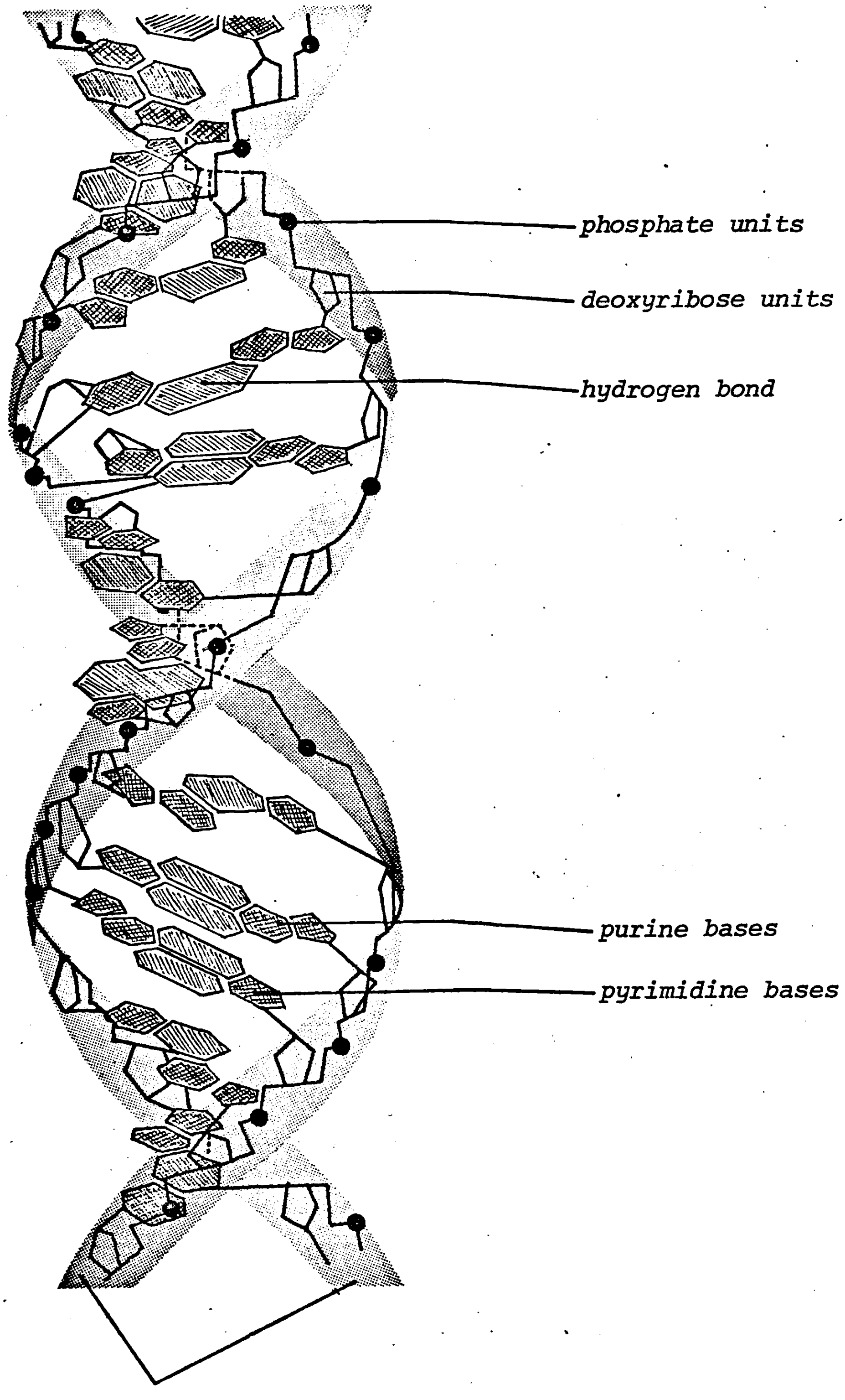


FIGURE 3

A = Adenine
B = Guanine
T = Thymine
C = Cytocine

This picture shows how the components of the DNA molecule are assembled together. It constructs a ladder, whose sides consist of attenuating units of deoxyribose and phosphate, while its steps are formed by bases, paired in a certain way, that is A with T, which are held together by two hydrogen bonds (indicated by two lines across the ladder sides), and G with C which are held together by three gydrogen bonds (the three lines connecting G and C in the picture).



these two ribbons, which contain phosphate units and deoxyribose units, construct what is called the back bone (sides of the ladder).

per turn in the A form is somewhat different from that of the B form. Similarly there are differences in the C form compared to the B form, but here it can be added that the weak H-bond is strengthened and the base stacking is reduced, because it usually occurs in concentrated salt solution. However, the form of interest in this report is the B form, because in this form the bases are arranged horizontally. In fact this form is very much like that suggested by Watson and Crick in 1953 (figure 1).

Table 1 summarizes the data on the three different forms.

Table 1

Form	Pitch	Residues per turn	Inclination of base pair from horizontal position
A (Na salt, 75% relative humidity)	28A	11	20°
B (Na salt, 92% relative humidity)	34A	10	0°
C (Li salt, 66% relative humidity)	31A	9.3	6°

The two strands make a complete revolution every 34A. It is accepted that there are 10 nucleotides, that is the monomer unit consisting of phosphate sugar and base, in each complete turn, so that in the B form of the DNA molecule each nucleotide takes up 3.4A of the strand. That is the distance between the planes of two superimposed bases belonging to the same chain is 3.4A. The adjacent bases are turned by 36° relative to each other in a plane perpendicular to the axis. The distance of a phosphorous atom from the fibre axis is 10A, and the distance between the adjacent phosphorous atom is 7.1A. The attachment of phosphate group and pentose sugars is not symmetrical, so that the sequence of attachment will appear different if read in one direction along the chain to the way it will appear if read in the other direction, as illustrate in figure 5 below. It is said that the two strands have opposite polarity.

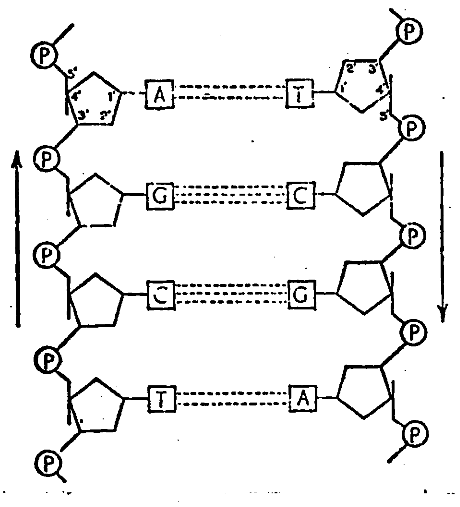


FIGURE 5

Shows the opposite polarity of the two strands caused by the different attachment of the phosphate groups and deoxyribose units in each chain (strand).

There are many exceptional forms of DNA, for example T_2, T_4 and some other viruses, but these are not of interest in this report.

2.2 Some considerations on the structure and properties of DNA in view of energy band calculations.

The first vital assumption that has to be made is that the DNA molecule can be treated as a periodic structure, while in reality it is most probable that the DNA molecule is not periodic. Refering to figure 6 below, if for the moment it is assumed that the bases are all equal, that is either all A's or T's or G's or C's and that there is no hydrogen bonding that interconnects the two sides of the ladder, and that

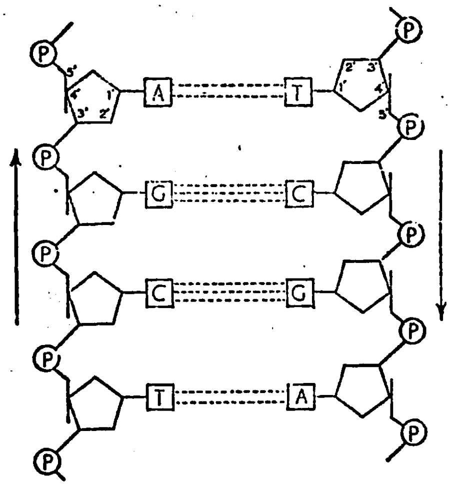


FIGURE 6

Schematic ladder model of DNA molecules.

A = Adenine
T = Thymine
G = Guanine
C = Cytosine

the strands (sides of the ladder) are not polarized, than it can be said that the molecule is a doubly periodic structure that goes on until the height of the total molecule is reached. This model can be treated as a one dimensional periodic structure, with a short period, say, L_1, that is the height of each step of the ladder, and a long period, say, L_2, that is the height of one turn of the strands. Unfortunately this simple structure does not exist, but nevertheless the idea can be applied to a model which is nearer to the real DNA molecule, as long as the assumption that it is periodic can be justified.

In figure 7, the ribbon model is once more represented, the two arrows along the strands indicating the direction of its polarity. The diameter of the helix

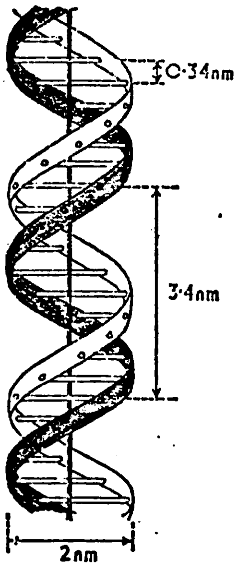


FIGURE 7

Ribbon model of DNA molecule.

is 20A, which means that its radius is 10A. The long period L_2 as mentioned previously, is 34A.
Watson and Crick assumed that the angle between the adjacent bases in the same chain is 36°. This means that there are 10 bases in one long period, the distance between adjacent bases being 3.4A. This is what was previously called the short period L_1.
In the following in particular, the B form of DNA is taken for further consideration. As previously discussed, this form is the most stable form of the DNA molecule, but the great advantage for calculation, is the fact that the bases are perpendicular to the imaginary axis of the double helix. The number of bases in DNA molecule is of the order of 10^5.
It is also vital to understand the role of chemical bonding and types of bond orbitals.
If ψ_a is a normalized atomic orbital of an atom a, which overlaps with ψ_b , a normalised atomic orbital of atom b, and r_{ab} is the internuclear distance between atoms a and b, then the physical concept that the formation of a chemical bond is due to the overlapping of the the two atomic orbitals can be written mathematically as:

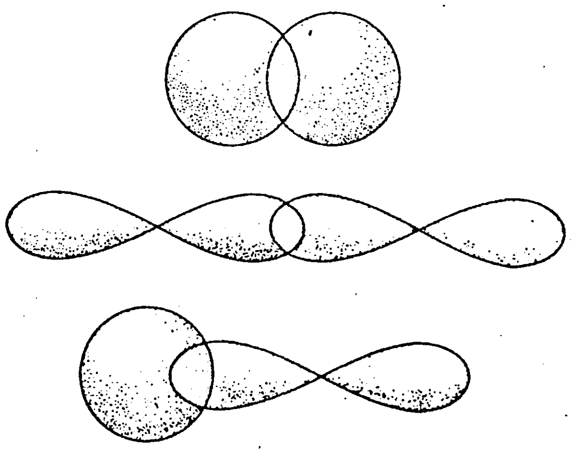
$$S_{ab}(r_{ab}) = \int \psi_a \psi_b \, d\tau \qquad \text{(overlap integral)}$$

where S_{ab} is a function of r_{ab}.

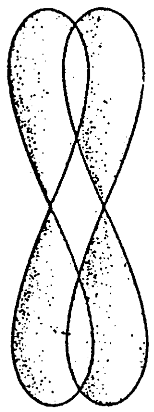
When $r_{ab} = 0$, then $S_{ab} = 1$, the normalisation condition. When $r_{ab} = \sim$, then $S_{ab} = 0$, which is the orthogonallity condition. When S_{ab} is not equal to zero it is a measure of the non-orthogonality of atomic orbitals. It is found that the overlap integral decreases rapidly as r_{ab} increases. That is why in the calculations that will be examined, all overlap integrals are assumed to vanish, exept those between nearest neighbouring atoms.

There are two types of bond orbital which arise from two different type of overlap, one is the so called σ (sigma) bonding and the other is called π (pi) bonding. The σ bonding is formed when the overlap is axial or frontal. Experimental data show that some properties of chemical bonds, for example bond length, bond energy, chemical reactivity,etc. are the same in different compounds, and hence the electrons forming a bond should be localized between the atoms they unite, forming a localized bond. These electron clouds give an idea of approximate size of the molecule. However, they do not give a significant contribution to the electronic character of the remaining bonds of the molecule. This type of bonding has no nodal plane.

The π bonding is formed when two p-orbitals overlap laterally. In order to satisfy the principle of maximum overlap, the electron clouds of the π atomic orbitals must be parallel. However, it is possible to calculate a bent bond if two adjacent orbitals are oriented so that their axis form an angle θ[2]. The electron cloud of a π orbital is not localized between the two nuclei forming the boding, but rather below and above the internuclear axis in such a way that it is necessary to consider both parts of the clouds as inseparable and constituting a single bonding orbital. The π bonding has a nodal plane that passes through the nuclei; This is a weak bonding and can easily be broken and disappear in an additional reaction. The electrons which form the σ orbital are called σ-electrons, while those which form the π orbital are called π-electrons.



σ- orbitals



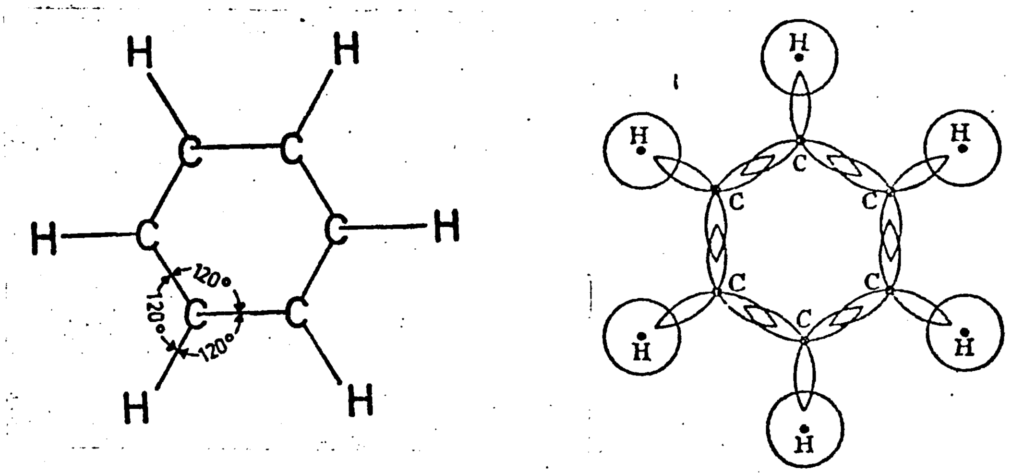
π-orbital

In conjugated and aromatic molecules, both types of bonding exist. The σ-electrons localized between the atomic nuclei constitute the molecular skeleton. They determine the geometric properties, such as bond length, bond energy, etc., but they contribute almost nothing

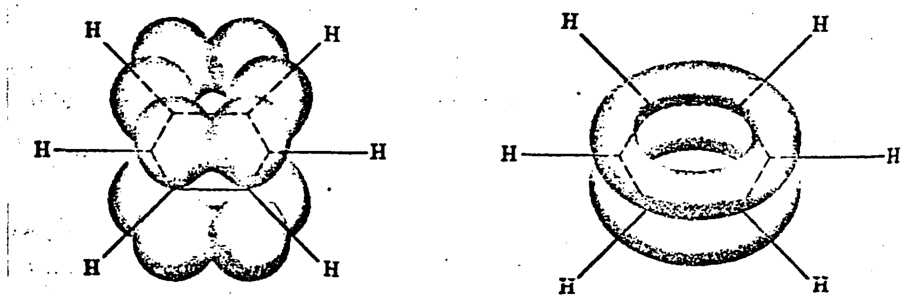
to the molecular chemical behaviour, while the π-electrons are responsible for chemical reactivity, carcinogenic activity, ultra violet absorption, etc.
However, the π-electrons in these compounds can no longer be said to be "localized", but they are "delocalized π-electrons". The prototype of aromatic compounds is benzene C_6H_6. According to the MO method, the MO wave function ψ_j can be written as a linear combination of the atomic orbitals (LCAO) ϕ_i where i and j are integers. Thus for benzene the MO orbital is

$$\psi = c_1\phi_1 + c_2\phi_2 + \dots\dots\dots\dots\dots\dots + c_6\phi_6$$

where the c_r are constants which are determined in such a way as to make the energy function stationary; r is an integer.
It is clear from the above equation that each of the 6 (six) π-electrons has to occupy one MO orbital which extends over all the carbon atoms, and is hence completely delocalized, that is the π-electrons are present all over the ring extending to all six carbon atoms, as sketched in figure 9.
The six hydrogen atoms are situated in such a way that the valence angle is 120° (figure 9a) and they are all in the same plane namely that of the carbon atoms. This implies that the carbon atoms are hybridiz-



(a) σ-Bond network in Benzene. The picture on the right shows the π-hybrids.



(b) (c)

FIGURE 9

Orbital model for benzene. The fundamental framework of the molecule is provided by σ-bonds shown in part a. At each carbon is located an unhybridized p_z orbital which can overlap with p_z orbitals on both neighboring carbons as shown in part b. The result is the double doughnut-shaped molecular orbital shown in part c. This is actually only one of the new molecular orbitals that forms.

ed in the trigonal state (sp^2) see figure 9b. The unhybridized p atomic orbitals (there are six of them) must overlap as much as possible in order that the bond may be as strong as possible. This can only be achieved if the molecular skeleton is planar and the unhybridized p atomic orbitals are perpendicular to it, and also if the bond length between the carbon atoms is between that of double bond and single bond.

In figure 9c the complete configuration of benzene is shown. The π-electrons move freely in the two doughnut shapes above and below the nodal plane.

Generally speaking a delocalized π-electron system is present in a molecule when the internuclear distances are between the values of that of a single bond, C—C, which is 1.54A, and that of a double bond C=C, which is 1.33A. The delocalized energy (resonance energy), can be defined as the difference between the energy of localized electrons and the total energy of π-electrons, or it is the energy difference between the calculated energy and the actual internal energy, as MO energies lead directly to obtaining the resonance energy.

