Now we have the complete blueprint of human life which may enable us to lead a life of one hundred years' youth!

Human Genome Project : Dreams and Dollars

J P Chaudhuri

The most spectacular achievement of this decade is the completion of human genome project. On June 26, 2000, scientists announced the completion of a "working draft" of the human genome at press conference in Washington, USA and London, UK. This means that almost every gene of an individual is sequenced. Theoretically, we now know the complete blueprint of human life, which may enable us to enjoy 100 years of youth, to conquer Alzheimer (dismentia), cancer and AIDS, to get embryos repaired by genetic engineering, or to fight famines and kill mosquitoes of Calcutta (Kolkata) and New York applying genetics and biotechnology. Briefly we have reached the door to the *Brave New World.*

Genes and Genome: All genes of a cell or of an individual are collectively known as genome. Each diploid organism like us has got two sets of genomes — paternal and maternal. Genes, the genetic units, are functional units of DNA (Deoxyribo-Nucleic-Acid). The most apt definition of gene is "one gene one enzyme", which means one gene codes for one enzyme or protein. Proteins, enzymes or non-enzymes, may be regarded as functional and structural units of an organism. Enzymes and proteins are polypeptide chains comprising amino acids of a definite order, dictated ultimately by the DNA sequence.

DNA is a polymeric chain, whose unit monomers are called nucleotides, each of which are

Dr J P Chaudhuri, Tumour Cytogenetics Lab, Kinderpoliklinik, LMU, Pettenkofer Str. 8a, Rm 092, D 80336 Munich, FRG
email : Jyoti.chaudhuri@lrz.uri-munchen.de

