Hargobind Khorana

In just 150 years, the field of genetics has risen from nothing at all to an area that is central in managing health, food grain and ecology. Charles Darwin's Origin of Species, published in 1859, was the first scientific examination of the mystery of the inheritance, and now, the wonder has been laid bare, with the documenting of the human genome. The human genome is the blueprint of the human genetic heritage, billions of units of information stored within the structure of a giant molecule found in every living cell. The detective story had many personalities and Prof Hargobind Khurana, Nobel laureate, was one of the most important ones.

Before 1859, the species were considered unchanging and fixed, to stay as 'God had created them'. Charles Darwin's' discovery, that species evolved, driven by natural selection, was based on fossil records and geographical distribution of species. The theory said changes in conditions of climate, vegetation, food supply resulted in particular variations in a species having an advantage for survival, and over centuries the species changed to these variants. The theory was a strong theological challenge but it made no suggestion of how the variations might arise.

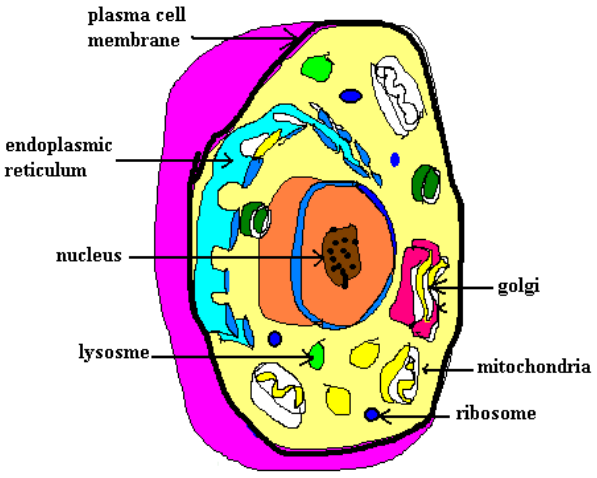
In the meantime, Gregor Mendel, an Augustinian monk and teacher of high school students in Austria, discovered important features of heredity. With painstaking records of the results of cross-breeding pea plants, Mendel found that the traits passed on by parents to offspring were passed in 'factors' or units – one for each specific trait and with equal contribution by each parent. These hereditary factors did not combine, but were passed intact; each parent transmitting half of its hereditary factors to each offspring. Certain factors were "dominant" over others and different offspring of the same parents received different sets of hereditary factors.

Mendel's work, published in 1966, had shown that heredity passed through discrete units, which we now call genes. In later years other scientists refined the theory of how genes were inherited, but the nature of the genetic material was still undiscovered.

The discovery of nuclein

At about the same time, in 1868, a Swiss biologist called Friedrich Miescher analysed the nuclei of human white cells and found that they contained a phosphorus containing substance, which he called nuclein.

The nucleus is a separate compartment in the cells of all animals, plants, fungi



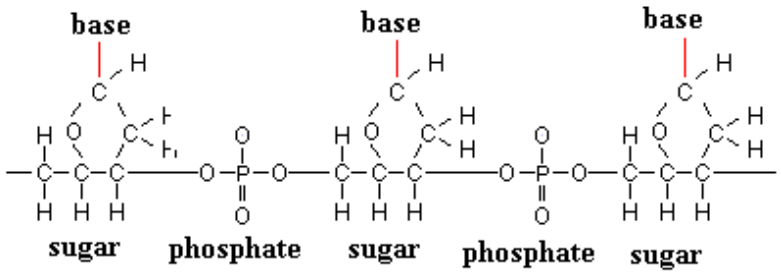
He found that nuclein consisted of an acidic part and a protein part. We now know that that the acid part was DNA and the protein part helps package the DNA. But at the time, that nuclein or nucleic acid may play a part in cell inheritance was just a suspicion, almost dismissed in the face of the seeming lack of chemical diversity in nucleic acids.

It was only in 1943 that it was shown that it was the nucleic acid in the cell that carried genetic information. Avery and others at the Rockefeller Institute showed that injecting the nucleic acid from one strain of bacteria into another changed the recipient bacterium into the donor strain. It appeared that the genetic component in the material injected was taken up by and became part of the recipient. The study of genetics now made a transition from a 'describing and classifying' subject to an analytical science. The field of molecular biology was born. In 1952, an experiment using a radioactive marker to identify individual atoms showed that it was indeed the nucleic acid portion, and not the protein coat, of a virus, which entered the host cell and provided the genetic information for replication of the virus inside the host.