The component of interest in the DNA bases are A, T, G, and C and they are aromatic as has been discussed previously. Hence the delocalized π-electron system is present. Therefore in the planning of energy band calculations three main points have to be borne in mind, they are:

- interaction between π-electron orbitals of single bases, for example that of A and T or G and C.
- interaction between π-electron orbitals of the adjacent superimposed bases belonging to the same chain.
- interaction between superimposed base pairs.

This geometrical consideration of the molecule in general will be of importance when proceeding with the energy band calculations for DNA.

References:

1. Watson, James D. and Crick, F.H.C., Nature, 171, 737-738 (1970)

2. Molecular Orbital Theory for Organic Chemistry, Streitwieser, A., John Wiley and Sons, New York, 1961

Background Reading:

The Biochemistry of the Nucleid Acids, J.N. Davidson, Chapman and Hall.

Introduction to Organic and Biological Chemistry, John R. Holum, John Wiley & Sons Inc.

Valence Theory, J.N. Murrel, S.F.A. Kettle, J.M. Tedder, John Wiley & Sons Ltd.

Valence, C.A. Coulson, Oxford University Press.

CHAPTER 3

General Survey of Methods employed in DNA energy band calculations

The method which has mostly been used and that will be mainly discussed is the Molecular Orbital Method. The general geometry of the molecule and its constituents, the type of bonding and the type of electrons involved in the structure, have been discussed previously. Consider, first of all homonuclear molecules, that is those which have carbon (C) atoms only. In the MO method the molecular orbitals ψ ; can be written as linear combination of atomic orbitals ϕ_k where k = 1,2,.....,n, namely the number of C atoms in aromatic or conjugated system.

Hence:

$$\psi_j = \sum_{k=1}^{n} c_{jk} \phi_k \qquad ....(1)$$

where c_{jk} is the coefficient of the atomic orbitals ϕ_k in the molecular orbital ψ_j.

This is what is called Linear Combination of Atomic Orbital - Molecular Obital Method or abreviated LCAO-MO.

It is generally assumed that the best set of coefficients c_{jk} is that which minimizes the total energy. This leads to the so called secular determinant, equivalent to a secular equation for the energy. In such a complicated problem, the equation is based on a number of approximations.

The better the approximation method used, the better the results should be.

In MO theory, the σ-electrons and the π-electrons are considered to be independent from each other. This is justified because of two reasons. One is that the electrostatic interaction between σ- and π-electrons has been proved, by several workers, to be negligible [1,2,3], the other is that the exchange between σ- and π-orbitals is also negligible, except for electron spin resonance.

Before proceeding to discuss some particular methods associated with the MO theory, some fundamental mathematical formulae will be outlined.

MO theory depends on the overlapping of the atomic orbitals in order that the linear combination of atomic orbitals can be formed. Therefore first of all the overlap integral has to be considered.

The overlap integral is:

$$S_{ks} = \int \phi_k \phi_s \, d\sigma \ldots\ldots\ldots\ldots \quad (2)$$

where k and s are different atoms.

Secondly, the energy operator, the Hamiltonian H is:

$$H = T(i) + U_{core}(i) + \frac{1}{2} \sum_{ij} \frac{e^2}{r_{ij}}$$

where

$T(i)$ is the Kinetic Energy operator for electron i

$U_{core}(i)$ is the potential energy operator for electron i in the field of the core. U includes all electrons other than π-electrons.

$\frac{1}{2} \sum_{ij} \frac{e^2}{r_{ij}}$ is the electrostatic repulsion between π-electrons.

It has to be borne in mind that the inner core electrons are lumped with the nuclei, simply reducing the nuclear charge.

$\underline{H}$ describes the motion of the electrons for a fixed position of the nuclei. It depends on the position of the nuclei, and not on their momentum.

Thirdly, the Coulomb integral. For the time being it is assumed that the molecule contains only carbon atoms, therefore the so called Coulomb integral in MO theory, given a symbol α, represents the energy of the 2p electron of the Carbon atom which forms the π bonding. It thus includes also the kinetic energy besides the Coulomb energy. This energy is the energy of the π-electron in the presence of the other nuclei and it is a measure of the electronegativity of that atom. It should be borne in mind that the value of α is altered by the presence of any atom other than Carbon. The Coulomb Integral, α, depends on:

- The nuclear charge
- The type of orbital

It is usually written as: $\alpha_r = \int \phi_r \, \underline{H}_\pi \, \phi_r \, d\tau$ ....(5)

Its numerical value is negative and it represents the energy required to remove an electron from the π-orbital to infinity.
Fourthly, the so called Resonance Integral, given the symbol β and usually written as:

$$\beta_{ks} = \int \phi_k \underline{H}_\pi \phi_s \, d\tau \quad \ldots\ldots\ldots\ldots \quad (6)$$

Physically this represents the energy of interaction of two atomic orbitals ϕ_k and ϕ_s.
Its numerical value is negative and depends on the interatomic distance of atoms k and s.

1. The simplest and most daring approximation method in MO theory is the so called Hückel Approximation method, which was developed by Hückel in 1931[4]. This approximation is normally abbreviated HMO (Hückel Approximation Method).

   In his approximation the following assumptions are made:

   a. The relationship of the Hamiltonian $\underline{H}$ to the complete Hamiltonian of the molecule is not specified, the electron repulsion term is neglected.

b. The secular equation is simplified by making the following approximations:

- The overlap integrals S_{ks} are taken to vanish. A normalized atomic orbital can always be taken, that means $S_{kk} = 1$.
- The Coulomb Integral α_r is assumed to be the same for each carbon atom, this means setting all diagonal elements corresponding to the Hamiltonian $\underline{H}_{\pi}$ equal to a constant, say, α .
- The Resonance Integral β_{ks} is assumed to be constant between two adjacent atoms and is equal to Zero, if the atoms are not directly bonded (not adjacent). This means that the matrix elements between the directly bonded atoms (adjacent atoms) are constant, say, equal to β and the rest are Zero (neglected).

The neglect of the interelectronic repulsion among π-electrons has the result that the HMO method fails to explain spectra and it gives ionisation potentials not in agreement with experimental results.

To assume that $S_{ks} = 0$ is rather difficult to accept, because the calculated overlap between two π-orbitals for carbon at the distance of adjacent atoms in an aromatic hydrocarbon is 0.25[5].

However, when the results of the Hückel Approximation are compared with those of modern Semiempirical Methods, they agree fairly well.

The Molecular Orbitals extend over several identical nuclei, therefore it seems justified to assume that the Coulomb Integral α is the same for all nuclei.

The assumption that the Resonance Integral β can be taken to be constant when atom k and atom s are directly bonded, and otherwise equal to zero, is also justified because β depends mainly on the internuclear distance k-s.

The simplest example for the application of the Hückel Approximation in a single molecule is Benzene.

Benzene has a plane ring, therefore the electronic wave function must be either symmetric (σ-electrons) or antisymmetric (π-electrons) under reflection in the plane. The π molecular orbitals are built out of p atomic orbitals a_j. The correct LCAO is

$$\sum_j a_j e^{ikj}$$

If the function is to be single valued $e^{6ik}=1$ where $k = 0 \, , \pm\frac{\pi}{3} \, , \pm\frac{2\pi}{3} \, , \pi$

The corresponding energies go in this order. One finds that there is one electron per atom in the π orbitals, six in all. They fill the lowest states $k = 0 \, , \pm\frac{\pi}{3}$ (two electrons with opposite spins in each state).

Since the six atoms are treated exactly in the same way, there is no need of introducing single and double bonds.

The energy values are eigenvalues of a matrix involving the already mentioned α and β integrals. The β integrals are considered only for the nearest neighbours.

The secular equation for benzene is

$$\begin{vmatrix} \alpha - E & \beta & 0 & 0 & 0 & \beta \\ \beta & \alpha - E & \beta & 0 & 0 & 0 \\ 0 & \beta & \alpha - E & \beta & 0 & 0 \\ 0 & 0 & \beta & \alpha - E & \beta & 0 \\ 0 & 0 & 0 & \beta & \alpha - E & \beta \\ \beta & 0 & 0 & 0 & \beta & \alpha - E \end{vmatrix}$$

For more complicated molecules there are several values of the α and β integrals. Other examples will be given when considering in detail the DNA calculations. When, instead of a single molecule, a periodic chain is considered, the conceptual modifications to the Hückel method are quite small.• Consider a chain with N primitive cells and n atoms in each primitive cell. Then the combination of atomic orbitals, assumed, as an example, to be 2p orbitals is

$$\psi(\underline{r}) = \sum_{jl} c_{jl} \, \psi_{2p} \left[\underline{r} - (\underline{r}_1 + j\underline{a}) \right]$$

where

$\underline{a}$ is the primitive translation vector

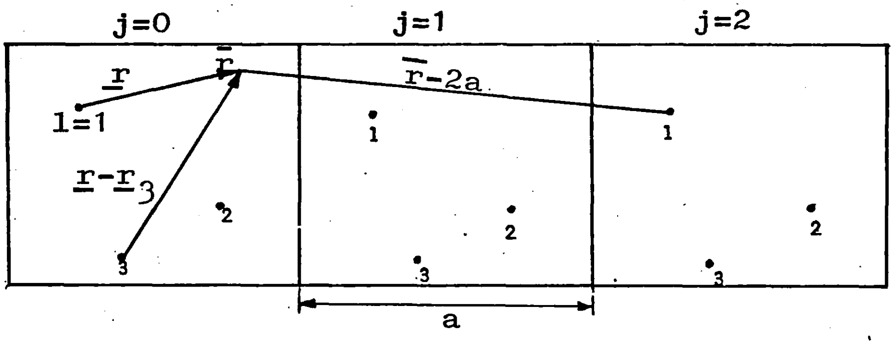
j denotes the different primitive cells $0 \leq j \leq N-1$

denotes the different atoms in one primitive cell

$$1 \leq l \leq n$$

$\underline{r}$ is measured from the atom $j = 0 \quad l = 1$

The coefficients c_{jl} will have to be determined



$\underline{r}$ any point

By applying the Born-von Karman periodicity condition one finds in the usual way that

$$c_{jl} = c_{lk}e^{ikja}$$

One then uses the above wavefunction ψ to calculate the expectation value of the energy of a de-localized electron moving in the periodic potential of the one-dimensional crystal

$$H = \int \psi^* H \psi \, dV$$

and one minimises it by imposing the conditions $\frac{\partial H}{\partial c_{lk}} = 0$ for all coefficients.

The actual potential does not come explicitly in the calculation, but only through the parameter β of the Hückel simplified method.

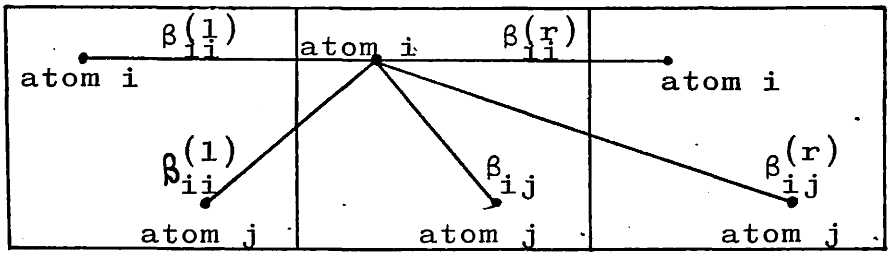
The minimum condition provides a system of algebraic homogeneous equations for the unknown c_{lk}. For non-zero solution the determinant of the coefficients of the equations must vanish.

With the abreviations

$$\gamma_{ii} = \alpha_i + \beta_{ii}^{(r)} \; e^{ika} + \beta_{ii}^{(l)} \; e^{-ika}$$

$$\gamma_{ij} = \beta_{ij} + \beta_{ij}^{(r)} \; e^{ika} + \beta_{ij}^{(l)} \; e^{-ika}$$

(r) stands for right and (l) stands for left



The determinant is

$$\begin{vmatrix} \gamma_{11}-E & \gamma_{12} & \cdots\cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22}-E & \cdots\cdots & \gamma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{n1} & \gamma_{n2} & \cdots\cdots & \gamma_{nn}-E \end{vmatrix} = 0$$

There are few other methods developed from the simple Hückel method, for example:

2. Mulliken-Wheland method. The difference between this method and the simple Hückel method lies in the overlap integral S. Instead of taking it to be equal to zero, Mulliken and Wheland took it to be equal to 0.25 for nearest neighbours.

3. Goepert-Mayer and Sklar in 1938[6], showed that by including the repulsion term in the Hamiltonian, this approximation can be made more reliable.
However this method is considered as too theoretical and also by introducing the electrons repulsion term, one is faced with the electron repulsion integrals, which involve a very laborious calculation.

There are many more methods based on MO theory, which have been developed, one of the most important being the Self-Consistent Field Molecular Orbital Method, abbreviated (SCF-MO). This method was first suggested by Roothaan's. This method is sometimes also called SCF-LCAO-MO method, because it uses LCAO-MO's as its basis. However, most biological molecules are heteronuclear molecules. This means that there is at least one or more heteroatoms in the π -system in addition to carbon ones. In the DNA molecule these are usually Nitrogen(N) or Oxygen(O) atoms. The MO method which includes heteronuclei was first introduced by Pauling and Wheland[7] and therefore the method is known as Pauling-Wheland method.

1. Pauling-Wheland method. One feature of this method is that the Coulomb integral α_{XC} and and the Resonance Integral β_{XC} for the heteroatom X and the Carbon atom C are both expressed in terms of the Resonance integral for Carbon-Carbon atoms. The values of the Coulomb Integral α_C and the Resonance Integral β_{CC} of benzene are usually taken as standard values.

   An auxiliary inductive parameter, η, is used to take into account the inductive effect of heteroatom X on neighbouring Carbon atom C_i, where i indicates the position of the Carbon atom with respect to the heteroatom X. Therefore the new parameters can be written in the form:

$$\alpha_X = \alpha_C + \zeta_X \beta_{CC}$$
$$\beta_{CX} = \rho_{CX} \beta_{CC}$$
$$\zeta_{c_i} = \eta\zeta_X$$

   A variety of methods is available for estimating the parameter δ_X and ρ_{CX}

2. Pariser and Parr method [8,9]

This is a semiempirical model and it is based on the method of antisymmetrized products of molecular orbitals in the LCAO approximation, including configuration interaction, but it incorporates empirical elements. The Hamiltonian is written as

$$H = T(i) + U_{core}(i) + \tfrac{1}{2} \sum_{ij} \frac{e^2}{r_{ij}}$$

as has been discussed previously.

The wave functions is a linear combination of anti-symmetric products

$$\phi_\Lambda = \frac{1}{\sqrt{n!}} \begin{vmatrix} (\phi_1 \alpha)^1 & (\phi_1 \beta)^1 & (\phi_1 \alpha)^1 \\ (\phi_1 \alpha)^2 & (\phi_2 \beta)^2 & \cdots\cdots \\ (\phi_1 \alpha)^3 & \cdots\cdots & \cdots\cdots \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \end{vmatrix}$$

where

Λ is an index which indicates the particular configuration, i.e. the way the electrons are assigned to particular molecular orbitals

$$\phi_i = \sum_p C_{ip} X_p$$

the X_p are the atomic orbitals centred around the various nuclei.

and, for instance $(\phi_1\beta)^3$ is the one-electron function for electron 3 in orbital ϕ_1 with spin function β (spin down)

The expectation value of the energy

$$E_\Lambda = \phi^*_\Lambda H\phi_\Lambda \, d\tau_1 \, d\tau_2 \, d\tau_3 \ldots\ldots$$

for the wave function ϕ_Λ can be expressed as

$$E_\Lambda = \sum_i I_i + \tfrac{1}{2} \sum_{ij} (J_{ij} - K_{ij})$$

the sum being over the occupied molecular orbitals where the core energy for molecular orbital ϕ_i is

$$I_i = \int \phi(1)(T+U)(1)\phi_i(1) \, dv_1$$

where ϕ_i is either $(\phi_i \alpha)$ or $(\phi_i \beta)$

The Coulomb integral between M.O. i and j is

$$J_{ij} = \int \phi_i(1) \, \phi_j(2) \, \frac{e^2}{r_{12}} \, \phi_i(1) \, \phi_j(2) \, dv_1 dv_2$$

The exchange integral between i and j

$$K_{ij} = \int \phi^*_i(1) \, \phi^*_j(2) \, \frac{e^2}{r_{12}} \, \phi_j(1)\phi_i(2) \, dv_1 dv_2$$

The K'_{ij} appearing in the formula for the energy are equal to K_{ij} if ϕ_i and ϕ_j have the same spin, but $K'_{ij} = 0$ if ϕ_i and ϕ_j have different spins.

In general a single ϕ_Λ will not be a good approximation to the wave function ϕ_s of a certain state because that state is a "mixture" of several configurations, for example, there may be a finite probability of the appearance of excited states.

Then one writes

$$\phi_s = A_1\phi_1 + A_2\phi_2 + \cdots\cdots$$

and one carries out a "configuration interaction" calculation almost universally by the variational method.

If

$$S_{mn} = \int \phi_m^* \phi_n dv$$

and

$$H_{mn} = \int \phi_m^* \phi_n dv$$

then the energy levels are given by the roots of the secular equation

$$\begin{vmatrix} H_{11} - ES_{11} & H_{12} - ES_{12} & \cdots\cdots \\ H_{21} - ES_{21} & E_{22} - ES_{22} & \cdots\cdots \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \end{vmatrix} = 0$$

The integrals H_{mn} can be expressed through generalisation of the core energy integrals, Coulomb and exchange integrals. The Exchange integrals are denoted by

$$(pq/rs) = \int \phi_p(1)\phi_r(2) \frac{e^2}{r_{12}} \phi_s(2)\, dv_1\, dv_2$$

We must then have semiempirical methods of evaluating all these integrals, and also some criterion for deciding the extent of configuration interaction that has to be included.

With configuration interaction there is a core integral in each matrix element

$$I_{ij} = \int \phi_i^*(1) \left[(T+U)(1) \right] \phi_j(1) dv_1$$

or in terms of the atomic orbitals

$$I_{ij} = \sum_p \sum_q C^*_{ip} C_{jq} H^{core}_{pq}$$

$$H^{core}_{pq} = \int X^*_p(1)\left[T + U\right] X_q(1)dv_1$$

In any case we need the LCAO coefficients C_{ip} and the atomic orbitals matrix elements H^{core}_{pq}

The rather drastic assumption made here is that the C_{ip} are arbitrary, that is they may be chosen in the most convenient way, subject only to the ortho-normalisation condition. The excuse for this is that large deviations from the correct (self-consistent) values may be compensated by adding more and more configurations.