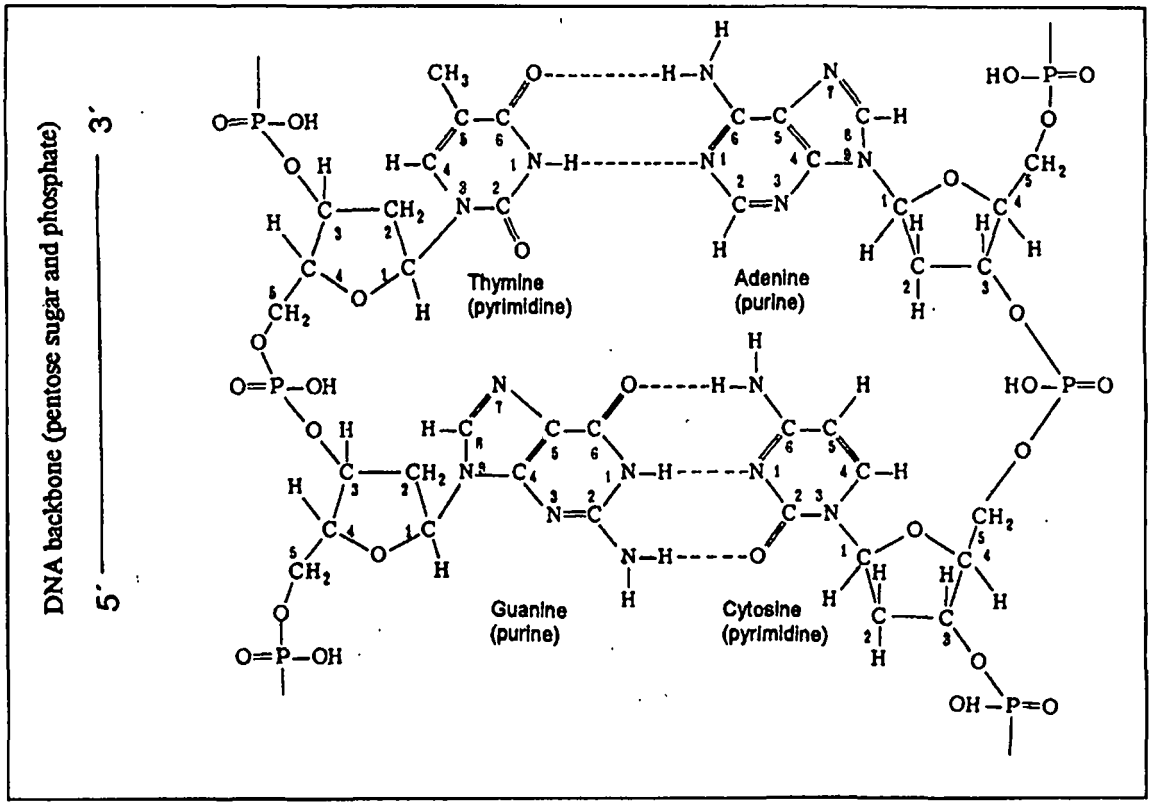


Fig 1. *The backbone of DNA consists of phosphoric acids and desoxyribose sugars. Each sugar carries a base, a purine (adenine or guanine) or a pyrimidine (cytosine or thymine), as a side chain. Complementary DNA strands are held together by specific base pairings — adenine to thymine and cytosine to guanine. These antiparallel strands twine around each other to form a right handed helix which makes one complete revolution with every ten base pairs.*

comprised one phosphoric acid and one desoxyribose (a pentose sugar) with a side chain, consisting of one of the four different bases called Adenine, Cytosine, Guanine and Thymine (Fig 1). When Watson and Crick (1951) discovered that the DNA has a defined sequence of these four bases, it struck Gamov, a mathematician, that three of these four bases A,C, G and T at a certain order may code for an amino acid. The different combinations of three out of four bases make 64 possible combinations which can cover the 21 amino acids known in biology. Soon thereafter, the concept of base triplets, coding amino acids, could be finally established by Khorana and coworkers (1964).

DNA double helix : To attain the necessary stability, each of the long DNA strands is supported by a complementary strand. Two strands hold each other by building base pairing : adenine binds to thymine with the hydrogen

