Journal of Mathematics, Physics and Astronomy, New Series Volume 2 (2006-2007)

The Mathematical Modeling And The Structure of Six Transitions of DNA

Türkay Yeğenoğlu*
Ruhr University, Institute of Physiological Chemistry, Bochum, Germany

(Accepted 25 October 2007)

Knots, actually binding two ends of yarn together, are constituting one of the most important theories in the science of Mathematics. For about 75 years mathematicians are investigating how two end points can bind after passing through each other. After finding the model of one-transition, two-transition and three-transition, a short while ago the secrets of the model of the simple five-transition knot could be solved. The solution of this secret will not only be helpful to the science of Mathematics but probably also a great deal to the area of biology. The DNA molecules which make it possible that characteristics of personality pass from one generation to another are a six-transition knot. The geneticists are convinced that the changes in the form of knots lead to differentiation of the species. Scientists who solved the secrets of the five-transition knot emphasize that this will lead to inventions concerning the construction of DNA if secrets of the six-transition knot are found. After the solution of higher transition numbers of knots are found genetics as well as molecular chemistry will profit from the related applications.

1. INTRODUCTION

This doctoral thesis is divided into two sections. The first section describes the mathematical modeling of six transitions of DNA. DNA is the basis of our genotype and its structures can vary. For approximately 75 years, mathematicians have been researching how twdo end points manage a transition and how they link up again after having passed through one another. Consequently, the form of DNA can feature transition points, known as transition knots. We have known of the mathematical modeling of the 1-5 transition knots of DNA since 1986 when it was first presented (7). This presentation of the five transition knots is illustrated in the following figure:

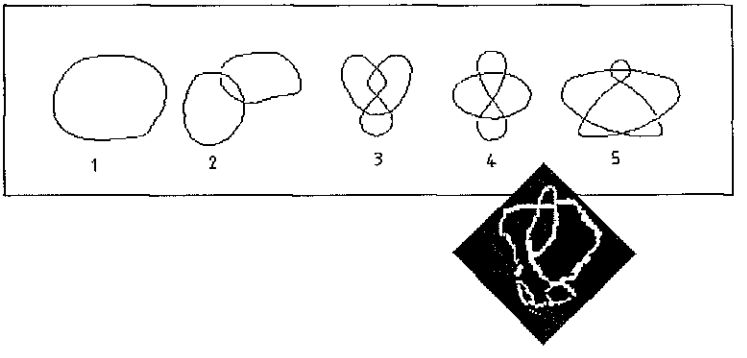


Fig. 1 DNA molecules which transfer personal characteristics from one generation to another consist of one six transition knot. Geneticists are convinced that changes in the shape of the knot would lead to a differentiation in the genus "species". The basis of the thesis is DNA with six transition knots and is explained here mathematically for the first time.

2. METHODS

The sequence space is presented graphically (7) (see 3.1). This form of presentation was transferred to DNA and taken over in this thesis with triangles (pyramids) instead of squares.
Gold Section Method (see 3.2).
The Gold Section method can be frequently observed in nature, for example, in plants (leaves, flowers, branches), in crystals and snow-flakes and in many other life-forms.
The following figure shows the sectional structure of a snow-flake.