No doubt, theoretical chemists know how to put a lot of common sense in their arbitrary choice.

The diagonal elements of H^{core}_{pq} are the α integrals (improperly also called Coulomb Integral) of the simple method. The non-diagonal elements are the β integrals (resonance integral) assumed to be the same between a given pair of atoms, no matter what molecule the atoms are in. They are treated as basic empirical quantities.

As for α - integrals, the Hamiltonian operator in them is

$$(T+U)(1) = T(1) + U_p(1) + {}_{p}\Sigma\, U_q(1) + \sum_{r} U^*_r(1)$$

where the atoms q are charged and atoms r are uncharged so that a $*$ denotes a neutral atom.

$$U_q(1) - U^*_q(1) = - \int X^*_q(2) \frac{e^2}{r_{12}} q_q(2)\, dv_2$$

Introducing the ionisation potential

$\left[T(1) + U_p(1)\right] \quad X_p(1) = W_p X_p(1)$ we then can write

$$\alpha_p = H^{core}_{pp} = W_p - \sum_{q \neq p} \left[(pp/qq) + (qipp)\right] - \sum_{r} (ripp)$$

where $(qipp) = - \int U^*_q(1) \quad X^*_p(1) X_p(1) dv$

This is called Coulomb penetration integral between X_p and the neutral atom q.

This formula is used by Ladik for his matrix F though he gives different names to various integrals. Finally there are the integrals J_{ij} and K_{ij}, written together as (pq/rs), Coulomb repulsion integrals. These are complicated and it is suggested to simplify them by taking

$$\sum_{p} C^*_{ip} C_{jp} = \delta_{ij}$$

So that atomic orbitals replace molecular orbitals in the integrands.

Then (pq/rs) = linear combination of integrals over atomic orbitals of the form (pp,qq). Tables of these integrals are in existence.

Thus the Pariser-Parr method makes use of the following quantities:

1. Atomic ionisation potential
2. "Band Resonnance" Integrals
3. Coulomb repulsion integrals (pp/qq)
4. Coulomb penetration integrals (qipp)

All or some of them may be regarded as empirically adjustable parameters. Indeed this is the essence of the method, that quantities generally regarded as purely theoretical may be considered as semi-empirical, that is may be adjusted until agreement with selected experimental data is reached.

In applying this approximation it appears that some but not very much configuration interaction should be included. It was found that the resonance integrals obtained theoretically were in a good agreement with those obtained experimentally. With small adjustment it was found that the resonance integral can be carried over from molecule to molecule. The theory can be applied to molecules containing tripple bonds. Adaption to

σ -bonds is feasible. Rationalisations of the electronic repulsion integral and penetration integral curves are possible in terms of varying effective charges but this is not recommended.

3. Pariser-Parr and Pople approximation (PPP-approximation).

   This was first introduced by Ladik[10]. It is the semiempirical method, initially devised for molecules with de-localized π -electrons, which includes, in a simplified way, Configuration Interaction (CI).

   It should be mentioned that in a recent survey by R.E. Christoffersen[11] this method is not even mentioned, since programmable extensions of it are now available making its original form obsolete. It is thus very possible that despite the added complications, the method is still too simple to deal with such a complicated structure as DNA.

   In a crystal like that of DNA the interaction between molecules in different cells is probably small, thus it is legitimate to consider only a small number of selected configurations.

We remember that the Hückel approximation for a one dimensional crystal consisted in writing, for an orbital p (2p 3s 4d ....)

$$\psi_p = \Sigma_{jl} \quad c_{p,jl} \quad \psi_p \quad [\underline{r} - (\underline{r}_l + j\underline{a})]$$

where $\psi_p [\underline{r} - (\underline{r}_l + j\underline{a})]$ is the p atomic orbital of the l^{th} atom in the j^{th} elementary cells. The coefficients c satisfy the Bloch conditions

$$c_{p,jl}(k) = e^{ikj} c_{p,l}(k)$$

and they are the eigenvectors of the matrix eigenvalue problem

$$(\gamma) c_{k,l} = \varepsilon_l (k) \quad c_{k,l}$$

where l denotes the various bands the elements of the matrix γ are

$$\gamma_{ls} = \beta_{ls} + \beta_{ls}^{(r)} \quad e^{ika} \quad + \beta_{ls}^{(left)} \quad e^{-ika}$$

$$\beta_{ls} = <\psi_l (r - r_l - ja) \, |H| \, \psi_s (r - r_s - ja) >$$

where j is j+1 in (right)

j-1 in (left)

With the PPP method the matrix (γ) is replaced by a much more complicated matrix (F) whose elements, first for the case of a single molecule, are

$$F_{11} = -I_1 + \tfrac{1}{2}p_{11}(I_1 - E_1) + \sum_{s=1}(p_{ss} - z_s)\ \gamma_{1s}$$

The F_{11} may be considered as replacing the α's of the Hückel method, while the β's are replaced by

$$F_{s1} = \beta_{1s} - \tfrac{1}{2}p_{1s}\ \gamma_{1s}$$

where

I_1 = ionization energy of atom 1

E_{11}= electron affinity of atom 1 , i.e. ionization energy of the negative ion.

$$P_{11} = 2\sum_{p=1}^{n} |c_{p1}|^2 = \text{"charge density"}$$

c_{p1} is the coefficient of the atomic orbital p in the molecular orbital 1

$$P_{s1} = 2\sum_{p=1}^{n} c^*_{p1}\ c_{ps} = \text{"bond order"}$$

Z_s = effective nuclear charge of atom s

β_{1s} = energy integral as before

$= \int \phi_1^* (1) H_{core} \phi_s (1) dv_1$

γ_{1s} = Coulomb integral as before

$$= \int \phi_1^*(1)\ \phi_s^*(2) \frac{e^2}{r_{12}}\ \phi_1(1)\ \phi_s(2) dv_1 dv_2$$

The charge densities and bond orders contain the coefficient to be determined. Thus initial values of the c's must be guessed, or worked out by means of a simpler calculation (e.g. Hückel).

The eigenvectors of the matrix (F) should be the same, if they are not the calculation must be repeated with different initial values until self-consistency is achieved.

The β -integrals are taken as adjustable parameters (generally taken different from zero only between nearest neighbours).

The γ -integrals are approximated in a more or less plausible way.

If, instead of a single molecule, a one dimensional chain is considered, the elements of the matrix (F) must be generalized in the same way as, in the simple method, the elements β_{1s} were generalized to the elements γ_{1s} involving $\beta^{(right)}$ and $\beta^{(left)}$.

Here the p's the β 's and the γ 's must be generalized. The initial values for the $p^{(right)}$ and $p^{(left)}$ can again be found by a Hückel calculation.

In a self-consistent procedure one should calculate the charge density for each k-state and add them up, an impossible procedure. Nothing is said about this; one assumes that the charge densities in the p's states are kept constant for all states of a band.

So far only self-consistency for one configuration has been taken into account. Next some CI must be included by adding to ψ further sums over different atomic orbitals, generally corresponding to excited states. Spin should also be considered by introducing wave functions which have the appropriate symmetry for singlet, triplet,........, states. It was assumed that only interactions with singlet excited configurations were important. There is in principle no difficulty in setting up the inter-configurational matrix elements, but the determinant of the secular equation becomes very large. It is suggested than an averaging procedure like that of Belezney and Biczo[12] may simplify matters and still provide enough accuracy.

References:

1. Parks, J.M. and Parr, R.G. J. Chem. Phys., 32, 1657 (1960)

2. Fueno. T., Ann. Rev. Phys. Chem., 18, 303, (1961)

3. Sinanoglu, O. and Martensen, E.M. J. Chem. Phys., 34, 1078 (1961)

4. Huckel, E., Z.Physik, 70, 204 (1931)

5. Ruedenberg, K., J. Chem. Phys., 34, (1961)

6. Goeppert-Mayer, M., and Sklar, S.L., J.Chem. Phys., 6, 645 (1938)

7. Pauling, L., and Wheland, G.W., J.Am.Chem.Soc., 57, 2086 (1935)

8. Pariser, R., and Parr, R.G., J. Chem. Phys., 21, 466, (1953)

9. Pariser, R., and Parr, R.G., J. Chem. Phys., 21, 767 (1953)

10. Ladik, J., Acta Phys., Hung. Tom. XVIII. Fasc. 3, 186, (1964)

11. Ab initio Calculations on large molecules Ralph E. Christoffersen.

Chapter 4

4.1 Survey of Calculations carried out by Ladik and others.

Ladik's idea is that one can start by calculating the electronic structure of an infinite chain consisting of only one kind of base. This is then followed by treatment of a model of an infinite double helix consisting a pair of bases, i.e. only A-T, or G-C. The idea is to follow this with calculations on more complicated periodic structures. He states "it seems probable that with the aid of the results obtained for different complicated periodic model sequences it will be possible to get a rather good approximation of the electronic structure of a real, non periodic DNA molecule."

The first step is to work out the electronic structure of single bases. The method used initially was the simple Hückel Method.

Diagram 1 represents the different steps for the approximation of the electronic structure of DNA. Each line in the figure indicates schematically a single base in side view.

The dotted line indicates the smallest periodically repeated unit.

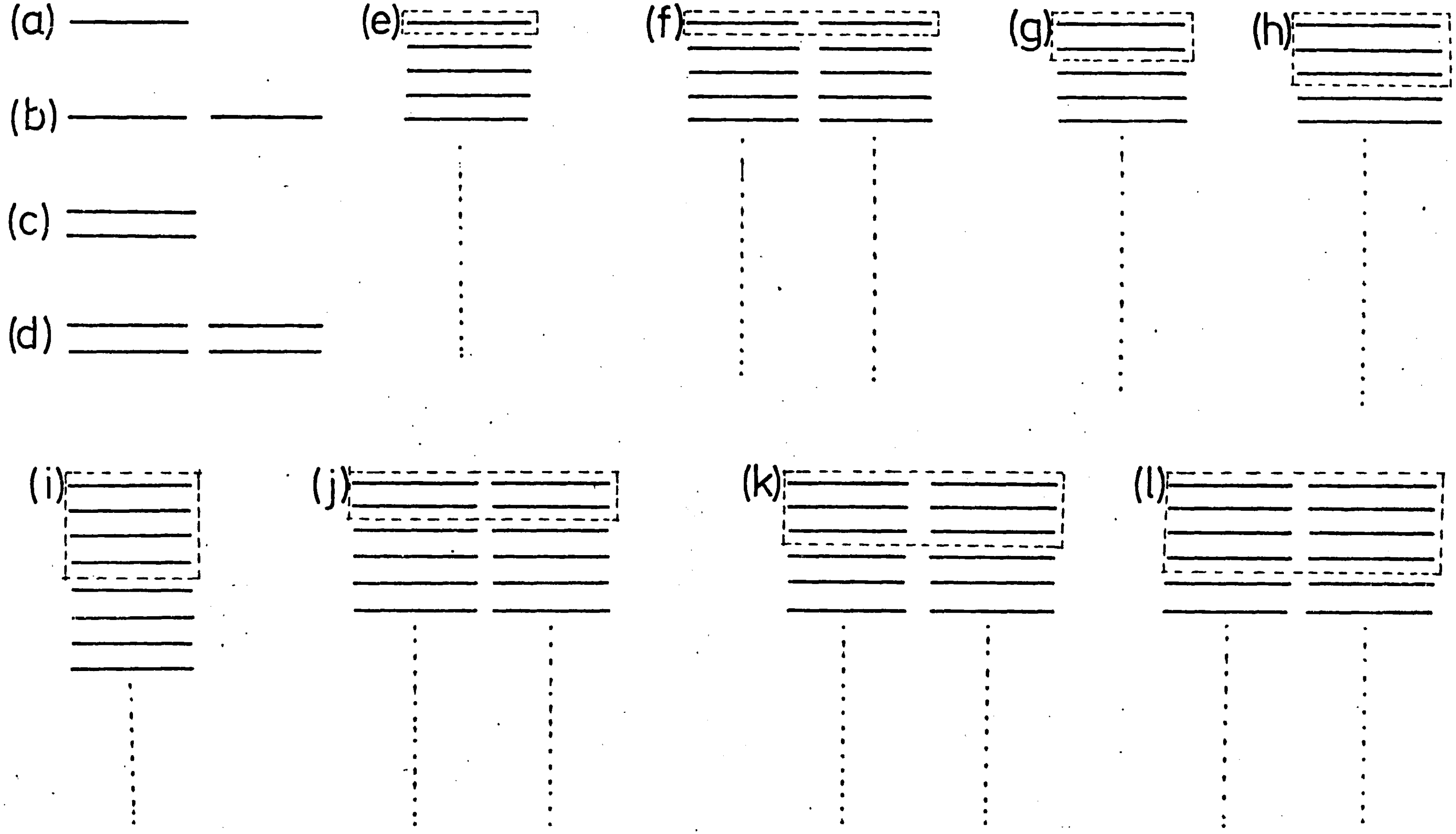


DIAGRAM 1

(a) single base

(b) base pairs

(c) two superimposed bases

(d) two superimposed base pairs

(e) infinite chain of the same nucleotide

(f) infinite chain of the same base pairs

(g) infinite chain of periodically repeated two different bases,

e.g. ATATATAT........

(h) infinite chain of three different bases periodically repeated.

(i) infinite chain of periodically repeated four different bases.

(j) infinite chain of periodically repeated two different base pairs.

(k) infinite chain of periodically repeated three different base pairs.

e.g.

A G T A G T A G T A G T ....

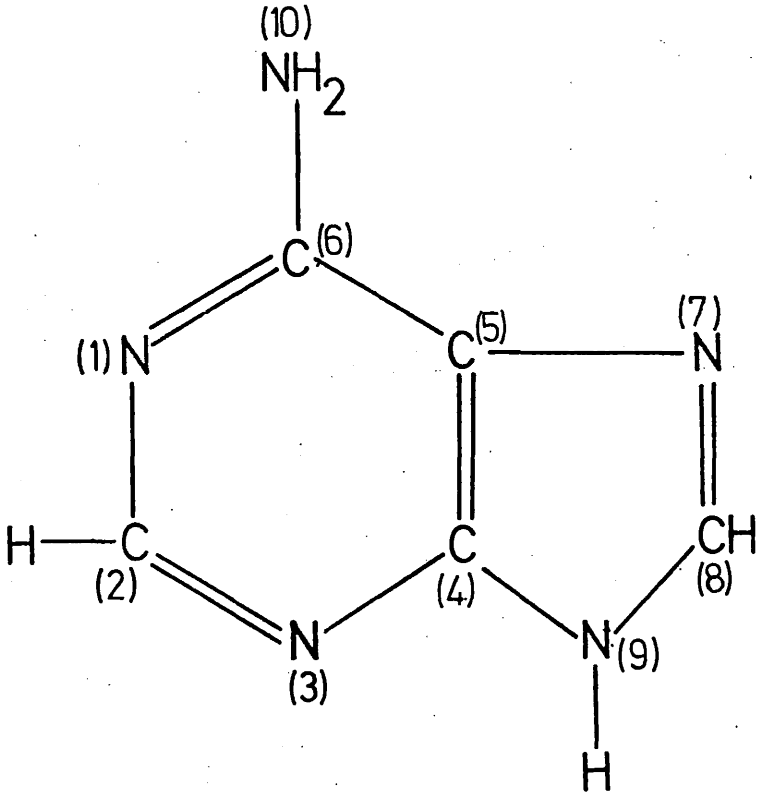
T C A T C A T C A T C A ....

(l) infinite chain of periodically repeated four different base pairs.

Energy levels of one base:

As explained before, the energy values are the eigen values of a matrix involving the overlap integrals α and the resonance integrals β , where the latter are considered for nearest neighbours only.

The values for α- and β-integrals were taken from previous calculations on aromatic molecules. For instance, for Adenine



Molecular structure of molecule.
The numbers between brackets indicate
the position of atoms in the molecules.

Atom 1 α = 0.58 β_{CC}
2 = 0.14
3 = 0.58
4 = 0.07
5 = 0.07
6 = 0.00
7 = 0.58
8 = 0.14
9 = 0.58
10 = 0.90

α is expressed in unit β_{CC}, which as has been discussed before is Carbon atoms resonance integral.

For atom number 6, α is taken conventionally to be equal to zero.

β (1,2) is 1.00
(2,3) 1.00
(3,4) 1.00
(4,5) 1.00
(5,6) 1.00
(4,9) 1.10
(5,7) 1.00
(7,8) 1.00
(8,9) 1.10
(6,10) 1.00
(1.6) 1.00

β (4,5) is taken to be equal to 1.00 conventionally.

The secular equation is then

$$\begin{vmatrix}
(0.58-E) & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & (0.14-E) & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & (0.58-E) & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & (0.97-E) & 1 & 0 & 0 & 0 & 1.1 & 0 \\
0 & 0 & 0 & 1 & (0.07-E) & 1 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & (-E) & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & (0.58-E) & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & (0.14-E) & 1.1 & 0 \\
0 & 0 & 0 & 1.1 & 0 & 0 & 0 & 1.1 & (0.58-E) & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & (0.9-E)
\end{vmatrix} = 0$$

This produces eigenvalues (energy levels)

$E = 2.77$

2.13

1.75

1.15

1.03

0.53

E = -0.79
-1
-1.39
-2.02

All to be multiplied by β which is a negative quantity. The absolute value of β is just over 3.3eV.

The first 6 levels are bonding and are filled, the last four levels are antibonding and empty.

The first excitation energy is thus 0.53 + 0.79 = 1.32.