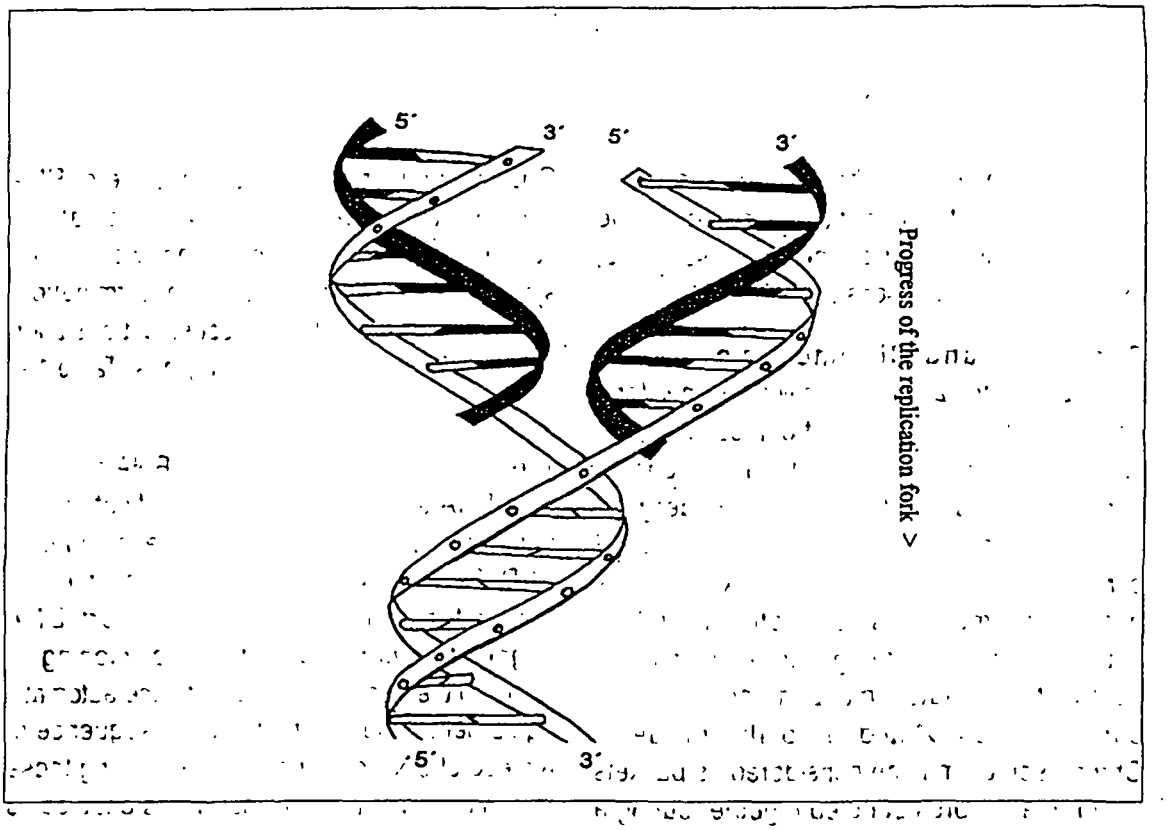


Fig 2. *Replication takes place following the strand of a DNA double helix. Each of the single strands serves as a matrix for the synthesis of a replicate resulting into two double helices. In this way, the quantity of DNA is doubled before a cell may undergo a miotic division.*

bonds, while guanine binds to cytosine with three hydrogen bonds. The bases adenine and guanine are purines, the larger molecules, while cytosine and thymine are pyrimidines, the smaller molecules. Base pairing in a double stranded DNA is restricted to A ≡ T and C ≡ G only, because of the bonding specificity. At the same time this kind of base pairing holds the two complementary strands equidistant from each other, because purine-pyrimidine and pyrimidine-purine have the same dimension. Two strands of DNA twine around each other to form a right-handed helix. The DNA double helix may dissociate and reassociate under physiological conditions enzymatically. Such strand separation and subsequent pairing may also be induced *in vitro* thermally.

When DNA is reproduced the strands of a double helix separate and each of the single strands serve as a matrix for the construction of a replicate (Fig 2). In this way, after one replication cycle, out of one DNA double helix two double helices are produced. One of the

two strands of a double helix, called codogenic strand, carries the genetic information. The other strand is code-wise meaningless and is, therefore, called anti-sense strand.

Chromatin and Chromosome : DNA associated with basic proteins like histones form chromatin, the major content of a cell-nucleus. When the cell prepares itself for a division, the chromatin mass of the nucleus undergoes condensation forming a definite number of chromosomes. A human cell, for example, has got 46 chromosomes, 23 maternal and 23 paternal. Strictly speaking, each of the parents contributes 22 autosomal and one of the sex chromosomes *X* and *Y* to the progeny. Chromosomes may be considered as packets of chromatin fibres bundled together during a cell division so that an efficient and equitable distribution of the genetic information to the two daughter cells is guaranteed.

Gene Expression : The genetic information of a gene is expressed following two essential steps : transcription and translation. The base sequence of a gene is transcribed from the corresponding section of a DNA producing a RNA chain. The base sequence of the RNA chain is then translated into the sequence of amino acids for the targeted protein of enzyme.

RNA molecules are also polymers of nucleotides like DNA, but they are much shorter, and therefore, stable though single stranded. The pentose sugar of RNA is a ribose, which holds one of the four bases-Adenine, Cytosine, Guanine and Uridine. The sequence of RNA with its base triplets have 64 combinations covering the 21 amino acids and two to three signs of punctuation like "stop" for termination, which are important for a successful translation from RNA to proteins or enzymes (Table 1—opposite page).

The reverse path, from protein to RNA and then from RNA to DNA, led to the identification of a number of genes. With modern techniques of PCR (Polymerase Chain Reaction) in a thermo-cycler for quick synthesis of short DNA segments, improved methods of cloning of genes in the form of plasmids, etc, the automatic equipment to determine the base sequence of a piece of DNA or gene and assembling these information with powerful softwares enabled us to decipher the base sequence of a complete genetic set, the genome, of an individual. Moreover, earlier analysis of the genome sequences of baker's yeast *S. cerevisiae*, the nematode *C. elegance* and fruit fly *D. melanogaster* paved the way for human genome project. A major factor behind this spectacular speed of success was indeed the fact that from the very beginning every bit of information was placed in a databank open to all through digital internet connections.

