# *Music of Life*

*Development of Molecular Biology: A Personal Account*

*Dedicated to Dr. Debendra Mohan Bose*
*(1885 – 1975)*
*Director, Bose Institute, Calcutta*
*(1938 – 1967)*
*Who induced the author to carry out researches in Life Sciences*

# *Music of Life*

*Development of Molecular Biology: A Personal Account*

Debi Prosad Burma

Vigyan Prasar

Published by

**Vigyan Prasar**

(An autonomous organisation under the Department of Science & Technology, Government of India)

C-24, Qutab Institutional Area

New Delhi 110 016

(Regd. office : Technology Bhawan, New Delhi 110 016)

Phones : 26864157, 26864022 Fax : 26965986

E-mail : vigyan@hub.nic.in

Internet : http:// www.vigyanprasar.com



***Music of Life** : Development of Molecular Biology: A Personal Account*

*Author* : Debi Prosad Burma

*Overall Supervision* : Dr. Subodh Mahanti

*Typesetting & Pagemaking* : Ms. Sonu

*Production Supervision* : Sumita Sen

The cover page was designed by Shri Mrityunjoy Chatterjee, Thot Shop Communications Pvt. Ltd., Kolkata, free of charge for the sake of spread of scientific knowledge.

**ISBN : 81-7480-106-5**

**Rs. 150/-**

Printed in India by Saurabh Printers Pvt. Ltd., NOIDA

# *Music of Life*

*What is it?*
*Is it what we listen to every day?*
*Is it what we muse of, day-in and day-out*
*Do hate and enjoy?*
*Is it the same which was born*
*Billions of years ago?*
*Where and how?*
*Did the Almighty play the flute?*
*Or, was it born all at once with a big bang*
*As the Universe did?*
*Life and death are intermingled.*
*Creation and destruction joined hands*
*Worked ceaselessly till they gave birth*
*To life and its music*
*Life and death, inseparable as they are*
*Till death embraces life*
*But life continues even today*
*The music is well known*
*What is not known is its mysticism*
*No doubt it will be known someday*
*Before the life vanishes from the earth*
*Like a bubble*
*To be created again here or there*
*Or, somewhere in the universe, with no boundary.*

# *Contents*

# *List of Plates*

# *Acknowledgement for Reproduction of photograph, plates and Figures*

Photograph on inside cover page (courtesy of Dr. M Siddiqi, Director, Bose Institute)

Fig 1.2. Reproduction of Fig 1. from Nature, 171, 740-741 (1953). Reprinted by permission from *Nature* (reference citation) copyright (1953) MacMillan Publishers Ltd.

Fig 1.3 & 1.15(b) 2.12 & 2.13 Reproduction of Fig 3-12 & 3-14, 3-17(a),(b),(c) respectively from Biophysical Chemistry Part-1 *The Conformation of Biological Macromolecules* by Charles R. Cantor and Paul R. Schimmel published by W.H. Freeman and Company San Francisco (1980).

Fig 2.7, 2.8(a), 2.8(b), 2.9(c), 6.6 Reproduction of Fig 2.2, 2.6, 2.5, 4.4 & 14.7 from *Introduction to Protein Structure* by C.Branden & John Tooze published by Garland Publishing, Inc. NewYork and London, (1991) permission from Taylor & Francis Books, Inc., 29 West 35th Street, New York, NY 10000, U.S.A.

Fig 2.9(a) 2.9(b) Reproduction of Fig 3-2 & 4-1 respectively from *Biochemistry* by Lubert Stryer published by W.H. Freeman & Co. San Fransisco (1975) permission form Nancy Walker

Fig.2.14 Reproduction of Fig3.1 from *Structure, Function and Genetics of Ribosomes* Ed. by B. Hardesty and G.Kramer permission by Springer Verlag, New York and James Lake (Author)

Fig 4.11 Reproduction of the figure from the Web Site permission from V. Ramakrishnan (Author).