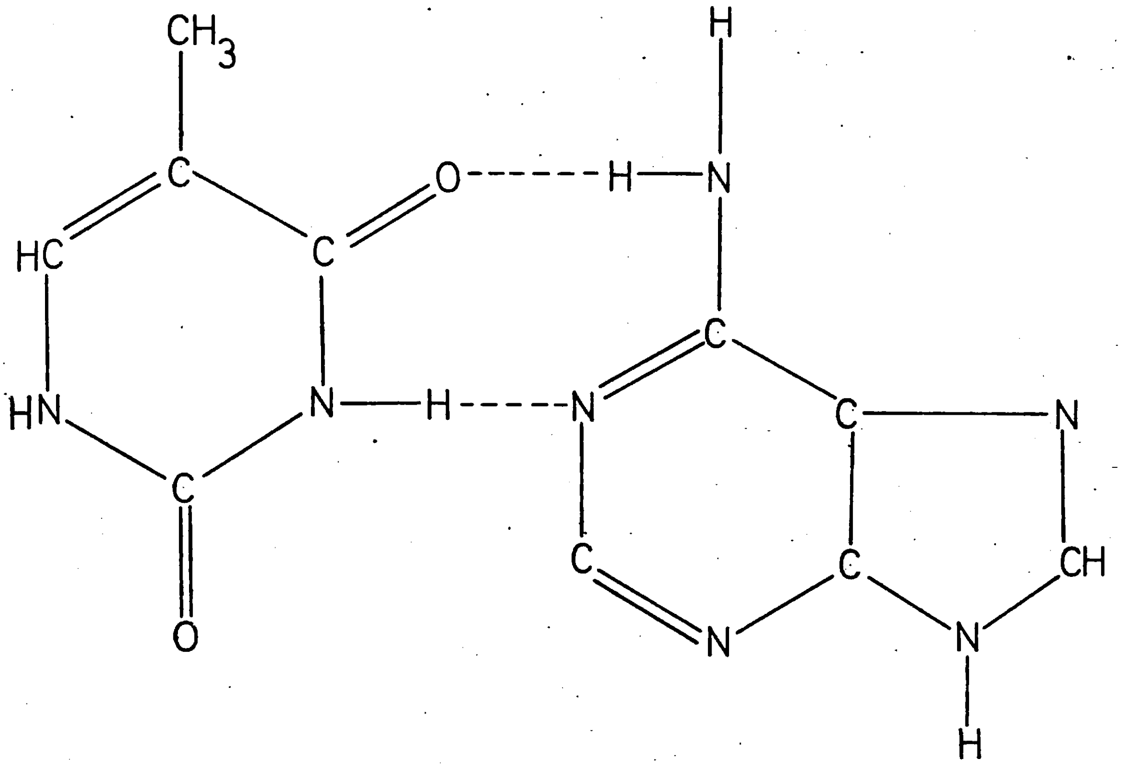
From the eigenvectors, one can calculate, using a standard way, the charge densities, oscillator strength, etc.

The excitation energy is in fair agreement with experiments; however the oscillator strength for this transition is calculated as 0.73 instead of the experimentally determined 0.30.

It seems to be a characteristic of this type of calculations that, while the energies may not be too bad, the wave functions are a poor approximation. Configuration interaction (CI) must be taken into account.

Energy levels of two adjacent bases:

The next step is the determination of energy levels and charge densities for a base pairs. For example Thymine-Adenine.



Molecular diagram for base pairs
Thymine - Adenine.

Here we have 20 atoms instead of 10, thus a larger secular equation. The additional problem is how to deal with the hydrogen bonds joining the two bases, that is whetner to consider the hydrogen atoms as separate centres or to represent the hydrogen bond by a β -integral. Both approaches were used in separate calculations, which is extended also to "anomalous" bases with a configuration slightly different from the standard one.

(a) With separate centres, the α-integrals for the two atoms X, Y, adjacent to H were replaced by $\alpha_X - 0.2$ and $\alpha_Y + 0.2$ while α_H was taken as $-0.60\ \beta$ and $\alpha_{XH} = 0.8\ \beta$; $\alpha_{HY} = 0.04\ \beta$.

(b) Without considering the Hydrogens as separate centres, the α were replaced as in (a) and β_{XY} was taken as 0.2.

All these values are only supported by more or less plausible arguments and may in fact be considered as largely arbitrary.

It is claimed that the energy levels are not very sensitive to the choice of this parameters.

The following table gives the energy levels for Adenine-Thymine calculated in the different ways. All are expressed in unit β_{CC}.

Without considering separate H-centres	With separate H-centres
E = 3.53	E = 3.55
3.11	3.12
2.77	2.78
2.16	2.20
1.80	1.80
1.75	1.77
1.74	1.76
	1.29
1.15	1.15
1.09	1.06
0.90	
0.53	0.62
0.45	0.52
	-0.49
-0.80	-0.72
-0.95	-0.98
-1.04	-1.09
	-1.22
-1.39	-1.41
-1.60	-1.63
-2.01	-2.03
-2.22	-2.22
-2.28	-2.34

With the first set, the levels can be seen to correspond to those of the two separate bases, e.g. the third and fourth levels correspond to the first and second levels of isolated Adenine. When the H are taken into account, there are in all 22 levels and an unambiguous correspondence is not always possible.

The first 12 levels are full. The activation energy is 1.25 β_{CC} with the first set values and 1.01 β_{CC} with the second set.

However the transition is between a level of one base to a level of other base, probably very unlikely. The activation energies within a single basis are about the same as in the previous calculation.

Energy levels of two superimposed bases:

The other configuration of two bases to be considered is when they are superimposed, with their planes parallel. Because of the Helix nature of DNA, it is assumed that they are rotated by 36° one with respect to the other.

Here the β-integrals between an atom i on one ring and an atom j on the other ring are not known. A very rough approximation was then adopted by assuming that the β-integral vary with internuclear distance R in the same way as the overlap integrals S. Both for σ-type and π-type interaction.

$$\frac{\beta_{ij}(R_1)}{\beta_{ij}(R_2)} = \frac{S_{ij}(R_1)}{S_{ij}(R_2)}$$

Then, knowing the value of β_{ij} to be used in the Hückel approximation for a distance R_1, it was possible to find the value for any distance R_2 by calculating the overlap integral S_{ij} as a function of R.

Because of the rotation of the bases about the long axis the 2p, this is just an example, orbitals on the two rings will not have the same axis. If Θ is the angle between the axis and the line connecting the two atoms, one will have

$$S_{ij} = \cos^2\Theta \int \psi_{\sigma i}\, \psi_{\sigma j}\, dV - \sin^2\Theta \int \psi_{\pi i}\, \psi_{\pi j}\, dV + \sin\Theta \cos\Theta \int (\psi_{\pi i}\, \psi_{\pi j} - \psi_{\sigma i}\, \psi_{\pi j})\, dV$$

ψ_σ is the p-orbital directed along the axis, ψ_π is p-orbital directed perpendicularly to the axis. The last integral vanishes because of orthogonality. Similar consideration hold when the two atoms are not Carbons.

About 1600 overlap integrals had to be calculated for the various combination of the four bases, in their standard and anomalous forms.

Having thus estimated the β-integrals between atoms in different rings, the setting up of the secular determinant is a straightforward. The β-integrals between atoms in different rings turn out to be about 1/10 the β-integrals between the same atoms on the same ring.

Rather surprisingly, the energy levels with the two bases superposed turn out to be not very different from those with the two bases adjacent. This is due to the fact that the energy levels are in any case not very different from those of isolated bases. However, the suspicion still remains that these results may be an artefact of the method and of the approximations employed.

The first excitation energy is found generally to be lower, with the bases superposed, than in the isolated molecules.

The oscillator strengths can be calculated and, though for the various reasons stated, one would not place great reliance on the calculation, the interesting result is obtained that these strengths are generally lower than in the isolated bases.

This is in agreement with the experimental fact that the absorbtion coefficient of DNA is only about 60% of the coefficient calculated by superposing the spectra of the single nucleotides.

Band structure of an infinite chain of a single nucleotide:

The number of bases in one chain of DNA is about 10^5, thus it is legitimate to consider the chain as infinite. It is not completely certain whether there is some periodically repeated unit; there probably is, though long and with a complicated structure. This unit is considered as the primitive cell of a one-dimensional crystal. As a first step it was assumed that it consisted of a single molecule of one of the four bases. Although the chain is called one-dimensional, each cell has a number of atoms in a plane perpendicular to the axis. As in the case of two superimposed bases, the helix nature of the chains was taken into account. The calculation was performed again with the simple LCAO-MO methods in the simplest Hückel model.

The determinant is, as has been discussed in chapter 3,

$$\begin{vmatrix} \gamma_{11}-E & \gamma_{12} & \cdots\cdots\cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22}-E & \cdots\cdots\cdots & \gamma_{2n} \\ \cdots & \cdots & \cdots\cdots\cdots & \cdots \\ \gamma_{n1} & \gamma_{n2} & \cdots\cdots\cdots & \gamma_{nn}-E \end{vmatrix} = 0$$

It is of the same type as for an isolated bases, except that the elements γ_{ij} are now functions of k, and thus the energy levels are given as functions of k and form energy bands in the usual way, one band for each row of the determinant, that is one band for each atom in the primitive cell, for instance 10 for poly-adenine.

In order to carry out the calculation in the simplest way, a drastic approximation was introduced here; the same α and β parameters were used as for single bases. In order to calculate the $\beta^{(\text{right})}$ and $\beta^{(\text{left})}$ it was again assumed that these energy integrals vary with distance between the two atoms in the same way as the corresponding overlap integrals, and the values of the overlap integrals calculated for two superposed bases were used.

The "valence" and "conduction" bands are highest filled band and the lowest empty band, deriving from the highest filled level and the lowest empty level of the isolated basis. In the case of Adenine these were at 0.53 and at -0.79 in units of β_{cc} which is a negative quantity.

Still for Adenine, it was found that the highest filled level at E = 0.53 spreads into a valence band extending from E = 0.58 for $k=\frac{\pi}{a}$ to E = 0.48 for k = 0, while the lowest empty level at E = -0.79 spreads into a conduction band extending from E = -0.75 for $k = \frac{\pi}{a}$ to E = -0.82 for k = 0.

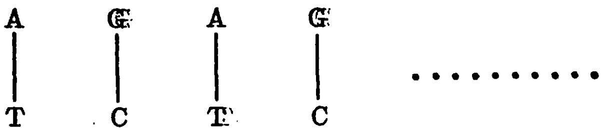
Incidently, it appears somewhat strange that two consecutive bands should have energy maxima and minima at the same point.

For other bands it was found that the edge is not necessarily at k = 0 or at $k = \frac{\pi}{a}$, but may be at an intermediate value.

The energy gap is thus indirect and of magnitude 0.48 + 0.75 = 1.23, a little less than the first excitation energy of the isolated basis; this is to be expected in view of the broadening of the levels.

Similar results were obtained for the other poly-bases.

Again one wonders how much these results are due to the type of approximation used. As for the width of the bands, this must be, within these approximations, of the same order as the β integrals between different bases. It was found when dealing with two superposed bases that these integrals could be estimated as about 0.1. Bands of this width could then have been anticipated, and the more quantitative results yielded by the calculation are open to much doubt.
The same considerations may explain the fact that the width of the higher bands is not larger than that of the lower ones, as is the general case with ordinary metals and semiconductors. Here all widths are about the same as the β integrals, and there are so many bands of comparable width because there are so many atoms per primitive cell. This was as far as the calculations had progressed up to 1964.
However the results for superposed bases, and the broadening calculated for poly-structures allowed some conjectures on the band structure of the real DNA, for instance in the model below, (only two base pairs).



or more complicated periodic models.

For the above model one would use the results obtained for the pairs A —— T and G —— C and the result that the average band width is about 0.1 β . One could then determine which level of the pairs broaden into bands, which overlap and form a wider band. These results were re-obtained later, in the next paper to be reviewed. It is not necessary to stress the highly speculative nature of the results obtained. In particular, from our present knowledge, it seems that there are no periodic sequences of the basis in DNA.

For what it is worth, the model indicates that DNA is an insulator with an energy gap of 2.7 eV to 3.5 eV. The lower value is obtained by considering separate H centres at the hydrogen bond joining the two bases of each pair, the higher value is obtained by not including the H-atoms as separate centres.

Experimental results by Eley and Spivey (1962) on the variation on conductivity with temperature had indicated an activation energy of 2.4 eV.

These are the results reported in the review article by Hoffman and Ladik (1963).

Another calculation in the Hückel approximation was published by Ladik and Bizco in 1964.

They consider infinite chains of different periodic models.

The models considered were:

poly(A-T); poly(G-C); poly $\begin{bmatrix}\text{A-T}\\\text{G-C}\end{bmatrix}$ or poly $\begin{bmatrix}\text{A-T}\\\text{C-G}\end{bmatrix}$;

poly $\begin{bmatrix}\text{A-T}\\\text{T-A}\end{bmatrix}$; poly $\begin{bmatrix}\text{G-C}\\\text{C-G}\end{bmatrix}$

The extension of the theory required to pass from a single system of two adjacent bases A-T to an infinite chain poly (A-T) is the same as explained in extending the theory from that of a single basis to that of an infinite chain of a single basis.

Thus the secular equation for poly (A-T) is obtained from that for a single (A-T) by replacing the α - and β- parameters by k-dependent combinations such as

$$\gamma_{ij} = \beta_{ij} + \beta_{ij}^{(right)} e^{ika} + \beta_{ij}^{(left)} e^{-ika}$$

It remains a 20 x 20 secular equations, but with complex elements.

Without considering the H-atoms of the hydrogen bonds as separate centres, the results for poly

A-T are, for the first two empty and the first two full bands:

E (poly A-T)	E(single A-T)
0.49 at $k = 0$ to 0.58 at $k = \frac{\pi}{a}$	0.53
0.41 at $k = 0$ to 0.49 at $k = \frac{\pi}{a}$	0.45
-0.84 at $k = 0$ to -0.77 at $k = \pi$	-0.80
-0.94 at $k = \frac{\pi}{a}$ to 0.93 at $k = 0$	-0.94

For poly G-C, similar results were obtained.

In the other four cases the complex matrices are of order 40 but all the principles and approximations on which the calculation is based remain the same. In the case of homo-chains the band extrema were always at $k = 0$ or $k = \frac{\pi}{a}$; in the more complicated structures the extrema were generally found at intermediate values of k, probably due to some interaction between bands. In these calculations all bands are "of the same symmetry" and no crossing of bands is possible.

Table 1 gives the limits of highest filled and lowest empty band, their widths and the width of the forbidden gap for various periodic models of DNA. The values of k at which the extrema occur are also indicated.

ENERGY-BAND CALCULATIONS FOR DNA MODELS

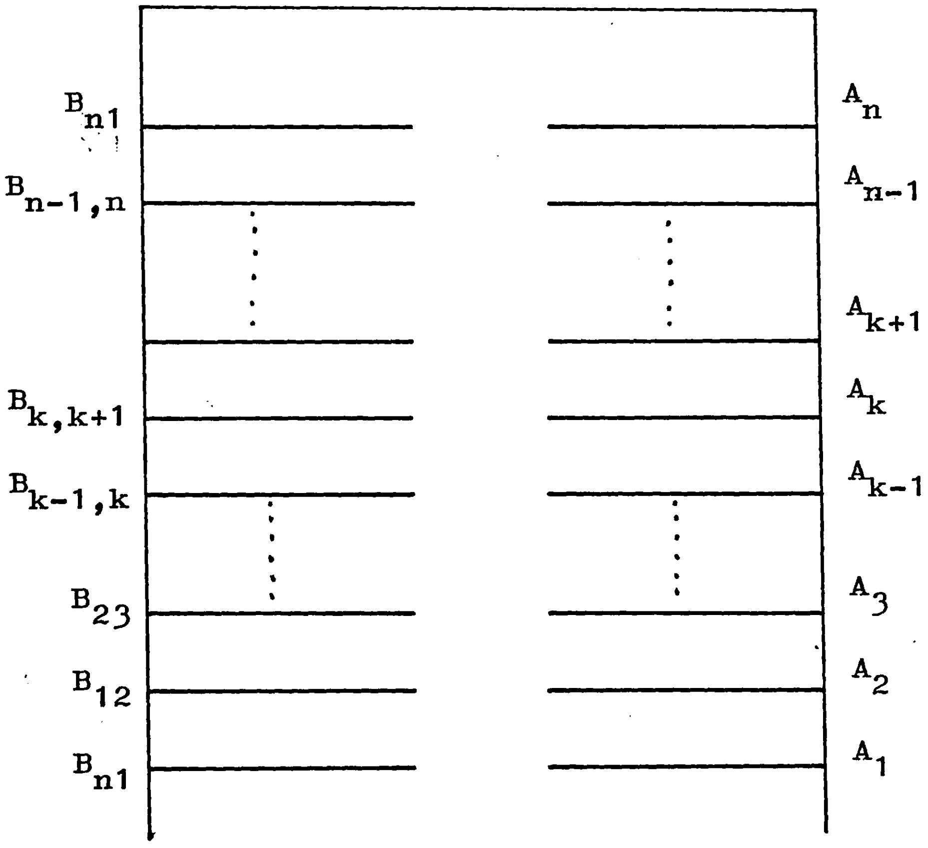
Table 1. The limits of the highest filled and lowest unfilled energy bands, their widths and the forbidden bandwidths (in β units) of different periodic models of DNA.

	$E(k)_{max}$	$E(k)_{min}$	k_{max}	k_{min}	ΔE	ΔE_G
poly T	0.464 -1.036	0.547 -1.014	0 0	π π	0.083 0.022	1.478
poly C	0.537 -0.867	0.631 -0.824	0 0	π π	0.094 0.043	1.361
poly A	0.480 -0.823	0.576 -0.749	0 0	π π	0.096 0.074	1.229
poly G	0.336 -1.122	0.433 -1.058	0 0	π π	0.097 0.064	1.394
poly (A-T)	0.411 -0.841	0.488 -0.766	0 0	π π	0.077 0.075	1.177
poly (G-C)	0.310 -0.854	0.385 -0.811	0 0	0.80π π	0.075 0.043	1.221
poly $\begin{bmatrix}T\\C\end{bmatrix}$	0.486 -0.846	0.503 -0.844	0	π 0	0.017 0.002	1.330
poly $\begin{bmatrix}A\\C\end{bmatrix}$	0.514 -0.786	0.522 -0.786	0 0	π π	0 0	1.300
poly $\begin{bmatrix}C\\G\end{bmatrix}$	0.386 -0.843	0.387 -0.843	0 0	π π	0.001 0	1.229
poly $\begin{bmatrix}T\\A\end{bmatrix}$	0.484 -0.786	0.493 -0.786	0 0	π π	0.009 0	1.270
poly $\begin{bmatrix}G\\A\end{bmatrix}$	0.372 -0.786	0.385 -0.783	0	π 0	0.013 0.003	1.155
poly $\begin{bmatrix}G\\T\end{bmatrix}$	0.374 -1.024	0.380 -1.024	0 0	π π	0.006 0	1.098
poly $\begin{bmatrix}A-T\\G-C\end{bmatrix}$	0.346 -0.815	0.351 -0.691	0 0.87π	0.70π 0.20π	0.005 0.124	1.037
poly $\begin{bmatrix}A-T\\C-G\end{bmatrix}$	0.342 -0.860	0.356 -0.699	0 0.43π	π 0.26π	0.014 0.161	1.041
poly $\begin{bmatrix}A-T\\T-A\end{bmatrix}$	0.414 -0.805	0.454 -0.708	0.13π 0.09π	0.57π 0.78π	0.040 0.097	1.122
poly $\begin{bmatrix}G-C\\C-G\end{bmatrix}$	0.332 -0.804	0.351 -0.748	0.15π 0.88π	0.34π 0.75π	0.019 0.056	1.080

The forbidden gap is not very different in various models. Some of the lower values are due to the fact that some bands are "combinations" arising from a combination of levels belonging to different bases.