This was progress, indeed. From Mendel's first discovery that there were 'factors' of each parent that participated in the traits of offspring, a number of experiments now showed that it was the nucleic acid that carried the genetic information in all living cells. But how in fact this information was carried and expressed was still a great, unanswered question.

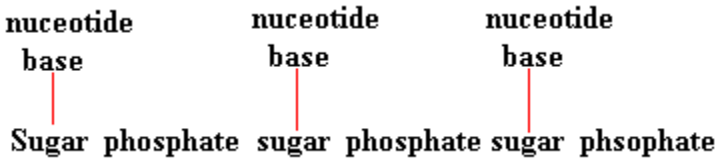
It was a really difficult question, at the time. How did the contents of one egg cell and one sperm cell produce a whole human being, different from any other? How did the millions of cells in his or her body know just what kind of cell they were and what proteins to produce? How could the information to tell a kidney cell it was a kidney cell, or a brain cell what cell it was, be stuffed into the nucleus of the cell?

The world of scientists, of course was hard at work and all kinds of studies were on, about the structure and properties of deoxyribonucleic acid, or DNA, which was this carrier of genetic information. One bit of information was that the nucleic acid molecule is a long 'backbone' attached to groups called 'bases' or 'nucleotide bases'.

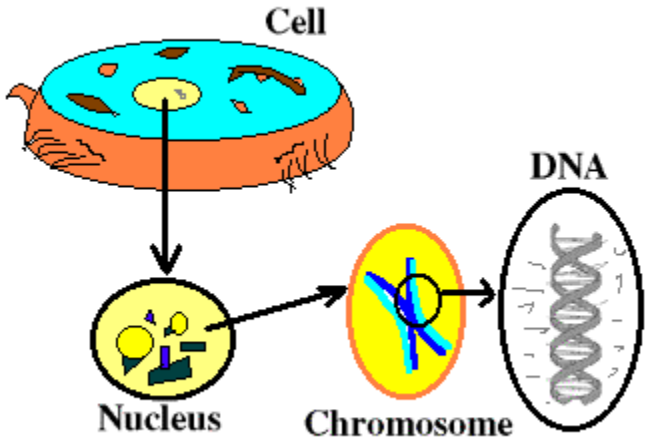


The backbone is a regular succession of sugar molecules, separated by 'phosphate' groups, or atoms of phosphorus and oxygen. The sugar molecules, of course, are hydrocarbons. The structure is like in the picture.

A convenient way of visualizing the nucleic acid molecule is like this”



The DNA molecule is built up in this way and has millions of units. Each mammoth DNA molecule forms one, threadlike ‘chromosome’ and there are 23 pairs of chromosomes in the nucleus of each human cell. The word, ‘chromosome’ came from the fact that the material in the nucleus could take a coloured stain, for observation, and had been named ‘chromatin’. DNA is ‘deoxyribonucleic acid’, from ‘deoxyribose’, the sugar molecule found in the backbone.



The next bit of information was that the bases, to which each nucleic acid molecule was attached, took only four forms in all the samples of DNA. These forms were adenine (A), guanine (G), thymine (T) and cytosine (C). But these four forms could be present in different proportions in the DNA of different organisms. Now an important observation made, in the late 1940s, was that the quantities of A and T or G and C always turned out to be equal, as if whenever there was an A there would be a T and whenever there was a G, there would be a C. This observation was an important clue to the structure of the DNA and later also in its function.

X ray studies

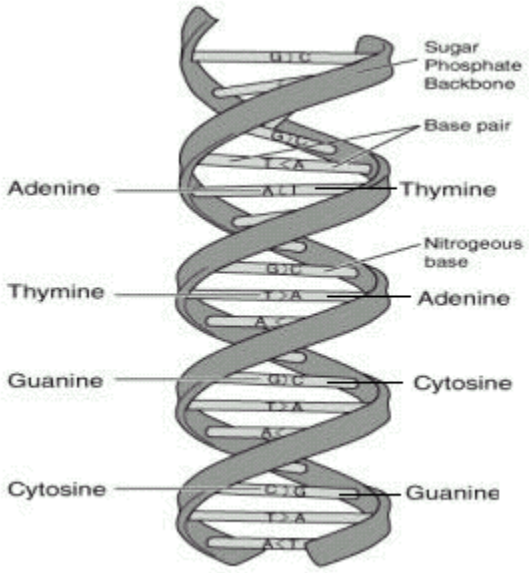
The last clue that enabled getting at the structure of DNA came from the X-Rays studies of Franklin and Wilkins in England. The use of X rays that most people are familiar with is their use to cast shadows of bones, to detect fractures. But X Rays are also useful in science because the X Rays consist of waves whose dimensions match the distance

between atoms in crystals and molecules. Regular features of molecules are then revealed by patterns in which X Rays passing through the molecules are scattered.