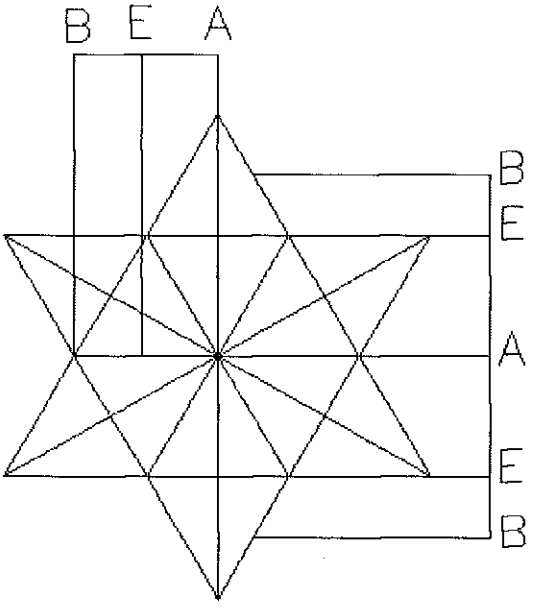


Fig. 2

I have used the sectional drawing of a snow-flake crystal represented in the Golden Section method as a basis model for DNA.
This form is most suitable for calculation the inner transitions using the Fourier sequence (see 3.3).
In order to model the transition points and inner bonding points the logarithmic term of six transitions of DNA had to be determined (see 3.4). This resulted in the model with the transition areas. The differential equation system (see 3.5)
from A. Karadeniz was applied in order to be able to use the works of J. Chotai (see 3.6). The systems used here are based on those of A. Karadeniz, however, I have developed them further independently. I have created an additional genetic model using J. Chotai's (see 3.5) likely ratio test, in which matrix calculus was used. The solution of a normal differential equation system with linear and constant coefficients of six transitions of DNA using matrices was recorded.
Finally, the mathematical drawing of six transitions of DNA was presented.

3. THE MATHEMATICAL STRUCTURE OF SIX TRANSITIONS OF DNA

3.1 The mathematical structure of six transitions of DNA with the sequence space metod.

The presentation of the mathematical modeling of DNA from 1-5 transition knots had been so far known to mathematicians only. Following on from there, I have undertaken the modeling of a six-figured DNA.
the following drawing shows a mathematical representation of DNA and the modeling of its 1-5 transition knots.

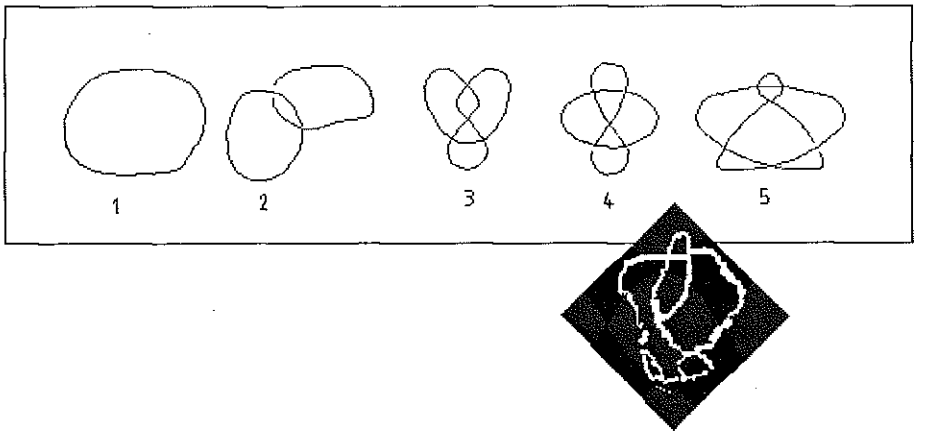


Fig. 3

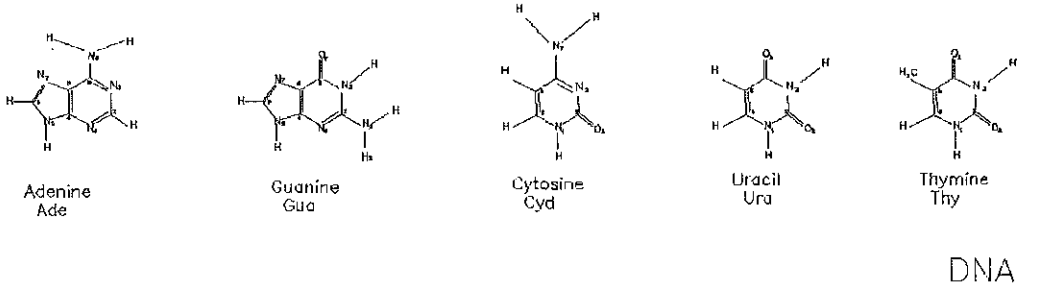


Fig. 4

Using the sequence space method, the diverse nucleotide sequences in similar viral genomes and their representation can be mapped.
The following figure illustrates this method (7).