The lowest value of E_G is 1.037β which is about 3.45eV. Experimental value for DNA from d.c. conductivity measurements are lower, it is about 2.4eV. However the simple Hückel approximation employed in the calculation does not take into account the existence of different multiplets; the calculated bands are "singlet" bands. It may be that the experimental value refers to a transition between, for example, a singlet and a triplet band, although these transitions are forbidden in first approximation. Clearly more accurate calculations will have to take into account the possibility of different configurations (Configuration Interaction) and this has been attempted by Ladik and co-workers in their 1968-1971 papers. Before examining these, another attempt carried out in the simple Hückel approximation will be examined.

Note that so far not much has been said about the fundamental problem of whether the interaction between base pairs in different planes is strong enough so that the π electrons may be considered as delocalized over the whole chain. Often one assumes that the different planes are held together by van der Waals forces and then there would not be much overlap between the π-wave functions in different planes. Still the calculated width of the bands is not negligible, and this indicates that there must be some overlap. The last calculation with the simple Hückel method was carried out by Belezney and Biczo (1964), The last step in Ladik's programme was to work out the energy levels of a periodic chain of A-T and G-C base pairs. This was carried out by Belezney and Biczo by means of an ad hoc method of averaging.



Let A_k be the matrix corresponding to the secular determinant for the k^{th} subunit.

$$A_k = \begin{pmatrix} a_{11}^{(k)} - E & a_{12}^{(k)} & \cdots\cdots\cdots\cdots & a_{1n}^{(k)} \\ a_{21}^{(k)} & a_{22}^{(k)} - E & \cdots\cdots\cdots & a_{2n}^{(k)} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots\cdots\cdots & \cdots\cdots \\ a_{n1}^{(k)} & a_{n2}^{(k)} & & a_{nn}^{(k)} - E \end{pmatrix}$$

where the a_{ij} represent the usual Hückel parameters α and β. Then only the interaction between nearest neighbour subunits is taken into account using matrices $B_{k,k+1}$ which contain the β-integrals between atoms belonging to neighbouring subunits.

$$B_{k,k+1} = \begin{pmatrix} b_{11}^{(k,k+1)} & \cdots\cdots\cdots\cdots\cdots\cdots & b_{1n}^{(k,k+1)} \\ & & \\ & & \\ b_{n1}^{(k,k+1)} & \cdots\cdots\cdots\cdots\cdots\cdots & b_{nn}^{(k,k+1)} \end{pmatrix}$$

Then the complete Secular Equation for the whole chain is:

$$\begin{bmatrix}
A_1 & B_{12} & 0 \cdots & 0 & 0 & 0 \cdots & 0 & B_{1N} \\
B_{21} & A_2 & B_{23} \cdots & 0 & 0 & 0 \cdots & 0 & 0 \\
0 & B_{32} & A_3 \cdots & 0 & 0 & 0 \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & A_{k-1} & B_{k-1,k} & 0 & 0 & 0 \\
0 & 0 & 0 & B_{k,k-1} & A_k & B_{k,k+1} & 0 & 0 \\
0 & 0 & 0 & 0 & B_{k+1,k} & A_{k+1} & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & B_{N-1} & B_{N-1,N} \\
B_{N1} & 0 & 0 & 0 & 0 & 0 & B_{N,N-1} & A_N
\end{bmatrix} = 0$$

where $B_{21} = B_{12}^{+}$

The number r of different subunits depends on the macromolecule; in the present model $r = 4$ (A-T, T-A, G-C, C-G). Let p_r be the probability of appearance of subunit r, then one can define an average:

$$\bar{A} = \sum_{i=1}^{r} p_i A(i)$$

and similarly if q_{ij} is the probability that units i,j are nearest neighbours one can define an average

$$\bar{B} = \sum_{i,j=1}^{r} q_{ij} \, B(ij)$$

$B(ij)$ is the same as B_{ij}

Then each matrix $(A_k) = (\bar{A}) + (a_k)$ where a_k is a deviation from the average, which is assumed to be a lot smaller than $(\bar{A})$. However when the determinant is developed it is clear that linear terms in the deviations will cancel out, thus, neglecting second order terms, we can replace the true secular equation by

$$\begin{bmatrix} \bar{A} & \bar{B} & 0 & \cdots\cdots & 0 & \bar{B}^{+} \\ \bar{B}^{+} & \bar{A} & B & \cdots\cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots\cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots\cdots \bar{B}^{+} & \bar{A} & \bar{B} \\ B & 0 & 0 & \cdots\cdots & \bar{B}^{+} & \bar{A} \end{bmatrix} = 0$$

So the model macromolecule has been replaced by an average homo-macromolecule, with all the subunits equal. The problem is then the same as that of poly-adenine, etc. previously treated by Ladik, except that now the periodic subunit is a very nebulous average of the true subunits. The average depends on the choice of the probability p, q. Initially, presumably all p were taken as $\frac{1}{4}$ (four subunits; G–C, C–G, A–T, T–A) and all q were taken as 1/6.

The error due to averaging may be reduced by a suitable choice of the set of basis functions used to calculate the matrices A_k. But in general one "average level" will correspond to two true levels. The results obtained for poly $\begin{bmatrix} \text{A-T} \\ \text{G-C} \end{bmatrix}$ were

		Ladik results
First empty band	$\left[\begin{matrix}-0.822\\-0.813\end{matrix}\right.$	$\left[\begin{matrix}-0.815\\-0.709\end{matrix}\right.$
First filled band	$\left[\begin{matrix}0.367\\0.440\end{matrix}\right.$	$\left[\begin{matrix}0.346\\0.350\end{matrix}\right.$
Energy Gap	1.18	1.055

They are compared with those of Ladik which are called "exact" by Belezney and Biczo while they are only a rough approximation.

It is worth mentioning that one method shows a wide valence band and a narrow conduction band, the other method yields the opposite results. The fact that the positions of the bands and the energy gap are similar with the two methods seems to be just a consequence of the position of levels in the isolated bases (different ones), a result much too simplicistic.

Calculations were repeated with different values of the probabilities p and q, corresponding to different (A-T)/(G-G) ratios found in different types of natural DNA. The results again do not mean much and could have been anticipated by looking at the genesis of levels.

The first application of the PPP (Pariser-Parr-Pople) approach to DNA-like structures was made by Ladik et. al. in 1968. They recalculated the band structure of the five polyhomonucleotides previously worked out by the simple Hückel method. The ionisation potentials and electron affinity values were taken from tables. For the γ-integrals a simplified expression was used while for the β integrals the same values were taken as in the previous calculations.

The results were rather shattering: the bands obtained in the present approximation are vastly different from those obtained in the Hückel approximation.

A comparison of results is made in the next section.

4.2 Comparison of results.

The tables below show the results of the two methods which have been discussed previously, namely the simple Hückel methods and the Pariser-Parr-Pople method.

Table 1

Results obtained using the Hückel Approximation for Adenine

	Level of isolated molecule in eV	corresponding band in eV
Full Bands	-9.22	-8.63 to -9.87
	-7.09	-6.04 to -7.54
	-5.83	-5.68 to -6.01
	-3.83	-3.81 to -3.89
	-3.43	-3.33 to -3.42
	-1.76	-1.60 to -1.92
	2.65	2.74 to 2.49
	3.33	3.41 to 3.28
	4.63	4.68 to 4.55
Empty Bands	6.73	6.74 to 6.69

Hence the energy gap according to this method is $E_G = 4.41$ eV for the isolated molecule. Similar results were obtained for the other bases.

Table 2

Results obtained using the PPP approximation

	Level of isolated molecules in eV	Corresponding band in eV
Full Bands	-16.73	-16.36 to -16.66
	-15.58	-15.18 to -15.64
	-13.48	-12.82 to -13.44
	-12.33	-11.84 to -12.08
	-10.68	-10.60 to -10.60
	- 9.85	- 9.18 to - 9.49
	- 1.86	- 1.41 to - 1.51
	- 1.31	- 0.67 to - 0.75
	+ 0.02	0.34 to 0.33
Empty Bands	2.01	2.54 to 2.48

We may note that the difference in energy between the Hückel and PPP molecular levels is (7.93 ± 0.48) eV for the six filled bands and (4.60 ± 0.08) eV for the four empty bands.

Clearly one of the methods is inadequate for dealing with the excited states. Within the occupied, or unoccupied levels, the results of the two methods are consistent, though there must have been a difference in fixing the zero of energy, corresponding to the different results for the filled and empty levels. The energy gap which was 4.09 eV with the Hückel method, has become 7.67 eV with the so called self-consistent procedure. The relation between this gap between the bands and the true energy gap is discussed below following such a large energy gap seems unrealistic for this type of structure. In the first paper the authors only point out in a foot note that the forbidden band widths obtained by the PPP method have nothing to do with the first singlet excitation energies of these system. It is then rather puzzling that they still carried out the calculation.

The width of the bands is generally smaller with the new method.

In the second paper, in conjunction with Avery and Packer of Imperial College, Ladik and Biczo worked out poly (A-T) and poly (G-C).

The method and the approximations were the same as for the single-base chains, but, as with the

simpler method, the computational problem was more serious. The β-integrals between atoms in the two different bases were given the same values as in the calculation with the Hückel method. Five iterations were necessary to produce self-consistency to within 10^{-3} in the charge densities and bond orders.
The following tables, constructed as the table on the previous page, gives results for poly (A-T), in eV, for the two highest empty bands and the two lowest empty bands.

Table 3
Results obtained using the Hückel approximation for poly (A-T).

Level of isolated molecule in eV	Corresponding band in eV
-1.76	-1.93 to -1.63
-1.50	-1.63 to -1.37
2.66	2.56 to 2.80
3.13	3.10 3.13

Table 4

Results obtained using the PPP approximation for poly (A-T).

Level of isolated molecule in eV	Corresponding band in eV
-10.86	-10.69 to -10.50
- 9.71	- 9.62 to - 9.32*
- 2.83	- 2.84 to - 2.77
- 1.74	- 1.57 to - 1.47

*The E-k relation (one-dimensional) was calculated in full for this band.

A similar wide disagreement occurs for poly (G-C). The same things as in the case of single-base chains happen here, in particular the width of the forbidden gap is very large, i.e. 6.5 eV.

The relation of this value to the true energy gap is discussed more fully than in the previous paper. It is pointed out that for an isolated molecule the singlet excitation energy between levels i and j is given by

$$1 \Delta_{i \rightarrow j} = \epsilon_j - \epsilon_i - J_{ij} + 2K_{ij}$$

where

${}_j$ and ${}_i$ are the energies of molecular orbital j and i respectively.

J_{ij} is the Coulomb Integral between Molecular Orbitals i, j.

K_{ij} is the Exchange Integrals between Molecular Orbitals i, j.

However, it may be shown that for an N-polymer the integrals J and K approach Zero as N goes to infinity if the electrons are delocalized. This is described exactly by Bloch-type wavefunctions. Thus there cannot be complete delocalisation, otherwise the energy gaps are all wrong.

The authors think that the difference between the true E_G and the calculated separation of the bands may be due to a correlation between the motion of the electron and that of the hole which it leaves in the valence band. There would be the formation of an exciton. The authors say that they have started calculations of this effect, but they do not seem to have been published.

The next paper (1970) continues applying the method to more complicated structures, that is poly (A-T;G-C); poly (A-T;C-G); poly (A-T;T-A) and poly (G-C;C-G). The maximum and minimum energies of 30 bands are given.

For poly $\begin{bmatrix}A-T\\G-C\end{bmatrix}$ results in eV for the "valence" and "conduction" bands, are compared with those obtained by the simple method.

	Hückel Method		PPP Method
	Ladik	Beleznev-Biczo	
Highest	-1.17	-1.47	-8.76
filled band	-1.15	-1.22	-8.73
Lowest	2.36	2.71	-2.660
empty band	2.71	2.74	-2.657
	E_G = 3.51	E_G = 3.93	E_G = 6.07

The same trends are observable as for the less complicated structures. The only new feature is that bands now come out to be very narrow. Even the width of the bands of $\begin{bmatrix}A-T\\T-A\end{bmatrix}$ are much smaller than those of the bands of (A-T) only. This therefore placed the hypothesis of non-localization of the electrons in a serious doubt.

In these more complicated structures electrons seem to stick to their original molecule, though, again, this may well be an artefact of the method of calculation. The conduction mechanism would be by hopping. Here, again the calculated "energy gap" is not related in a simple way to the experimental singlet excitation energy.

All the results obtained up to 1970 are discussed in Ladik's review paper in the International Journal of Quantum Chemistry (1971).

Here, he makes the point that the small width of the bands of the more complicated structures can be understood from the alteration of the sub-units which prevents strong interactions between similar levels.

It is worth noting that for these narrow bands, a calculation of the conductivity on deformation potential theory leads to results very far from the experimental ones.

It is mentioned that water impurities have a small effect on the band structure. On the other hand Mg^{2+} ions produce drastic changes.

Ladik's final conclusion is that:

It seems probable that in DNA the electrons are strongly de-localised only in such segments of the macromolecule in which either the same bases are repeated or there are change impurities. In other parts of the molecule the electrons should be described by the hopping models.

This conclusion puts in question the validity of the periodic models used in the calculations.

There have been attempts at calculating the charge carrier mobilities in the periodic structure considered, based on the calculated band structure, even though the mechanism of conduction in DNA does not seem to have been ascertained.

The first attempt (1965) by Ladik et. al. is much too simplified. The conductivity obtained is much too high to be realistic. It is of the order of 10^3 to 10^5 $\Omega^{-1}\text{cm}^{-1}$

The later effort by Suhai (1972) is more serious, applying the variational methods of transport theory for scattering by longitudinal lattice vibrations. This gave conductivities in the range of 30 to 100 $\Omega^{-1}\text{cm}^{-1}$ for band widths of range from 0.1 to 0.3 eV and 4 to 20 for band widths 0.02 to 0.1 eV. The mean free path was large enough to justify the hypothesis of delocalized electrons, particularly for the larger bandwidths.

References:

1. Hoffmann, T.A. and Ladik, J., Cancer Research 21, 474-484 (1961)

2. Ladik, J. and Appel, K., The Journal of Chemical Physics, 40, 2470-2473 (1964)

3. Beleznay, F., and Biczo, G., The Journal of Chemical Physics, 41, 2351-2355, (1964)

4. Ladik, J. and Biczo G., The Journal of Chemical Physics, 42, (1965)

5. Ladik, J., Biczo, G., and Elek, G., The Journal of Chemical Physics, 44, 483-485, (1966)

6. Ladik, J., Rai, D.K. and Appel, K., Journal of Molecular Spectroscopy, 27, 72-78 (1968)

7. Avery, J., Packer, J., Ladik, J., and Biczo, G., Journal of Molecular Spectroscopy, 29, 194-203, (1969)

8. Rozsnyai, B.F., Martino, F., and Ladik, J., The Journal of Chemical Physics, 52, 5708-5717, (1970)

9. Ladik, J., International Journal of Quantum Chemistry, 4, 307-317 (1971)

10. Sandor Suhai, The Journal of Chemical Physics, 57, 5599-5603, (1972)

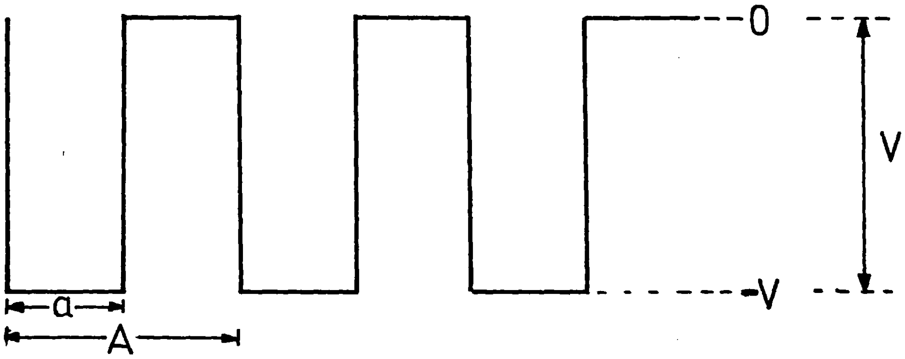
Chapter 5

5.1 Purpose of the Simple Model

From Ladik's work, one conclusion can be drawn and that is that a heteronucleotide model has narrower bands than a homonucleotide.

However, a very simple and exact method can be employed to arrive to that conclusion as will be described in this chapter.

Consider an infinite chain of square potential wells with period A. The depth of the well is -V.