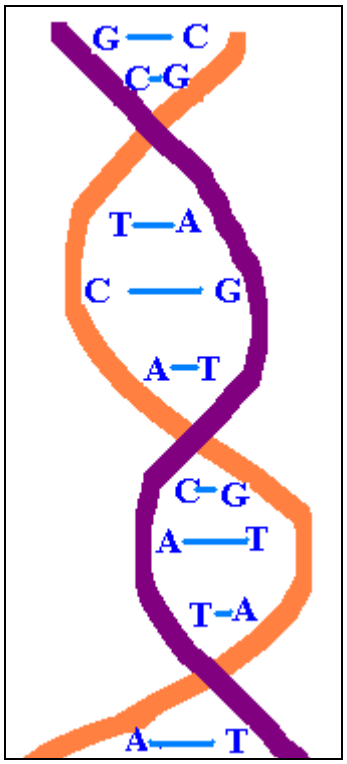
The work of Franklin and Wilkins soon showed that the structure of DNA displayed a 'periodic' nature along its length, one period repeating every 0.34 nm and a secondary one every 3.4 nm (a nanometer, nm, is a millionth of a millimetre).

Watson and Crick

Watson and Crick pieced all these clues together, of the repeating sugar and phosphate, the A, G, T, C appearing with equal quantities of A and T or G and C, and the periodicities along the length. And with a method rather like putting together a three dimensional jigsaw puzzle, they proposed the celebrated 'double helix' model, of two nucleic acid chains intertwined and connected by links of A and T or G and C (which accounted for there being an A for every T and a G for every C).



This was a stupendous breakthrough. The two strands being connected by complementary pairs of A-T or G-C makes sure that if the two strands were separated, then each half could build up the other half by adding exactly an A for every T and a G for every C, and vice versa! This makes possible the creation of exact duplicates of DNA, or self-replication, using the two halves as templates. The DNA in a cell of an organism could then divide, with one half of the DNA going in one half of the cell and the other half of the DNA in the other half of the cell and then generate two complete, identical cells. Watson and Crick shared the 1962 Nobel Prize for the discovery.



Har Gobind Khurana, born in 1922, in an obscure village in Punjab, in the then undivided India, was an unlikely candidate for center-stage in the drama that was unfolding. He was the youngest of the five children of the village 'patwari', or clerk who kept land records at the lowest end of Indian land tenure system. The village itself had a population of about a hundred and Khurana's father was poor. But he had regard for education and the children were sent to school.

Khurana attended the DAV school in Multan and later went to the Punjab University at Lahore, where he took his M.Sc. He was fortunate in gifted teachers both at school and in college and in 1945, Khurana was awarded a Government of India scholarship to go to England for his Ph.D. He took his Ph.D. from the University of Liverpool in 1945 and spent the next year with Professor Vladamir Prelog in Zurich. This was a period when Khurana's thought and philosophy towards science, research and work were strongly influenced and moulded in the European tradition.

After a brief visit to India in 1949, Khurana went back to Cambridge, till 1952. These were three crucial years, when Khurana became interested in proteins and nucleic acid. Khurana moved on to work at the British Columbia Research Council, in Vancouver and then to the Institute for Enzyme Research at the University of Wisconsin. All through, his interest was in the area of proteins, nucleic acids and DNA, the exciting developments that raced towards unraveling the mystery of heridity!

When Crick and Watson had proposed the double helix model in 1953, the framework of transmitting genetic information had been found. Here was a stable mechanism which

seemed to have the necessary complexity, as also a method to split in two and yet be able to replicate the original from the separate halves. It was elegant and versatile, but yet, there was no clarity about the code itself, just how did the chain of A, T, G and C lead to the specific and diverse functions of the billions of cells in a living organism?

The answers to these questions came from the work of three, Holley, Khorana and Nirenberg. It had been found by then that the functions of organisms were controlled by a kind of protein called enzymes, agents that promote the growth of specific things. The colour of a person's eyes, for instance, is the result of a specific enzyme that the person's body produces, which in turn promotes the production of the dye that leads to the colour of the eyes. Whether a person is diabetic, again, is a result of chemical agents that determine the rate of production of insulin. And whether these agents are present and in what numbers, depends on the instructions contained in the cells that produce the agents. The individuality of a person, in short was determined by what proteins his body was programmed to produce.