How to Construct a Sequence Space
One way to study the diverse nucleotide sequences in the genes of viruses is to map them into a multidimensional matrix called a Hamming sequence space. In this space, each point

represents a unique sequence, and the degree of separation between points reflects their degree of dissimilarity. The space can be most easily drawn for short sequences consisting of binary digits. For a sequence with just one position, there are only two possible sequences, and they can be drawn as the end points of a line (a). For a sequence with two positions, there are four permutations, which form the corners of a square (b). The variations on a three-digit sequence become the corners of a cube (c), and the variations of a four-digit sequence are the vertices of a four dimensional hypercube (d). Each higher-dimensional space is built iteratively by drawing the previous diagram twice and connecting the corresponding points. The sequence spaces for viral genomes are far more complex than these simple figures because they involve thousands of positions that can each be occupied by one of four different nucleotides.

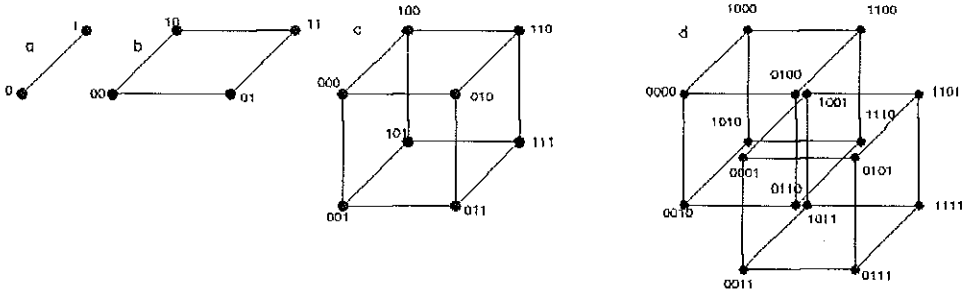


Fig. 5

3.1.1 How is a Sequence Space Constructed?

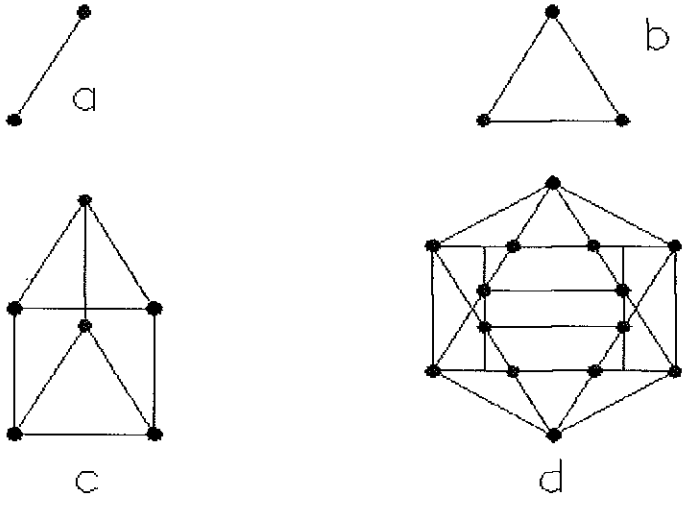


Fig. 6

In a multi-dimensional matrix called "sequence space" the cards showing the points of diverse nucleo-protein sequences can be mapped so that the nucleotide sequences in similar viral genomes can be investigated. Each point in this space represents a unique nucleo-protein sequence and the degree of separation between points reflects the degree of their dissimilarity. The space can be most easily drawn for short sequences consisting of binary digits. For a sequence with just one position, there are two possible sequences and they can be drawn as the end points of a line.

a. It can be a sequence with two positions, with four permutations, which form the comers of a square;

b. or the variations of a sequence with three digits, which are the corners of a triangle;

c. or the variations of a sequence with four digits;

d. or the comer points of a three-dimensional prism.

In each of the viral genomes "sequence spaces" there are thousands of positions which four different nucleoproteides can occupy. These are more complicated than simple models.

3.2 The Golden Section method as a mathematical rule of formal harmony of six transitions of DNA which can be illustrated by means of a sequence space method.

In this part of the mathematical modeling, the six-digit DNA is modeled applying the Golden Section method.