The solution of the Schrödinger Wave equation in the well is

$$= A \sin \gamma x + B \cos \gamma x \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

$$= A \gamma \cos \gamma x - B \gamma \sin \gamma x \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

where A, B are parameters to be determined

$$\gamma^2 = \frac{2m}{\hbar^2}(E+V)$$

where m is mass of the electron

$\hbar$ is Planck Constant

E is energy of the state considered

V is the depth of the wells

Eliminating A and B to express $\psi(a)$ in terms of $\psi(0)$.

$$\begin{pmatrix} \psi(a) \\ \frac{d\psi(a)}{dx} \end{pmatrix} = \begin{pmatrix} \cos\gamma a & \frac{1}{\gamma}\sin\gamma a \\ -\gamma\sin\gamma a & \cos\gamma a \end{pmatrix} \begin{pmatrix} \psi(0) \\ \frac{d\psi(0)}{dx} \end{pmatrix}$$

The matrix producing ψ and $\frac{d\psi}{dx}$ at one end of the well from their values at the other end may be called a transfer matrix. Similar formulae hold in a potential hill, V=0, except that γ may be imaginary if E is negative. In this case $\cos\gamma a$ is replaced by $\cos|\gamma|a$ and $\sin\gamma a$ is replaced by $i\sinh|\gamma|a$. If in a period A of the chain one has a number of different potential wells, the values of ψ and $\frac{d\psi}{dx}$ at one end of the period may be obtained from their values at the other end by application of a matrix b_{ij} which is the product of all the transfer matrices of the various segments. But by the Bloch condition, one has also

$$\begin{vmatrix} \psi(A) \\ \dfrac{d\psi(A)}{dx} \end{vmatrix} = e^{ikA} \begin{vmatrix} \psi(0) \\ \dfrac{d\psi}{dx}(0) \end{vmatrix}$$

so that

$$\begin{vmatrix} \psi(A) \\ \dfrac{d\psi(A)}{dx} \end{vmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{vmatrix} \psi(0) \\ \dfrac{d\psi(0)}{dx} \end{vmatrix}$$

$$= e^{ikA} \begin{vmatrix} \psi(0) \\ \dfrac{d\psi(0)}{dx} \end{vmatrix}$$

It should be noted that the matrix b_{ij} is unitary.

The determinantal equation is then

$$\begin{vmatrix} b_{11} - e^{ikA} & b_{12} \\ b_{21} & b_{22} - e^{ika} \end{vmatrix} = 0$$

$$e^{2ikA} - (b_{11} + b_{22})e^{ikA} + b_{11}b_{22} - b_{12}b_{21} = 0,$$

or, because b_{ij} is unitary, $e^{2ikA} - (b_{11} + b_{22})e^{ikA} + 1 = 0$

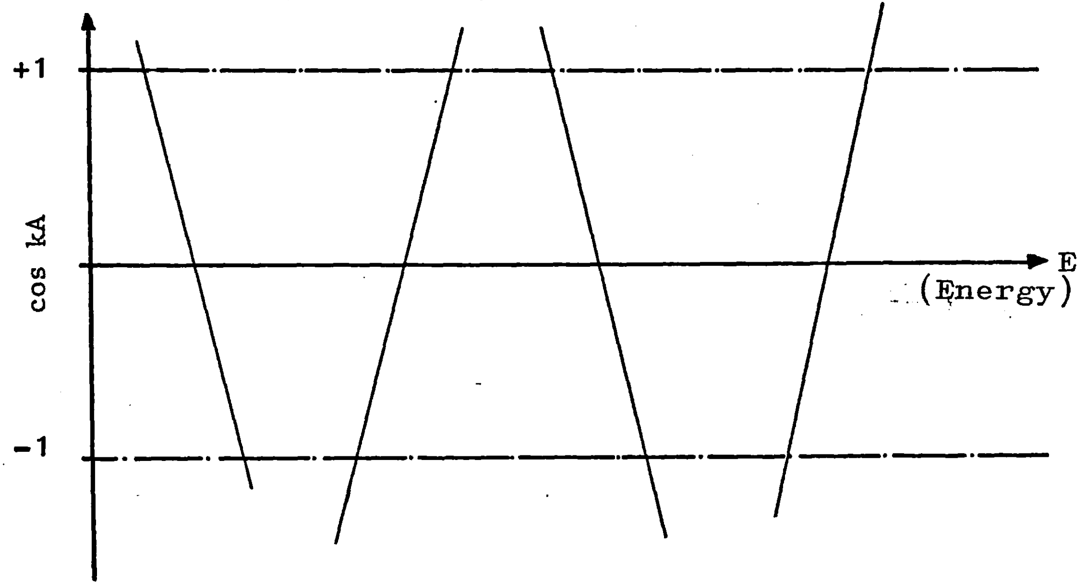
$$e^{ikA} = \tfrac{1}{2}(b_{11} + b_{22}) \pm \sqrt{\tfrac{1}{4}(b_{11} + b_{22})^2 - 1}$$

$$= \tfrac{1}{2}(b_{11} + b_{22}) \pm i\sqrt{1 - \left[\tfrac{1}{2}(b_{11} + b_{12})\right]^2}$$

$$= \cos kA + i \sin kA$$

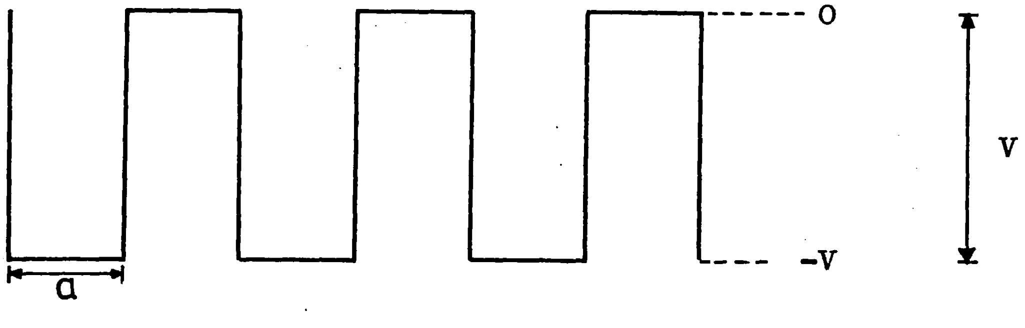
Therefore $\cos kA = \frac{1}{2}(b_{11} + b_{22})$

If $\cos kA$ is plotted against E(energy) one gets a group of the type illustrated in the figure below, in which the allowed energy bands are the ranges of E in which $|\cos kA| \leq 1$

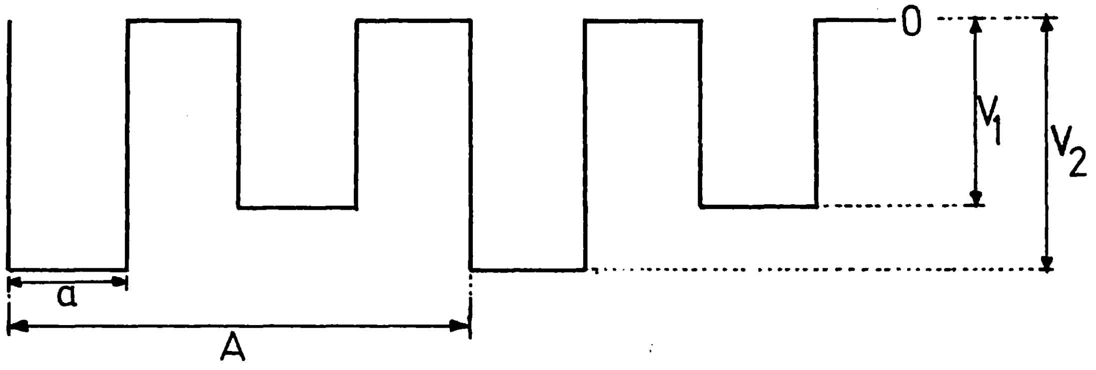


Assuming that a well with a certain potential V represents a base, either A, T, G or C.

For a homonucleotide chain, the model has the form



For a heteronucleide chain, the model is



When values for A, V_1 and V_2 are chosen

A computer programme was written to do the calculation and the results are plotted on the following figures.

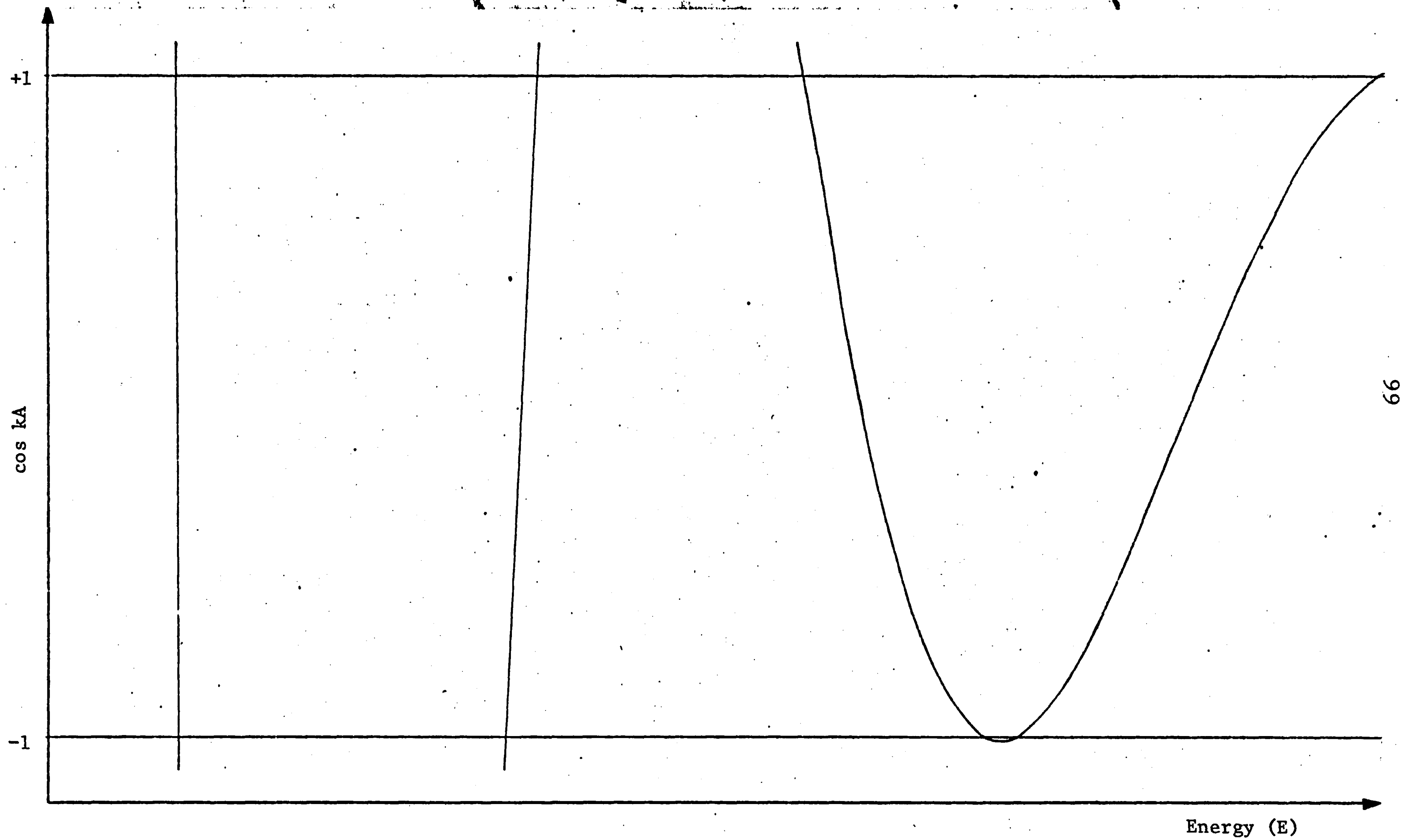


DIAGRAM 1

Energy diagram for homonucleotide chain, with a = 3.6 Bohr Radii and potential V = -1 (27.2 eV). Unit 2 Rydberg. The band edges and the band width are listed in table 1, page 104.

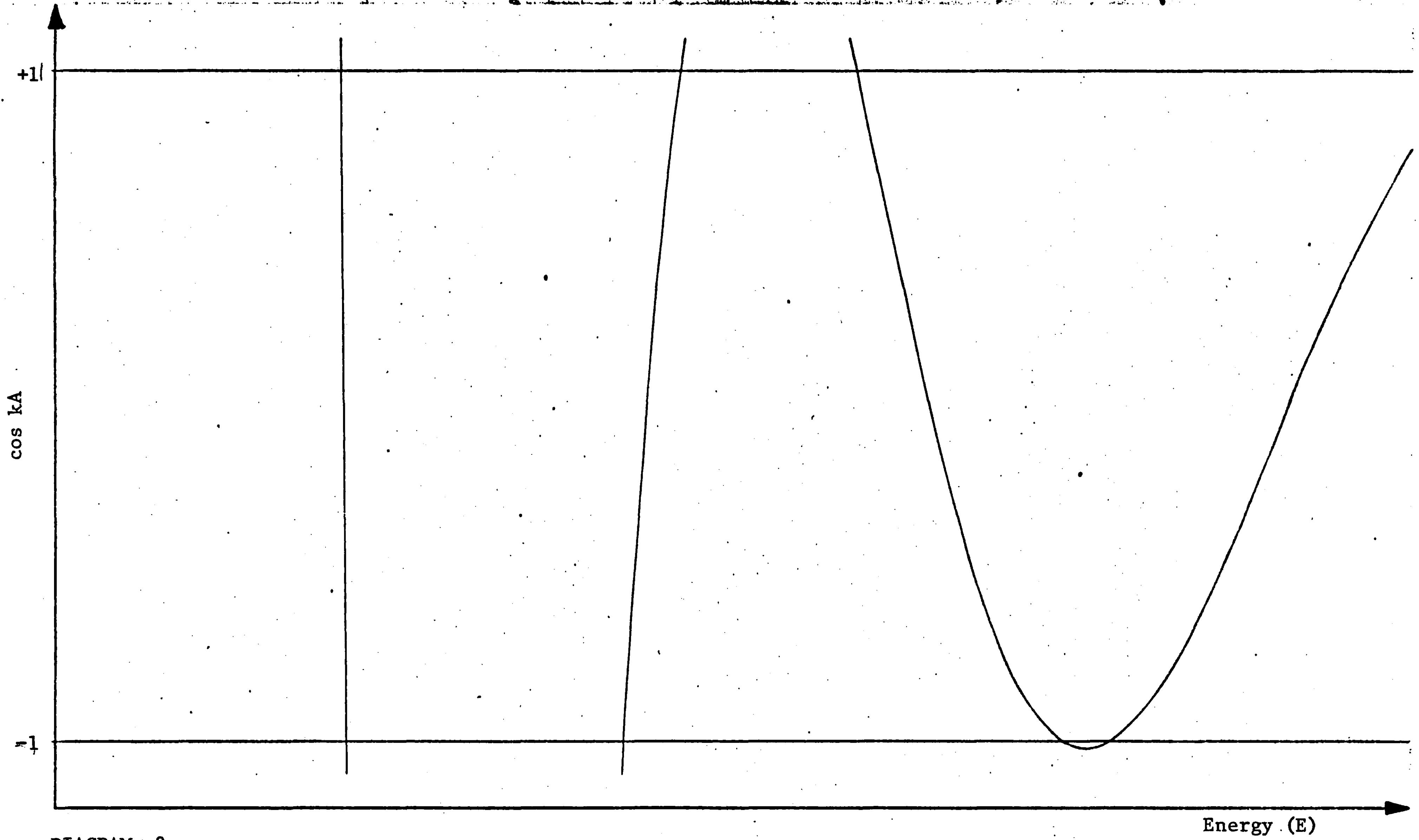


DIAGRAM 2

Energy diagram for homomucleotide chain, with a = 3.6 Bohr Radii and potential V = -0.75. Unit 2 Rydberg. The band edges and the band width are listed in table 1, page 104.

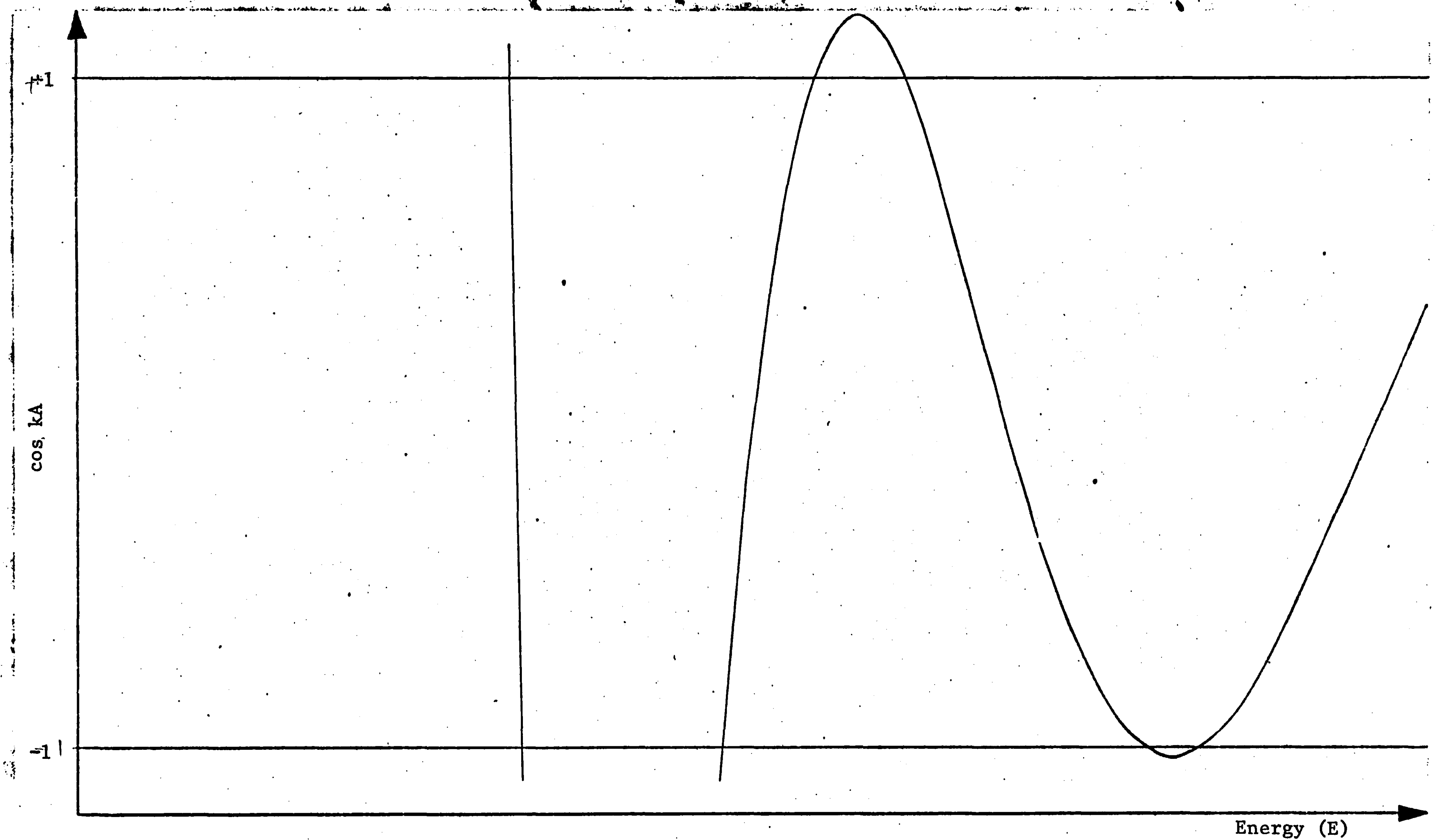


DIAGRAM 3

Energy diagram for homonucleotide chain, with a = 3.6 Bohr Radii and potential V = -0.50. Unit 2 Rydberg. The results are listed in table 3.