Proteins, themselves, it had been found, were composed of just twenty building blocks, called amino acids. Amino acids are organic molecules, containing from ten to thirty atoms. Different combinations, sometimes short strings, sometimes thousands of units long, of these twenty components construct all the millions of proteins that are found. Of these twenty building blocks, only some can be synthesized by the body, the rest need to be ingested as food. But from this stock of twenty amino acids, the code in the DNA ensures that complicated proteins are constructed. How this is done is the way the code works, and to frame words in the code and create the proteins the words spelt was deciphering the code!

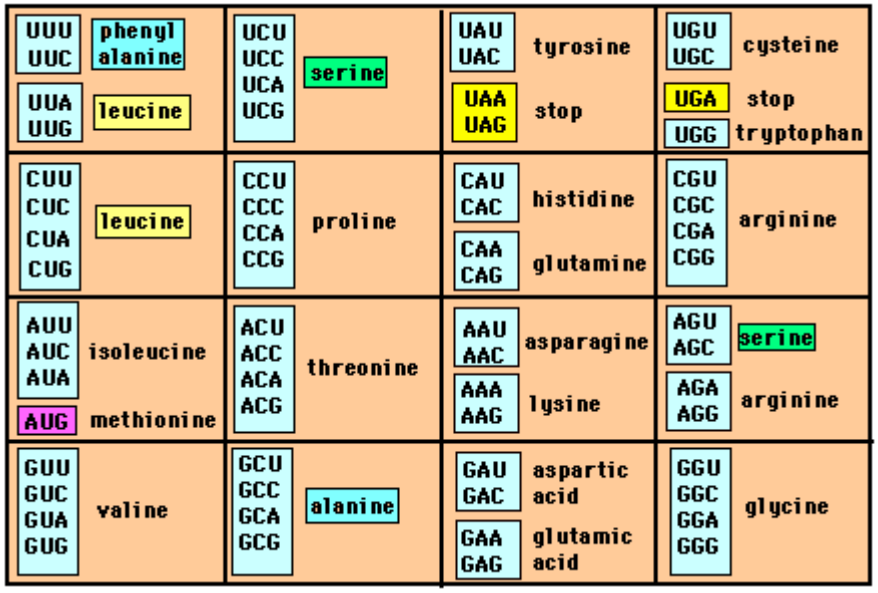
The way different proteins come about is by the sequences of the nucleotides, A, T, G, C, which dictate the amino acids to be included. Now, just four types of groups could individually code only for four amino acids. So, clearly, more than one nucleotide was involved. Two nucleotides could code for 4x4 = 16 amino acids, like this:

A-A	A-T	A-G	A-C	T-A	T-T	T-G	T-C	G-A	G-T	G-G	G-C	C-A	C-T	C-G	C-C
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

OR

1	A A	5	T A	9	G A	13	C A
2	A T	6	T T	10	G T	14	C T
3	A G	7	T G	11	G G	15	C G
4	A C	8	T C	12	G C	16	C C

This is still not enough, as we have twenty amino acids to code for. If we use three sets of nucleotide bases, then we code for 4x4x4 = 64 combinations, which is *more* than 20! Well, that is the way the code is constructed. In the deployment of the 64 codes, there are several different codes that mean the same amino acid. This seems to be nature's way of playing safe, that even if in the transmission of the code, there are some errors, the correct amino acid is still conveyed. Here is the code:



(Thymine, ‘T’ converts to uracil, ‘U’ outside the DNA molecule)

Most amino acids thus have more than one way of being coded, the codes being different in the sequence of bases. This seems to have evolved as the most stable way to eliminate errors, given the level of ‘redundancy’ available.

So in this way, every consecutive group of three nucleotides, or every ‘triad’, codes for a particular amino acid. The groups of three are also called ‘codons’. The codon AUG, which codes for Methionine is the ‘start’ codon and we can see from the chart that there are three ‘stop’ codons. The stop codon is where the sequence of amino acids for one protein stops and the where the next sequence starts. The nucleotides thus provide a complete alphabet to specify all kinds of proteins, using all the twenty amino acids in varying combinations and numbers. Each sequence of codons from ‘start’ to ‘finish’ identifies one protein and is called a ‘gene’. And the whole DNA is the code for thousands of proteins. The complement of twenty three pairs of chromosomes in the human cell then identifies all the proteins that determine the complete design, as of the human species, as well as the individual heritage of every living person.

From here Nirenberg, Khorana and Holley went on to do their most important work, which was analysing how exactly the information in the DNA was tapped and used in the formation of proteins.