This method of modeling represents a key system to the mathematical thesis of DNA in connection with the Fourire sequence which forms the next part of my work.

For this modeling method I have applied methods according to Preiß which graphically illustrate the schematic representation of the Golden Section method in a snow flake crystal (8).

This schematic representation and significant information concerning the Golden Section method are shown as follows:

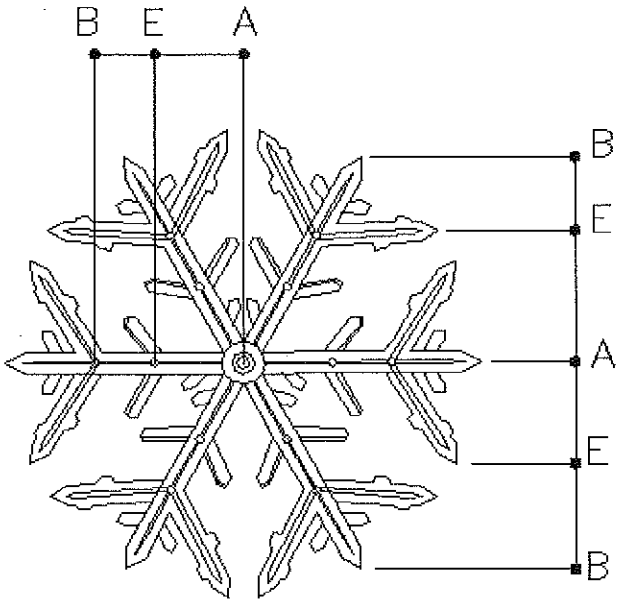


Figure 7

Graphicists and designers construct forms, whereby the world is not only to be understood to mean an object (a body). we draw forms, construct figures with a great variety of forms, design new types of script, and select formats for pictures and paper. The world form comes from the Latin "forma" and means outline, shape or appearance. It is subject to an order, a rule, we evaluate it according to formal and aesthetic appearance, according to its proportions. If these principles of design are taken into consideration, the results will be felt to be harmonious, beautiful, aesthetic and valid. If a graphical or plastic representation fulfills all these criteria, it is perfectly shaped.

Mathematicians, philosophers and artists in the times of the ancient Pharaohs occupied themselves with finding proportions in the human anatomy, to recognize it as "the measure of all things". In the pyramid of Memphis the first "Canon" about the proportions of the human body was discovered. Polyclitus, the great sculptor in the fifth century BC wrote the "Canon", a work discussing the ideal proportions of the human figure.

The Ancient Romans developed a theory of proportions and in the Renaissance Leonardo da Vinci, Michelangelo and above all, Dürer, defined their knowledge concerning the law of harmony. The geometrical configuration of a snow flake (Fig. 7) is a hexagon, a six-pointed star.

The following figure represents the mathematical modeling of DNA which I have created using the Golden Section method.

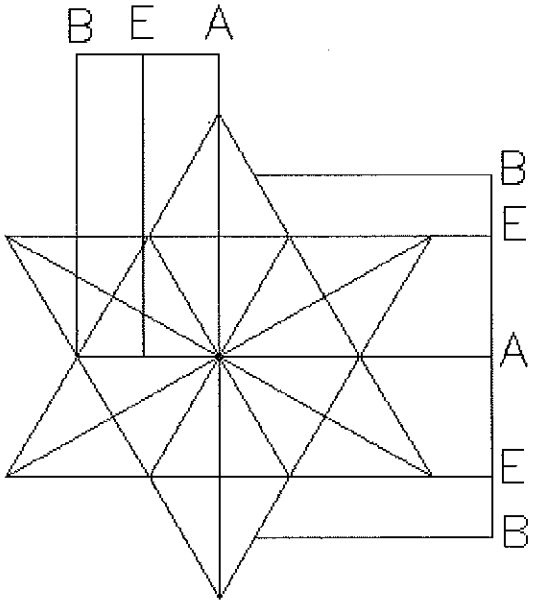


Fig. 8

3.3 Fourier Sequence

This section of my work consists of the modeling of the DNA using the Fourier sequence. The Fourier sequence is used today in the filed of telecommunications especially for transmitting frequencies for radio and TV and also for calculating the range of micro-waves.