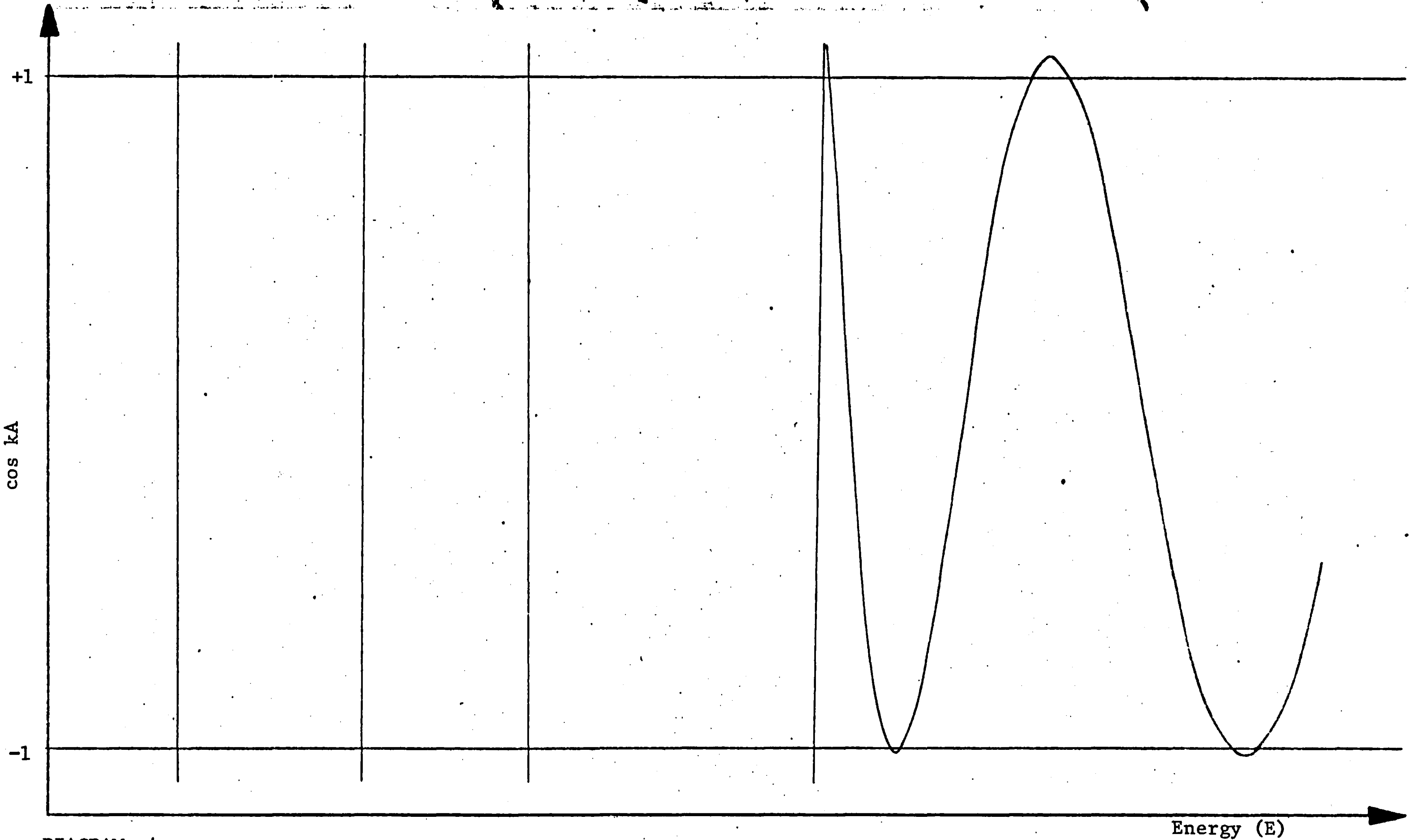


DIAGRAM 4

Energy diagram of heteronucleotide chain, with a = 3.6 Bohr Radii and potential $V_1 = -0.75$, $V_2 = -1$. Unit 2 Rydberg. The results are listed in table 4.

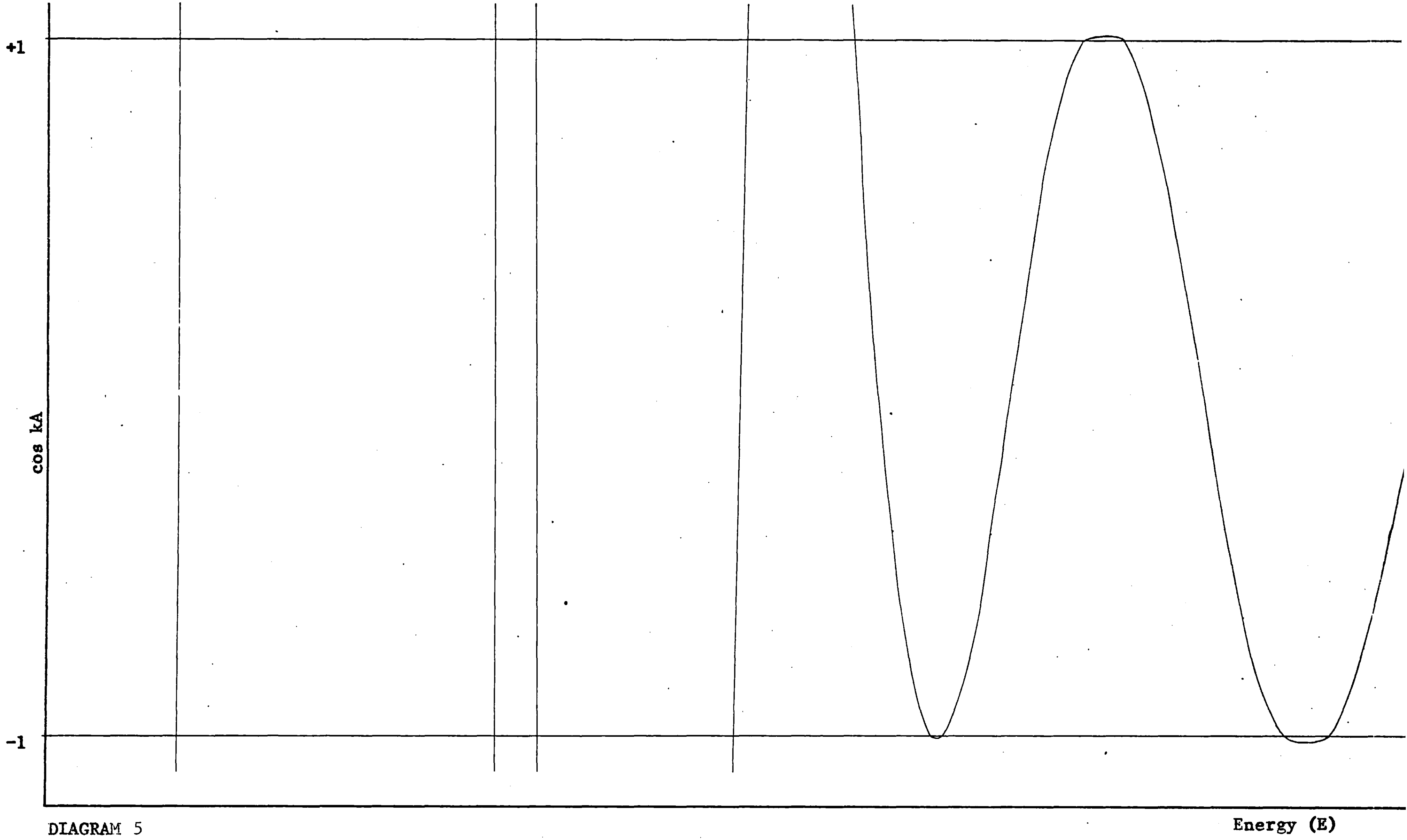


DIAGRAM 5

Energy diagram of heteronucleotide chain, with a = 3.6, Bohr Radii and potential, a mixture of V = -0.5 and V = -1.0 Unit 2 Rydberg. The results are listed in table 4.

5.2 Analogy with Ladik's results

1. A homonucleotide model with potential V=-1 27.2eV and where a=3.6 Bohr radii. The programme produces results as shown in diagram 1. The table below lists the band edges and band widths obtained from the diagram.

Table 1

Band 1	(A1)	E=-0.8054	to	E=-0.8095	ΔE=-0.0041
Band 2	(A2)	E=-0.313	to	E=-0.269	ΔE= 0.044
Band 3	(A3)	E= 0.133	to	E= 0.410	ΔE= 0.277
Band 4	(A4)	E= 0.453	to	E= 0.995	ΔE= 0.542

2. Another homonucleotide model with potential V=-0.75 and a=3.6 Bohr R. The results are shown in diagram 2, and the table below lists the band edges and band widths.

Table 2

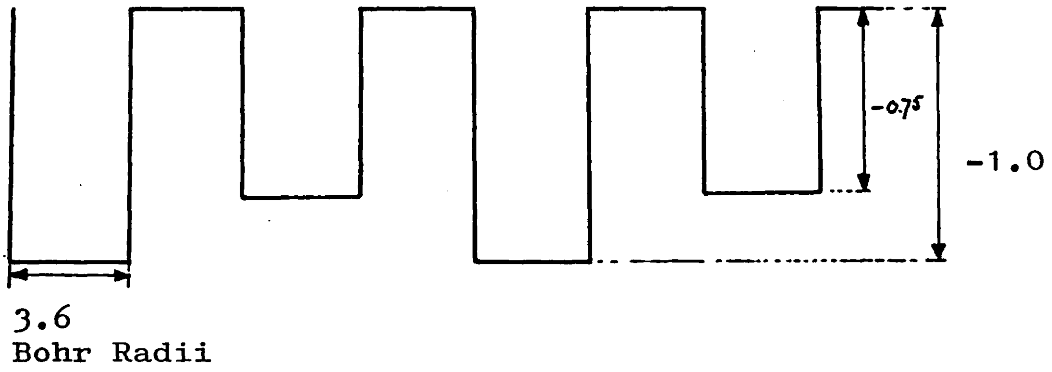
Band 1	(B1)	E=-0.5783	to	E=-0.5707	ΔE= 0.0076
Band 2	(B2)	E=-0.162	to	E=-0.080	ΔE= 0.082
Band 3	(B3)	E= 0.178	to	E= 0.485	ΔE= 0.307
Band 4	(B4)	E= 0.555	to	E not included in the graph.	

3. A homonucleotide model with potential $V=-0.50$ and $a=3.6$ Bohr Radii.

Table 3

Band 1 (C1)	$E=-0.3600$ to $E=-0.3420$	$\Delta E= 0.0180$
Band 2 (C2)	$E=-0.0450$ to $E= 0.0890$	$\Delta E= 0.1340$
Band 3 (C3)	$E= 0.2250$ to $E= 0.5850$	$\Delta E= 0.3600$
Band 4 (C4)	$E= 0.6650$ to not in the graph.	

4. A heteronucleotide which is a combination of models 1 and 2 in the following fashion.

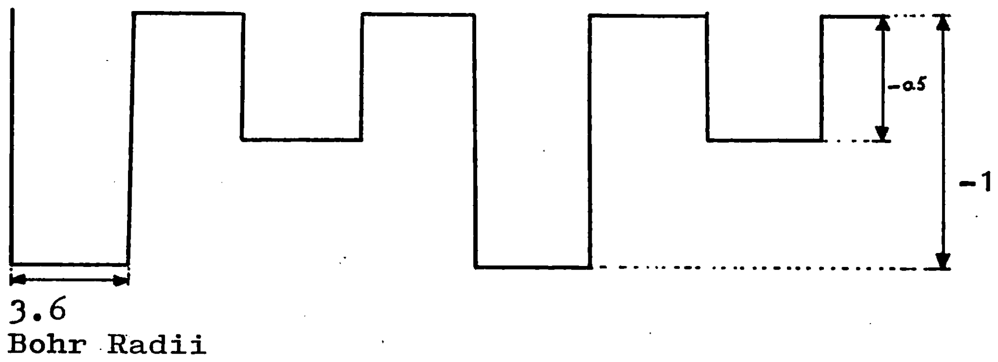


The results are shown in diagram 4 and the table overleaf lists the band edges and the band widths. The bands are also correlated to those (A1 A2... B1...) of the previous tables.

Table 4

Band 1	(A1)	E=-0.8084	to	E=-0.8084	ΔE= 0.0001	
Band 2	(B1)	E=-0.5358	to	E=-0.5345	ΔE= 0.0013	
Band 3	(A2)	E=-0.2939	to	E=-0.2909	ΔE= 0.0030	
Band 4	(A3)(B2)	E= 0.1313	to	E= 0.1425	ΔE= 0.0112	Hybridized
Band 5	(B2)(A3)	E= 0.156	to	E= 0.245	ΔE= 0.089	
Band 6	(B3)	E= 0.260	to	E= 0.449	ΔE= 0.189	
Band 7		E= 0.505	to	E= 0.750	ΔE= 0.245	
Band 8		E= 0.785	to	E not on graph		

5. A heteronucleotide model which is a combination of models 1 and 3 in the following manner.



The results are shown in diagram 5 and the table below lists the band edges and the band widths.

Table 5

Band 1	(A1)	E=-0.8073	to	E=-0.8073	ΔE≤0.0001
Band 2	(C1)	E=-0.3511	to	E=-0.3491	ΔE=0.0020
Band 3	(A2)	E=-0.2929	to	E=-0.2876	ΔE=0.0053
Band 4	(C2)	E=-0.0038	to	E=0.01797	ΔE=0.0142

5.3 Conclusion

It can be clearly seen from the results that the bands are narrowed in the case of heteronucleotides. The results are tabulated in tables 6 and 7.

Table 6

Homonucleotide		Heteronucleotide
$V=-1$ $a=3.6$ Bohr Radii	$V=-0.75$ $a=3.6$ Bohr Radii	$V=-1$ and -0.75 $a=3.6$ Bohr Radii
(A1) $\Delta E=0.0041$	(B1) $\Delta E=0.0076$	(A1) ΔE 0.0001
(A2) $\Delta E=0.044$	(B2) $\Delta E=0.082$	(B1) $\Delta E=0.0013$
(A3) $\Delta E=0.277$	(B3) $\Delta E=0.307$	(A2) $\Delta E=0.0030$
		(A3) $\Delta E=0.0112$
		(B2) $\Delta E=0.089$
		(B3) $\Delta E=0.189$

Table 7

Homonucleotide		Heteronucleotide
$V=-1$ $a=3.6$ Bohr Radii	$V=-0.50$ $a=3.6$ Bohr Radii	$V=-1$ and $V=-0.50$ $a=3.6$ Bohr Radii
(A1) $\Delta E=0.0041$	(C1) $\Delta E=0.0180$	(A1) $\Delta E\leq 0.0001$
(A2) $\Delta E=0.044$	(C2) $\Delta E=0.0660$	(C2) $\Delta E=0.002$
		(A2) $\Delta E=0.0053$
		(C2) $\Delta E=0.0142$

Table 8 gives the band edges for the various bands, and also the energy levels of the corresponding isolated wells.

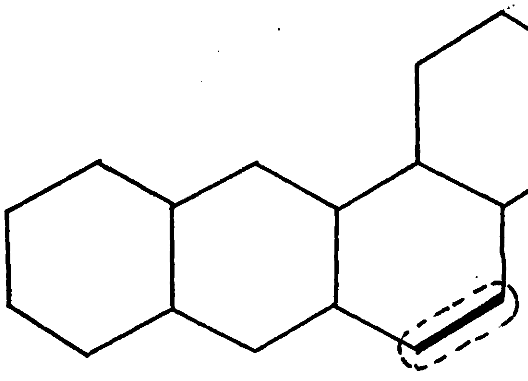
Table 8

Band	Energy of "atomic" level	Edges in Homostructure		Edges in Heterostructure 4		Edges in Heterostructure 5	
		From	To	From	To	From	To
A1	-0.8076	-0.8095	-0.8054	-0.8084	-0.8084	-0.8073	-0.8073
A2	-0.2921	-0.313	-0.269	-0.294	-0.291	-0.293	-0.288
A3	+0.381 (resonance)	0.133	0.410	0.156	0.245		
B1	-0.5746	-0.5783	-0.5707	-0.536	-0.534		
B2	-0.1312	-0.162	-0.080	0.131	0.143		
B3	+0.381 (resonance)	0.178	0.485	0.260	0.449		
C1	-0.3490	-0.3600	-0.3420			-0.351	-0.349
C2	-0.0188	-0.04	0.089			-0.03	0.00

In model 4 the correlation of the 4th and 5th bands with A3 and B2 is not straightforward. These two bands are hybridized. The same phenomenon was observed by Ladik in a number of his models.

Appendix 1

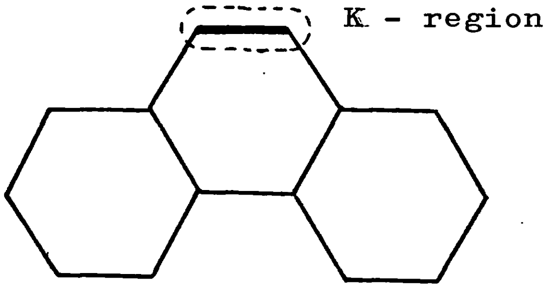
It has been claimed that there is a certain region, called the K region, in a condensed polycyclic molecule which, because of the high concentration of π-electrons, is able to possess carcinogenic properties. There are various arguments concerning the origin of these properties, but all accept that they lie in the π-electron distribution. However, they differ in the particular combination of derived quantities such as bond order, net charge, free valence, energy of excitation, resonance energy etc. They all depend on which method one takes to calculate the above quantities, namely the valence band method or the molecular orbital method.
The K region is an 'energy rich' region. Its very shape makes it exceedingly accessible to any approaching radical. It is less easily screened than most of the rest of the molecule. It might be called an 'exposed bond.'