The entire DNA, it was found, did not participate at a time in the formation of proteins. Bits and pieces of the DNA, genes that represent proteins, were copied on to similar molecules called RNA and shipped out of the cell nucleus, possibly to protect the DNA from damage. RNA is just like DNA, except that there is a small difference in the sugar found in the backbone and also in one of the bases. The bit of the DNA is thus copied on to a form of RNA called ‘messenger’ or mRNA, in a process similar to DNA replication. The mRNA then attach to structures called ***ribosomes*** in the fluid outside the nucleus of the cell. The point at which the mRNA ‘attaches’ is where the ‘AUG’ codon, the ‘start’

codon is found. The RNA does not ‘know’ how to find its way about, it just attaches when, in feverish, random motion within the cell, the correct parts of the mRNA and the ribosome come into proximity, and they ‘dock’, in a position where they fit, like a lock and key. It is somewhat like a golf ball wandering about the course and falling into the hole when it moves over it.

When the mRNA is in position, like this, another class of RNA called ‘transfer’ or tRNA attaches individual amino acids to the succession of codons. The tRNA are specific to each amino acid and the first one to attach to the RNA is the one that has the complement of the ‘AUG’, the ‘start’ codon. The cell contains enzymes that are chemically the correct ‘fit’ to help particular tRNAs transfer amino acids to the mRNA. Once the first, ‘start’ amino acid has attached, the tRNA carrying the amino acid coded by the next codon moves in, to place the next amino acid in position. The two adjacent amino acids then form a chemical bond, the start of a chain of amino acids that goes on to become the protein.

The process of adding successive amino acid continues till the ‘stop’ codon is reached. Here, enzymes get active and the mRNA is released from the protein that has been synthesised.

This understanding of the sequence of events came in steps, following a series of experiments to deduce what was happening down at the invisible, molecular level. Nirenberg and Heinrich Matthaei first developed a way to create simple proteins in a test tube, using mRNA that had been extracted from cells. Khorana carried forward Niernberg’s work with the synthesis of RNA, which, in turn, could build proteins. Holley followed up with on the transport of mRNA and the synthesis.

The first RNA synthesised was a chain of only ‘U’s (the RNA equivalent of ‘T’ in DNA). This was found to code for a protein that was a chain of the amino acid, phenylalanine. We can see from the genetic code chart that the codon, ‘UUU’ codes for this amino acid. Repeating a chain of ‘CA’ in the RNA led to the histidin-threonine chain, and so on.

Individual components of the codons had to be identified through carefully designed experiments. For instance, for a particular element appearing in one of the groups, a radioactive form of the element could be used. Once this was done, a particular strand of DNA, when the DNA had been broken up, could be traced by the radioactive ‘trail’. Painstakingly, using methods like this, the meaning of each codon was worked out and the meaning of nature’s hidden code was pieced together. The concise and enduring formula, for individuals of a species to develop, as an expression of the proteins their cells produced, was discovered.

Khurana did the bulk of the work himself and finally completed the decoding, the identification of the amino acids that each of the sixty four codons represented. Along the way, Khorana developed techniques of locating, navigating and marking spots along the length of the DNA molecule, ways of cutting and spicing strands of DNA, the enzymes

that could do this, the way to work with them. These were the forceps and scalpels of dissecting and assembling the fundamental components of genetics.

The scientific community soon realised the importance of the work that had been done. The human genome, or the entire code of the twenty three chromosomes, consists of nearly three billion genes. A handful was identified as being involved in specific human ailments. Khorana's techniques gave rise to an industry of generating the specific scraps to help an individual lacking those specific genes to recover lost function. The community and industry immediately saw the immense potential of mapping the entire human genome. The task was obviously too large for individuals. Biophysicists, microbiologists, information technology experts, the world over, were drafted into the global Human Genome Project.

The Human Genome now stands mapped, both by a public group as well as by a private enterprise. Humankind now has the most basic blueprint of its nature. Human cells, in principle, can now be programmed to eliminate disease or the tendency to disease. There is the possibility of creating the super race, there are fears of misuse. Applications in law and crime detection are commonplace. There are fears of invasion of privacy.

The understanding of the mystery of heredity, hesitatingly probed 150 years ago, may be the most important part of knowledge in the twenty first century.

Khorana is today a naturalised citizen of the US. He is married to Esther Elizabeth Sibler and has three children. From 1970, he has been a professor in the Massachusetts Institute of Technology, where he has continued to make major contributions in the field of molecular biology. Along with Holley and Nirenberg, Khorana was awarded the Nobel Prize for medicine in 1968.

---x---