I have dealt with this method especially because of the mathematical link between the building blocks (A, G, C and T) with which DNA is created and realized the mathematical modeling. In order to realize this method of modeling I have used, as already mentioned, the Golden Section method as a key system (9).
The following figure illustrates this modeling method

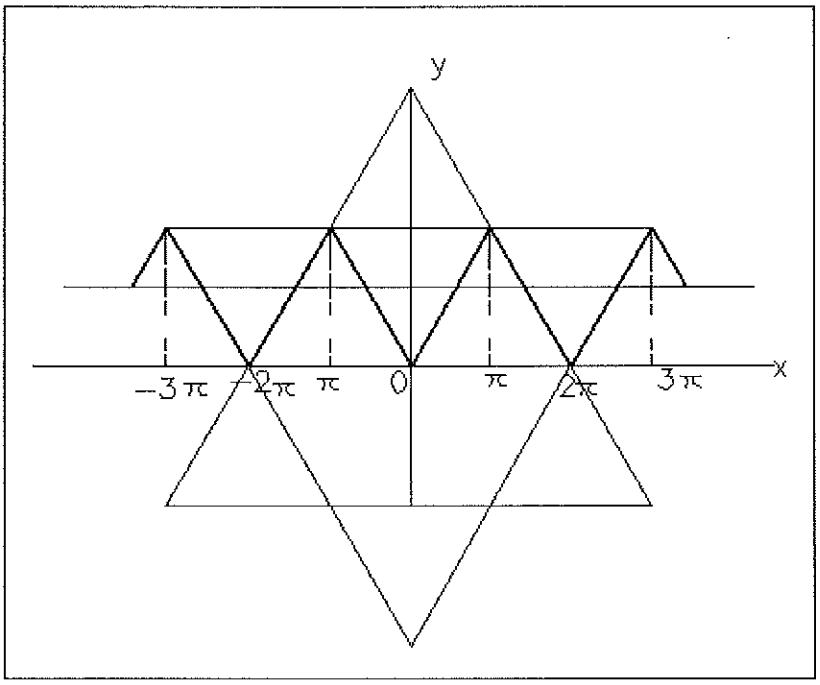


Fig. 9

3.4 The Logarithmic Term of Six Transitions of DNA
In this part of my work, the logarithmic term for six transitions of DNA must first be determined in order to model the transition points and inner connection points. This resulted in the model with the transition areas and the determination of the parable slopes (10).
As this section contains the most important part of my work and simultaneously leads to the conclusion, the above-mentioned modeling is represented by the following figure:

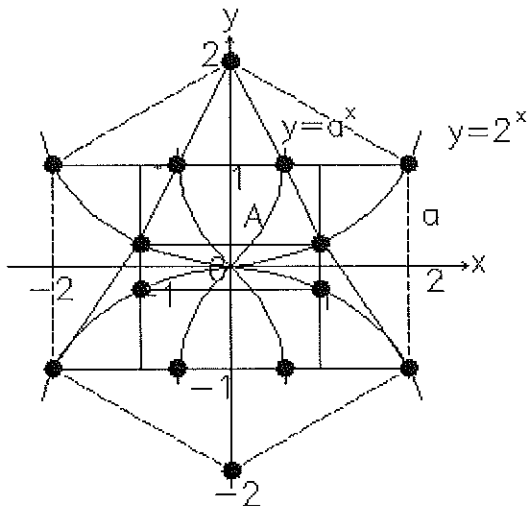


Fig. 10

3.4.1 Exponential Function and Logarithm X

The exponential function.

Crneate a graph for the function

$y=2^x$ for x e (0, 1, 2, 3, 4, -1, -2, -3, -4)

and draw the points in the system of coordinates

Place x for 1/2, 3/2 etc.

The function $x\text{-}a^x$ or $y=a^x$ defined in this way is called the exponentional function because the independent variable x is in the exponent.

The monotony $y=a^x$ is

a) $a>1$ and $x_1<x_2$, therefore $a^{x2}:a^{x1}= a^{x2-x1}$ with $x_2-x_1>0$.