Genetics in Focus : Genetics played a central role not only for our perennial endeavour for bigger, better and more bountiful products of agriculture and animal husbandry, but also to improve the quality of human being, say, by eugenics and by conquering the diseases. Every

	U	C	A	G	
U	UUU Phe	UUU Tyr	UAU	UGU Cys	U
	UUC	UCC Ser	UAC	UGC	C
	UUA Leu	UCA	UAA Stop	UGA Stop	A
	UUG	UCG	UAG Sec	UGG Trp	G
C	CUU	CCU	CAU His	CGU	U
	CUC Leu	CCC Pro	CAC	CGC Arg	C
	CUA	CCA	CAA Gln	CGA	A
	CUG	CCG	CAG	CGG	G
A	AUU	ACU	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC	AGC	C
	AUA	ACA	AAA Lys	AGA Arg	A
	AUG Met	ACG	AAG	AGG	G
G	GUU	UCA	UGA Asp	UUA	U
	GUC Val	UCC Ala	UGC	UUC Gly	C
	GUA	UCG	UGG Glu	UUG	A
	GUG	UCU	UGU	UUU	G

Table 1. *The base triplets of RNA coding the amino acids are the basis of translation of RNA into the corresponding protein.*

aspect of all living beings has got some genetic component, which stimulates us to learn more about genetic mechanisms and to harness them. In a sense, genetics holds the key to both diagnosis and therapy.

An important side effect of genome research may benefit the Informatics. Genetic code on DNA may soon serve as a superior means for saving large data volumes and for their efficient operations. Another industrial sector with great future is the DNA-based creation of biosensors.

Unicate, Duplicate and Clones : The complete genome or gene constellation and the base sequence of each of the genes of an individual are unique, so that every individual member has to be considered as unicate. The mono-zygotic twins, being identical, are duplicates. Using the techniques of cloning many members may be produced with identical genetic constitution resulting into clones. Utility of clones lies in the multiplication of individuals with an especially advantageous genetic constitution : for example, 64 cloned pilots for 64 Starfighters. Disadvantage of cloning, ie, asexual reproduction, is that the cloned individuals are generally unable to withstand evolutionary pressure like changes in the biological and/or physical environment. Organisms without sex are often doomed to rapid extinction not only because of accumulation of deleterious mutations but also because this process will continue ever-increasingly.

PCR and Forensics : A minute speck of DNA extracted from blood clots, saliva or ejaculate may be amplified quickly using Polymerase Chain Reaction (PCR) producing a good amount of DNA. By determining the so-called "DNA fingerprint" out of such samples, the persons involved in the crime may be identified. Now that we know the complete genomic sequence, criminal identification will be more efficient.

Genetic Diagnosis : A number of hereditary diseases, cancer as well as effects of environmental toxins could so far be traced by studying the structural and numerical status of the chromosomes. Using Southern, Northern and Western Blots, genetic analyses could be refined using DNA, RNA or the proteins, respectively. Further development was attained by amplifying the sequence of a part of a gene and defining the base sequence by means of an automatic sequencer. A parallel development is the so-called FISH (Fluorescent *in situ* Hybridisation) whereby a fluorophore tagged piece of known DNA as probe and the target DNA in a cell nucleus or in metaphase chromosomes are exposed to each other inducing strand separation followed by the pairing of the single stranded DNA again generating a fluorescent signal where the probe hybridises with the target DANN (Fig 3 at page 25). Based on this simple principle, a number of sophisticated techniques have been developed which enhanced the speed and accuracy of genetic diagnosis enormously. So we have the so-called DNA chip or matrix : on a small area of microscopic slide up to 1000 dots of DNA sequences, each specific for a gene, may be immobilised to serve as a composite

probe on to which a test DNA and a reference DNA with different labelling may be applied for hybridisation. In this way, the copy number of genes — none, one or more — may be determined with speed and accuracy. Similarly, chips or arrays have been developed based on DNA-RNA hybridisation to detect and quantify gene expression (Fig 4 at page 25). The 24-colour-karyotyping based on FISH also allows a quick and sure analysis of both structural and numerical changes of chromosomes. Our present knowledge of the complete sequence of human genome will further expedite the scanning of genetic constitutions.