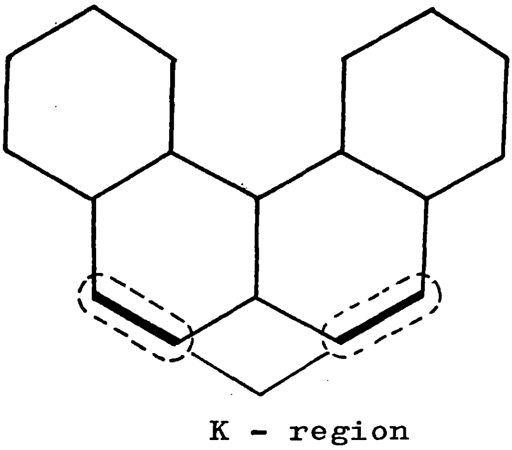
1,2 Benzanthracene

K - region

e.g.



Phenanthracene



3,4 Benzphenanthrene

The K region possesses unusual electrical properties. The high bond order marks it out at once as likely to be significant. The relative case with which charge can be moved to this region gives an explanation of the power of methyl substitution to enhance a carcinogenicity which is sometimes latent.

The carcinogenic properties are suggested to be correlated with:

a. Bond order

b. Charge density and distribution

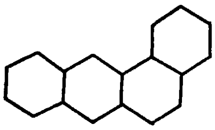
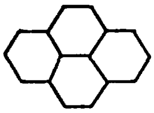
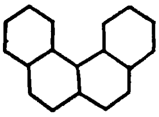
c. Free valence

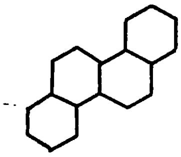
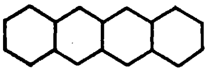
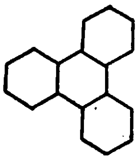
These quantities, a combination of which is called the 'electrical index' are all calculated using either the V.B. approximation or the M.O. approximation.

It has been suggested that there is a threshold value in the K region. When the electrical index of the K region exceeds this threshold value then the substance is carcinogenic.

a. Bond order

Table 1

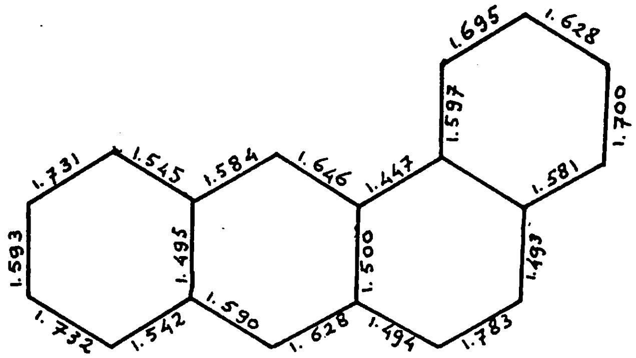
	Molecule	Bond Order	
		M.O.	V.B.
	1, 2 Benzanthracene	1.783	1.440
	Pyrene	1.777	
	3,4 Benzphenanthrene	1.762	1.442

	Molecule	Bond Order	
		M.O.	V.B.
	Chrysene	1.754	1.434
	Naphthacene	1.741	1.452
	Triphenylene	1.690	1.446

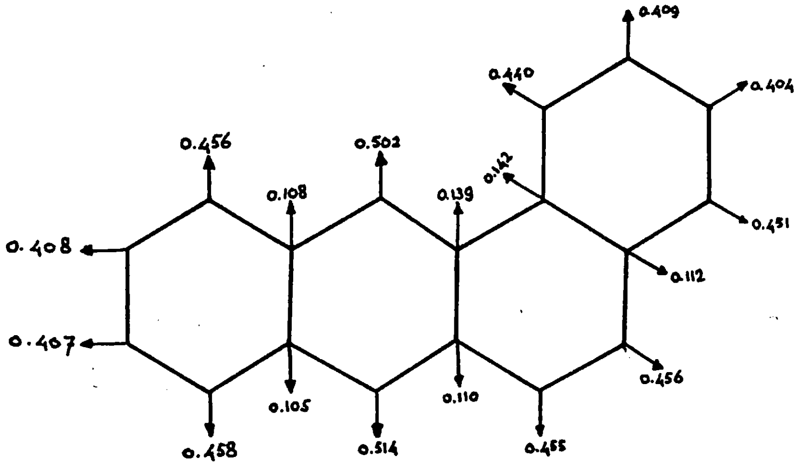
Considering the results of the calculation done one would be compelled to think that 1,2-Benzanthracene is very carcinogenic. In fact it is 3,4-Benzphenanthrene that is very strongly carcinogenic. The only way in which we could suppose that bond order was significant would be to assume that the K region has a high bond order, but certain other conditions have also to be fulfilled.

b. This has been proved to be hardly significant because the π-electrons distribute themselves equally over aromatic framework in both theories (V.B. and M.O.)

c. Take the result of the calculation on benzanthracine, using V.B. method.



Bond Orders



Free Valences

Immediately one can conclude that there is nothing special about the free valence in this region. A further objection arises from the fact that aza replacement to convert benzathracene into benzacridine, causes the sum of the free valences to increase. However this replacement is known to reduce carcinogenic power.

The idea of total charge in the K region was suggested and calculated by Mme. Pullman. The total charge in the K region is

$$2p_{12} + F_1 + F_2 + q_2 + p_1 + p_2$$

where p_{12} is the mobile bond order of the K region

F_1 and F_2 are free valences at the two atoms

q_1 and q_2 are the contributions of course structure to the atoms

p_1 and p_2 are the mobile order of the two bonds which connect this bond with the rest of the molecule, see diagram.

Pullman found that the threshold value is 1.291, above which carcinogeneity is almost certain to occur. The total charge at the K region, according to Pullman's theory is tabulated in table 2.

The mechanism of cancer-induction, according to Ladik, would be the unwinding of the helices of DNA caused by the repulsion of electric charges at the two ends. This polarisation of the helix would occur in an electric field because the mobile electrons in the conduction band and holes in the valence band would become separated. The hypothesis is put forward that

there is charge transfer from the K-regions of carcinogenic compounds to the DNA helix.
Of course several other factors should be taken into consideration. The above hypothesis was the initial cause for deciding to work out the electronic structure of DNA.

Coulson[1] suggested that radiation-induced cancer occurs when the energy of the radiation is higher than the so called "threshold energy". He suggested that this threshold energy is 3.4eV.
Hoffmann and Ladik[2] showed that the DNA molecule in its ground state is an insulator with an energy gap of 3.5eV. Energy from radiation can promote an electron from the valence band to the conduction band, thus causing the macromolecule to become a conductor.
Mason[3,4] suggested that the mobility of this electron has some connection with the occurence of cancer. However, the mechanism of this connection is not known for certain. It could be that there is a strong local electric field inside the cell, for example from the presence of dipolar molecules, from local differences in the ion concentration,etc. The field has to be strong along the axis to prevent quick recombination of the exciton.
In normal circumstances as DNA is an insulator, this

static electric field causes a deformation of the charge cloud of the individual nucleotide bases. However, when DNA becomes conducting, then this field, when it is in the direction of the axis of the macromolecule, will produce a migration of the π-electrons, thus a statistical probability exists for a net electric charge to appear on the ends of an excited or ionised DNA molecule, the molecule is polarised. It is understood that both parts of the molecule are polarised, because the delocalized π-electrons system belong to the double helices.

It has been suggested that polarisation of the molecule can cause replication of the double helix.

If the molecule becomes polarized there are charges of the same magnitude and identical signs at the ends of the chains. The potential energy of the repulsion due to these charges might induce the duplication mechanism. Thus, under the influence of radiation or carcinogenic agents the DNA molecule might duplicate at a time which is not determined by the coorperation of the whole organism. In this way, one may assume that in the cells of some tissue a great number of new DNA molecules can appear in an instant which can lead to mitosis of these cells. However, these mitosis is not determined by the growth regulation of the orgaism and is therefore undesired. It is then possible that this might induce a series of such irreversible biochemical reactions in the tissue in question, which converts these cells from normal into tumor cells.

Burch[5] suggested that the radiation induced cancer happened only when the radiation cause two specific chromosome breaks.

Butler[6] explained the mechanism of this break. However, the probability that the radiation will hit those special chromosomes at special place is very small indeed.

References:

1. Coulson, C.A. , Advances in Cancer Research, I: 1 - 58 , New York, Academic Press, 1953

2. T.A.Hoffmann and J.Ladik, Cancer Research, 1961, 21, 474

3. Mason R., Nature, 181, 820 - 822, 1958

4. Mason R., Disc. Far. Soc., 27, 129 - 33 , 1960
5. Burch, P.R. J., Nature, 185,135 - 42, 1960

6. Butler, J. A.V., Radiation Research, 4, 20 - 28, 1956

Total Charge/or density at K region in Pullman's Theory

Compounds	Total Charge	Carcinogenic Activity	
		Skin	Sub-cutaneous Tissue
Naphthacene	1.258	0	-
Anthracene	1.259	0	0
Triphenylene	1.260	0	-
-Benzacridine	1.260	0	-
1,2-Benzacridine	1.270	0	-
Chrysene	1.272	+	+
5-Methyl-3,4-benzacridine	1.273	0	-
Naphthalene	1.274	0	-
1,2-Benzanthracene	1.283	+	+
5,8-Dimethyl-3,4-benzacridine	1.284	+	0
5,7-Dimethyl-3,4-benzacridine	1.285	0	-
5,9-Dimethyl-3,4-benzacridine	1.286	0	+
Phenanthrene	1.291	0	-
8-Methyl-1,2-benzanthracene	1.292	+	0
5-Methylacridine	1.293	+++	-
3,4-Benzphenanthrene	1.293	+	0
7-Methyl-1,2-benzanthracene	1.294	+	+
6-Methyl-1,2-benzanthracene	1.294	+	-
9-Methyl-1,2-benzanthracene	1.296	++	+++

Compounds	Total Charge	Carcinogenic Activity	
		Skin	Sub-cutaneous Tissue
5-Methyl-1,2-benzanthracene	1.296	++	++
3-Methyl-1,2-benzanthracene	1.298	+	++
4-Methyl-1,2-benzanthracene	1.298	+	++
5,7,9-Trimethyl-3,4-benzacridine	1.298	+++	0
5,9-Dimethyl-1,2-benzacridine	1.302	+++	+++
5,8-Dimethyl-1,2-benzacridine	1.304	++++	+++
5,7-Dimethyl-1,2-benzacridine	1.304	++++	+++
10-Methyl-1,2-benzanthracene	1.306	+++	++++
5,6-Dimethyl-1,2-benzanthracene	1.307	+++	-
5,9-Dimethyl-1,2-benzanthracene	1.309	-	++++
8-Methyl-3,4-benzphenanthrene	1.309	+	0
6-Methyl-3,4-benzphenanthrene	1.310	+	+
4,9-Dimethyl-1,2-benzanthracene	1.311	-	+++
1-Methyl-3,4-benzphenanthrene	1.312	++	0
2-Methyl-3,4-benzphenanthrene	1.312	+++	+
5,7,9-Trimethyl-1,2-benzacridine	1.312	+++	++
7-Methyl-3,4-benzphenanthrene	1.313	+	0
5,10-Dimethyl-1,2-benzanthracene	1.317	-	++++
9,10-Dimethyl-1,2-benzanthracene	1.319	++++	+++
4,10-Dimethyl-1,2-benzanthracene	1.321	-	++
6,9,10-Trimethyl-1,2-benzanthracene	1.330	++++	++
5,9,10-Trimethyl-1,2-benzanthracene	1.332	++++	+++
5,6,9,10,Tetramethyl-1,2-benzanthracene	1.343	+++	+

APPENDIX 2

Computer Programm for Chapter 5

```
      PROGRAM BAND (INPUT,OUTPUT,TAPE5=INPUT,TAPE6=
      OUTPUT)
      COMPLEX EIGEN, WAVE
      COMPLEX AYE
      DIMENSION V(20),D(20)
      DIMENSION COSKA(400),EN(400)
      DIMENSION TR(2,2)
      AYE=(0.0,1.0)
      READ(5,100)  N
C     READ NUMBER OF SECTIONS NECESSARY TO MAKE UP
C     THE POTENTIAL OVER ONE PERIOD. THIS IS THEN
C     REPEATED INDEFINETELY
  100 FORMAT(I10)
      WRITE(6,101)  N
  101 FORMAT(1H1,50X,*NUMBER OF SECTIONS IN ONE
      PERIOD OF POTENTIAL*,I10)
     +0)
      DIST=0
      DO 1 J=1,N
      READ(5,102)    D(J),V(J)
  102 FORMAT(2F10.2)
C     THE LENGTH OF EACH SECTION AND THE CONSTANT
C     POTENTIAL THERE.
```

```
        DIST=DIST+D(J)
  1     WRITE(6,103)   D(J),V(J)
  103   FORMAT(20X,F10.2,*AT POTENTIAL*,F10.2)
        READ(5,104)  E1,E2,M
C       THE STARTING AND FINISHING ENERGY OF THE SCAN
C       AND THE NUMBER OF POINTS AT WHICH TO CALCULATE
  104   FORMAT(2F10.2,I10)
        WRITE(6,105)  E1,E2,M
  105   FORMAT(*STARTING*,E20.3,*FINISHING*,E20.3,
      + *NO. OF POINTS*,I10)
        DDD=DIST/N
        DO 40 K=1,200
        EN(K)=0
  40    COSKA(K)=0
        DO 2  K=1,M
        E=E1+FLOAT(K-1)*(E2-E1)/(M-1)
        CALL UNIT(TR)
C       THAT SETS THE INITIAL TRANSFER MATRIX TO UNITY
        DO 3   J=1,N
  3     CALL TRANS(D(J),E-V(J),TR)
C       EACH CALL MULTIPLIES THE MATRIX BY THE
C       TRANSFER MATRIX FOR ONE SECTION
        WAVE=CLOG(EIGEN(TR))*(0.0,-1.0)/DIST
        COSKA(K) = 0.5*  (TR(1,1)+TR(2,2))
  20    WRITE(6,108)     E,WAVE,COSKA(K)
  108   FORMAT(10X,*ENERGY*,E20.8,*WAVENUMBER*2E20.8,
       + *HALFSUM*,E20.8)
```

```
      EN(K)=E
 2    CONTINUE
      CALL GRAFIT(COSKA,EN,200,200,0,1)
      STOP
      END

      SUBROUTINE UNIT(A)
      DIMENSION A(2,2)
      DO 1  I=1,2
      DO 1  J=1,2
      A(I,J)=0
      IF(I.EQ.J)       A(I,J)=1.0
 1    CONTINUE
      RETURN
      END

      SUBROUTINE FILL(D,S,TR)
C     PUTS TRANSFER MATRIX IN TR FOR SECTION OF
C     LENGTH D , S=E-V   PRESENTLY RESTRICTED TO S
C     POSITIVE.
      DIMENSION TR(2,2)
      IF(S)  1,2,3
 3    CONTINUE
      GAMMA=SQRT(2.0*S)
      TR(1,1)=COS(GAMMA*D)
      TR(2,2)=TR(1,1)
      TEMP=SIN(GAMMA*D)
```

```
        TR(1,2)=TEMP/GAMMA
        TR(2,1)=-TEMP*GAMMA
        RETURN
  2     CALL UNIT(TR)
        TR(1,2)=D
        RETURN
  1     GAMMA=SQRT(-2.0*S)
        E=EXP(GAMMA*D)
        TR(1,1)=(E+1,0/E)*0.5
        TR(2,2)=TR(1,1)
        ES=(E-1.0/E)*0.5
        TR(1,2)=ES/GAMMA
        TR(2,1)=ES*GAMMA
        RETURN
        END

        SUBROUTINE MULT(A,B)
C       MULTIPLIES MATRIX A BY B LEAVING FINAL ANSWER
C       IN A DIMENSION A(2,2), B(2,2), C(2,2)
        DO 1   I=1,2
        DO 1   J=1,2
        C(I,J)=0
        DO 1   K=1,2
  1     C(I,J)=C(I,J)+A(I,K)*B(K,J)
        WRITE(6,100)  ((A(L,LL),LL=1,2),(B(L,LL),LL=1,2),
       +(C(L,LL),LL=1,2),L=1,2)
  100   FORMAT( /,2(3(10X,2E15.7)/))
```

```
      DO 2    I=1,2
      DO 2    J=1,2
  2   A(I,J)=C(I,J)
      RETURN
      END

      SUBROUTINE TRANS(D,S,A)
C     MULTIPLIES TRANSFER MATRIX A BY MATRIX FOR
C     NEXT SECTION
      DIMENSION TR(2,2), A(2,2)
      CALL FILL(D,S,TR)
      CALL MULT(A,TR)
      RETURN
      END

      COMPLEX FUNCTION EIGEN (TR)
      DIMENSION TR(2,2)
      COMPLEX ROOT
C     THIS FINDS THE EIGENVALUE OF A REAL 2NX 2
C     MATRIX
      EIGEN=ROOT(1.0,-TR(1,1)-TR(2,2),TR(1,1)*TR(2,2)
     +-TR(2,1)*TR(1,2))
      IF(EIGEN.EQ.(0.0,0.0))  EIGEN=(1.0E-200,1.0E-200)
      RETURN
      END

      COMPLEX FUNCTION ROOT(A,B,C)
```

```
C       FINDS THE ROOTS OF A REAL QUADRATIC PHASE
C       CONVENTION IS IF REAL ROOTS CHOOSE THE LARGER
C       IF IMAGINARY THE ONE WITH POSITIVE IMAGINARY
C       PART
        REAL A,B,C,DISC
        DISC=B*B-4.0*A*C
        IF (DISC.LT.0)    GO TO 1
        ROOT=CMPLX((-B+SQRT(DISC))/2.0/A,0.0)
        RETURN
   1    ROOT=CMPLX(-B,SQRT(-DISC))/2.0/A
        RETURN
        END
```