Therefore, for $a>1$ and $x_1<x_2$, a^{x1} is $< a^{x2}$.

b) $0<a<1$ ans $x_1<x_2$ and $x_1<x_2$ is therefore correspondingly $a^{x1} < a^{x2}$.

The exponential function $y=a^x$ increases monotonously for $a>1$ and decreases monotonously for $0<a<1$.

The constant function $y=1$ is obtained for $a=1$.

a) The exponential curves with the equation $y=a^x$, $a>0$ runs completely above the x axis.

b) For $a>1$, the curves increase monotonously (to the right).

For $0<a<1$ they decrease monotonously.

c) If a is a form variable then all curves pass through the point A (0/1) through the equation $y=a^x$.

d) Curves with the equation $y=a^x$ and $y=(1/a)^x$ are symmetric to one another or the y axis.

Proof: $y=(1/a)^x = 1/a^x = a^{-x}$ stems from $y=a^x$ by exchanging +x with -x.

e) The chart of $y=a^x$ has the negative x-axis as asymptote for $a>1$; and the positive x-axis for $0<a<1$.

3.5 Using matrices to solve a normal differential equation system with linear and constant coefficients of six transitions of DNA whose logarithmic term has been described.

I have used this section as a key system in order to achieve the genetic modeling of DNA (3.6) in the following section.

To do so, I established a connection to work by Chotai which I used as a backup for the DNA and by using matrices of a normal differential equation system with linear and constant coefficients created a key system, thus achieving the genetic modeling (11).

With the help of matrices, this key system is shown briefly in the following figure.

The formula was created with the help of matrices as a key system for Likelihood Ratio tests. It is shown here in an abbreviated form.

$$\frac{1}{a^{x_i}} = \sum_{j=1}^{n} a_{ij} X_j + f_i^{(t)}$$

3.6 The genetical and mathematical modeling of six transitions of DNA

3.6.1 Likelihood Ratio Test

It was shown for the fixed sample size approach that

$P[z(a^x)]<1/A$ when $a= 0.5$

holds for every choice of a^x. The proof relies on the fact that the expectation $E[\text{antilog } z(a^x)]= 1$ when the true $a=0.5$. Now since $z(a)\geq z(a^x)$ by definition and $z(a)>z(a^x)$ with positive probability for each a^x. I have

$E[\text{antilog } z(a)]>E[\text{antilog } z(a^x)] = 1$

when $a=0.5$. So the proof breaks down if I replace $z(a^x)$ by $z(a)$ in the above inequality. It has sometimes been claimed due to a logical slip that the inequality continues to hold for $z(a)$ that this claim is not true in general can be seen from the following example.

Example: Suppose that I have a binomial random variable with success probability a. $0 \leq a \leq 1$. For each of n trials the probability of obtaining r successes is given by

$$P(r:a) = \left(\frac{n}{r}\right) a^r (1-a)^{n-r}$$

the maximum likelihood estimate of a given by a = r/n with

$$P(r:a) = \left(\frac{n}{r}\right)\left(\frac{n}{r}\right)^r \left(1-\frac{r}{n}\right)^{n-r}$$

(n:a) at a=1 where this llikelihood attains a value of 1 in both cases. For $n = 2^x$. I therefore have antilog $z(a) = 2^x$. If r=0 or r=n so the probability that $z(a)>2$ when the true $a = 0.5$ is given by P (0:0.5) + P (n:0.5) = (1/2) + (1/2) = 1

letting $1/2<a<1$

This completes the example

I now give arguments to indicate the context of linkage analysis. The inequality $P[z(a)>2^xA]1/A$ is often likely to hold when $a = 0.5$ if I consider the fixed sample size approach.