Guided by the human genome directory and choosing from these protocols, pre-implantation diagnosis has already been so rationalised that mothers may rest assured to receive only the genetically healthy embryos.

Gene Therapy : A detected loss, duplication or other abnormalities, like changes in the base sequence, of a gene or genes causing an ailment, may then be treated by a genetic compensation. If a gene is lost or found defective, an ersatz gene may be supplied to the affected tissue. A duplication or multiplication (genetically known as amplification) of a gene may possibly be countered by applying adequate anti-sense sequences of DNA or their products at RNA or polypeptide level. Various routes and vehicles are being tried to target the genes or their products at the site affected. Generally, the benefits of the genome project are likely to come from pharmacogenetics. More rational drug designs are expected following genetic identification of cellular processes associated with a disease. Misuse of this aspect may also spread for earning quick dollars.

Genetic Engineering and Biotechnology : So far the diabetic patients were satisfied with insulin extracted from animal source. Today human insulin gene, obtained from healthy human cells and amplified *in vitro* using PCR or plasmids, may be introduced into bacteria like *E. coli*, which then grows in a bio-reactor under controlled conditions producing human insulin in bulk. Another example of gene- and bio-technology is the way to manage phenylketonuria (PKU) patients. They lack an enzyme to metabolise the amino acid phenylalanine. Restricting phenylalanine in the diet is helpful for the PKU patients. An adequately prepared human lactalbumin gene may be introduced into fertilised bovine eggs giving rise to a so-called *transgene* cow. Such a cow produces milk poor in phenylalanine and provides an ideal diet for the PKU patients. Now that every gene of human genome is known, theoretically any individual may be helped efficiently by preventive correction of the diseased genes and by augmenting other genes for any required skill or proficiency.

Genome Transparency and the Future : Before recruiting a person for a job, the necessary gene data may be demanded by an employer. Similarly, health and life insurance policies may also be restricted to keep out persons with higher risks. One day the Big Brother may also start to run data bank of the genes of citizens and try, for example, to prune

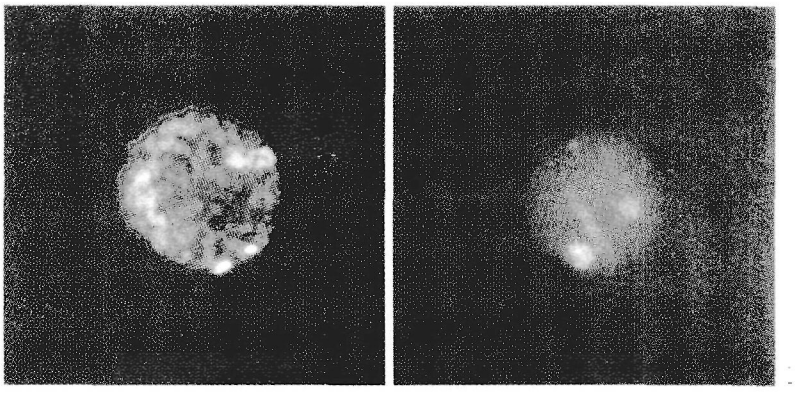
the unproductive persons like poets, philosophers and philanthropists. All these are still dreams of a *Brave New World.* We tend to overemphasise the genes. Every person is a product of three equally important factors : Genetics (genes and their function), Environment (including diet and medicine) and Training (physical and mental). Without proper training and food, no performance can be achieved, say by a soccer player, in spite of the best suited genetic constitution. Similarly, hormonal imbalance may induce sex reversal or wrong sex in spite of right chromosomes