For small samples I do not have any general proof for the inequality. But I give some calculations that seem to indicate its validity in the context of linkage analysis. Since the recombination a obeys $00 \leq a0 \leq 0.5$, 1 set $a = 0.5$ if r/n exceeds 0.5 and a = r/n otherwise. Thus, the hypothesis H_0:a = 0.5 is rejected when the number r of recombinants is small

enough compared with the total number n opportunities. So this a one-sided test. Now when I obtain r recombinants.
$z(a) = r\ 2^x\ r + (n-r)2^x(n-r)+n2^x(2/n)$ since the derivative of z(a) with respect to r given by log[r/(n-r)] is negative for $r<n/2$ it follows that z(a) is decreasing in r for $r < n/2$. Therefore the quality $z(a)>2^xA$ is equivalent to the number of recombinants not exceeding a certain number r depending on the given $A>1$. This probability is given by

$$2^{-n}\sum_{j=0}^{r}\left(\frac{n}{j}\right) \quad \text{when a} = 0.5$$

since antilog $z(a)>A$ here it suffices to show that this probability is less than 1/antiolg z(a) so I have to show that

$$\left(\frac{r}{n}\right)^{r}\left(\frac{n-r}{n}\right)^{n-r}\sum_{j-0}^{r}\left(\frac{n}{j}\right)\langle 1$$

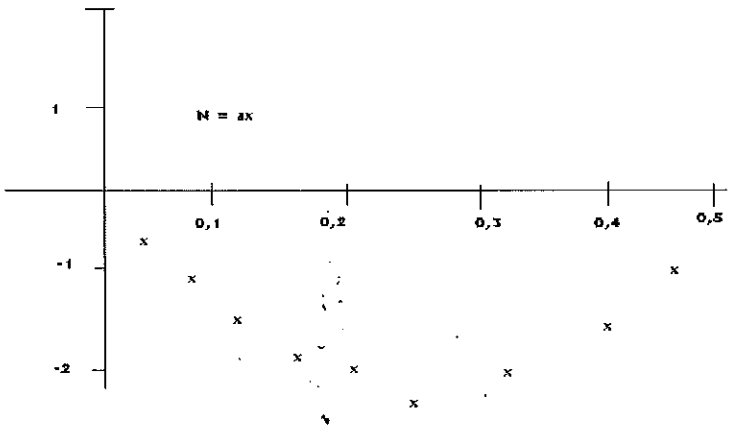
examining the tables of binomical coefficient for $n \leq 2^x$, I find that for roughly $r < n/3$. The inequality

$$\sum_{j-0}^{r-1}\left(\frac{n}{j}\right)\langle\left(\frac{n}{r}\right)$$

seems to hold. Moreover, for various n, the probability P(r; r/n) of obtaining exactly r recombinants when a = r/n does not usually exceed 1/2. These two points imply that

$$\left(\frac{r}{n}\right)^{2}\left(\frac{n-2}{n}\right)^{n-r}\sum_{j-0}^{r}\left(\frac{n}{j}\right)\langle 2p^{(r;r/n)}\langle 1$$

roughly holds. I have thus given a rough justification for the inequality $P[z(a) > 2xA] < 1/A$ in the context of linkage analysis, provided I interpret it in terms of the fixed sample size approach and in a one-sided manner.
$N = a^x \qquad z(a^2)$



Lod scores in linkage analysis

Fig. 11

4. RESULTS AND DISCUSSION

The following figure shows the result of my work. It represents the parabel slopes of DNA , especially the points where the slopes cross and where they link up.

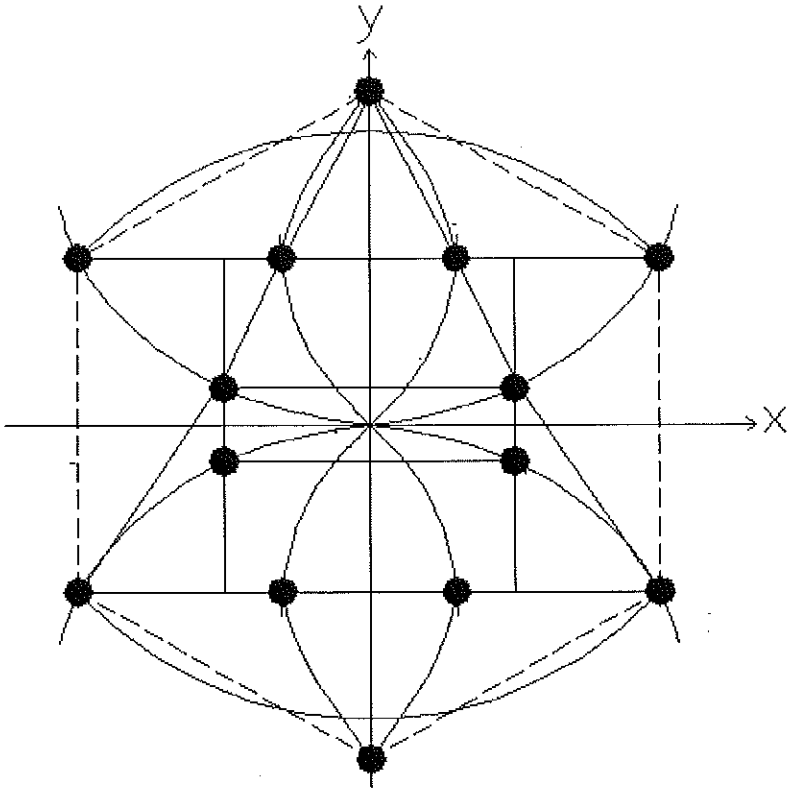


Fig. 12

This (Qy.18) is the mathematical representation of the DNA of a normal, healthy cell.
Other forms may occur under pathological conditions. How far this is the case in vivo has to be demonstrated in experimental work.

5.SUMMARY

So far six transitions form of DNA were reported. Based on mathematical model calculations a sixth form is postulated in this thesis.
This form of DNA has to be proved by experimental techniques.

REFERENCES

1. C.R. Calladine (1982), Mechanism of sequence-dependent stacking of bases in B-DNA. J. Mol. Biol. 161, 343-352

1a. R.E. Dickerson (1983), Base sequence and helix structure variation in B- and A-DNA. J. Mol. Biol. 166, 419-441

2. B.N. Conner, T. Takano, S. Tanaka, K. Itakura and R.E. Dickerson (1982) The molecular structure of d(^{1}CpCpGpG), a fragment of right-handed double helical A-DNA. Nature 295, 294-299

2a. Z. Shakked, D. Rabinovich, W.B.T. Cruse, E. Egert, O. Kennard, G. Sala, S.A. Salisbury and M.A. Viswamitra (1981), Crystalline A-DNA: The X-ray analysis of the fragement d (G-G-T-A-T-A-C-C). Proc. Roy. Soc. London Ser. B. 231, 479-487).

3. S.B. Zimmerman and B.H. Pheiffer (1979), A direct demonstration that the ethanol-induced transition of DNA is between the A and B forms: an X-ray diffraction study. J. Mol. Biol. 135, 1023-1027.

4. A.H.J: Wang, S. Fuji, J.H. van Boom and A. Rich (1982), Molecular structure of theoctamer d (GGCCGGCC): modified A-DNA. Proc. Nat. Acad. Sci. USA 79, 3968-3972.

5. Z. Shakked and O. Kennard (1983), The A-form of DNA, in A. McPherson and F. Jurnak, Eds. Structural Biology, John Wiley & Sons, New York.

6. S.R. Holbrook, J.L. Sussman and S.H. Kim (1981), Absence of correlation between base-pair sequence and RNA conformation. Science 212, 1275-1277

7. Scientific American, (July 1993), How to construct a Sequence Space, Page 34

8. Walter Preiss (1977), Angewandte Graphik, Verlag K. Thiemig, München, 11-12

9. Prof. A. Karadeniz (1985) Hohe Mathematik, Band 2, Verlag Caglayan-Istanbul, 81-82

10. Prof. W. Schweizer (1968) Lambacher Algebra 2, 4. Ed., Klett Verlag, Stuttgart

11. Prof. A. Karadeniz (1983) Hohe Mathematik, Band 3, 3. Ed. Verlag Caglayan-Istanbul, 117-118.

12. J. Chotai Department of Preventive and Social Medicine, University of Umea, Journal: On the lod score method in linkage analysis, 364-366.

13. Journal: Milliyet, 23 April 1986