Moreover, we have to learn a lot also in the context of the gene therapy Simple supplementation of gene failed often in want of a suitable route or vehicle to introduce the gene and to guide it to the target The total number of genes in our genome is still being debated. Present estimates range from 35000 to 150000 depending on the way of defining and recognising genes We may know the base sequence of almost 99 per cent of the genes of human genome, but we know little about their regulation. Chromosomal mutations like inversions may cause headache to the geneticists, because, being inverted, these sequences may preferably be accommodated in the anti-sense strand Likewise, the organisation of the genomes within a cell nucleus is still riddling the scientists Genes on each chromosome have a linear, ie sequential, order But we have 23 chromosomes in each genome Are not the genes on different chromosomes parallel to each other? Once sequential and once parallel — what a paradox! It is still a long way, till the Big Brother may claim to have all the strings of genetics in his hands. But before that, many dreams will be fulfilled and lots of dollars will be earned.

Addenda from the author dated 13 February 2001

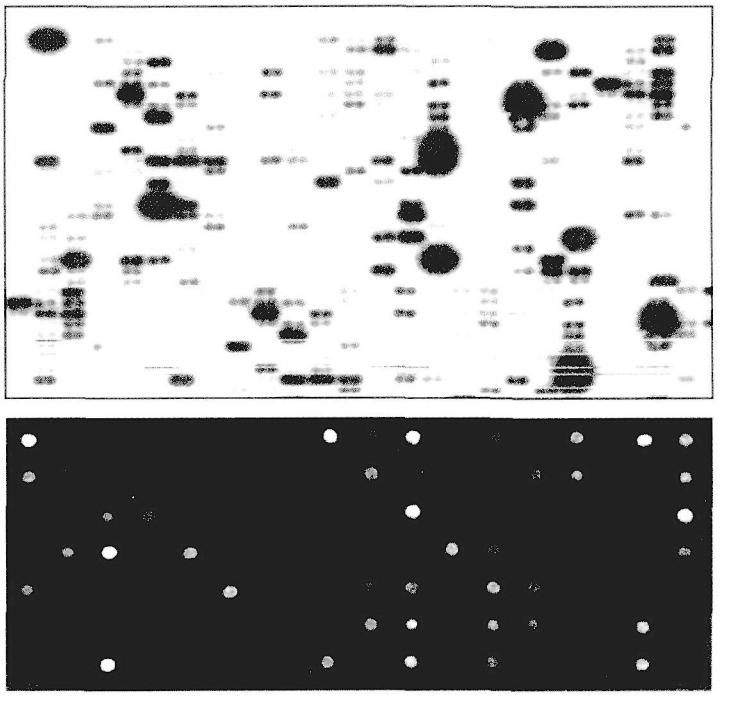
Today, the 13th February, 2001, from Berlin, the revised human genome chart has been released by Mrs Edelgard Bulmahn, the Minister for Science, Education and Research of the German Federal Government The total number of genes (ie, gene loci are now estimated to be 30 to 40 thousand only, and not 1000000 as had been calculated by some geneticists earlier It is interesting to know that one Indian scientist, Chakravarty by name, who engineered gene-manipulated E coli *bacteria in his laboratory in USA in late seventies, once calculated the human gene number and came to a value of 40000 as has been mentioned in an essay in the journal* Science *last year*

Fig. 3



FISH (Fluorescent in situ *Hybridisation) on the cell nuclei allows detection of chromosomal entities also at interphase, when chromosomes are otherwise not discernable. Left : A normal nucleus with chromosome 7 disomy (whole chromosome painting, green) and chromosomes X and Y, showing red and green signals respectively. (Right) An abnormal nucleus with monosomy 7 (the smaller signals represent chromosome 7 centromere, green and subtelmere, red) and chromosomes X and Y (the larger signals, red and green respectively).*

Fig. 4



4 DNA Chips (Atlas Array) : *Up to a thousand gene specific sequences of single stranded DNA pieces are immobilised on a microscopic slide or on a nylon membrane. RNA from test and/or reference tissues, labelled with radioisotopes of fluorophores, following hybridisation with the DNA-dots arrayed on the chip, allow us to determine gene expression. Pseudocolours are assigned to reflect the fluorescence intensity.*

(By courtesy of CLONTECH Lab Inc., USA)