Fig 4.12 Reproduction of Fig 2 from *Science* 289, 905-920 (2000) Reprinted by permission form American Association for Advancement of Science, U.S.A. and Peter Moore (Author)

Fig 4.13 & 4.14 Reproduction of Fig 2 & 5 respectively from *Science* 285, 2095-2104 (1999) Reprinted by permission from American Association for Advancement of Science, U.S.A. and Harry Noller (Author).

Fig 2.14, 5.2 5.3, 5.8 5.9, 5.10 Reproduction of Figs. 8,2,4,5,6 & 7, of RNA world and Ribosomes *Current Science* 63, 550-560, (1992) by D.P. Burma Reprinted by permission from Current Science Association, Bangalore.

Fig 7.2 Reproduction of Fig 9 from *The Blind Watchmaker* by R. Dawkins published by W.W.Norton & Company, New York,London (1987,1986) Permission by the Author.

Fig 7.3 Reproduction of Fig on pages 160 & 161 of the *Origin of Species* of Charles Darwin (1979). The Crown Publishing Group, 1745 Broad Way, 13th Floor, New York – 10019, U.S.A. permission from Random House Inc. 1745 Broad Way, New York, NY – 10019, U.S.A.

# *Foreword*

The Publication Programme of Vigyan Prasar has taken shape in the last few years. Vigyan Prasar has brought out publications on a variety of topics of science and technology. Popular Science Classics, India's Scientific Heritage, Natural History, Health, and Do-It-Yourself are some of the series that evolved over the years. Our emphasis has been on bringing out quality publications on various aspects of science and technology at affordable prices. Further, Vigyan Prasar is putting in efforts to bring out publications in major Indian languages for various target groups.

The present book, *The Music of Life: Development of Molecular Biology—A Personal Account*, attempts to narrate the history of molecular biology as it took shape through the second half of the twentieth century following the discovery of the structure of DNA by Francis Crick and James D. Watson in 1953. The author, Professor D. P. Burma, has over 50 years of teaching and research experience in this field. He is one of those who helped molecular biology take roots in the country. The narration is sprinkled with the author's interactions with some of the scientists who gave rise to the field of Molecular Biology including Watson and Crick. Thus the account is both authentic and historical making the book highly readable. The book will be useful not only to those who are engaged in teaching and research in molecular biology but also to the new entrants in the field and the general readers.

New Delhi
October 13, 2004

**V. B. Kamble**
Director
Vigyan Prasar

# *Preface*

The book is basically the outcome of 50 years of teaching and research experiences of the author. Two very basic questions may arise in the mind of a reader before and even after going through its pages. Was there a need for this book at all? It is definitely not a textbook although it has all its elements, with similar information. However, its style can be easily distinguished from that of a textbook. If I am asked to summarise the contents of this enterprise it may be emphatically stated, 'This depicts the history of molecular biology as it developed through the second half of the 20th century following the discovery of the structure of DNA'. In spite of that it deviates from proper historical recording. That raises the second obvious question 'Who are expected to read this book?' Before answering that it must be pointed out that the hindrances a reader will face while going through the book are the digressions frequently made to record some of the incidents in which the author was directly or indirectly involved. The argument may be advanced that these incidents are not terribly important. But those definitely smell of some of the heroes and heroines involved in the drama of unraveling the mysteries of life at the molecular level. In a word, the style of story telling adopted here may have some impact on the budding scientists who normally do not have the opportunity of knowing those from text books and even from the proper historical records. Finally, doubt may be cast by raising the following issue 'Why will the readers be forced to learn about some of my personal experiences (rather encounters) when my own contributions in molecular biology are not that significant? I can defend myself to some extent by way of clarification that I was lucky enough to witness the development of molecular biology during the half-century. Though mostly from a distance, I felt the impact and did enjoy the excitement, which I want to share with the younger generation. They, in turn, may look forward to similar or more rewarding ecstasies in their career. The personal experiences, however, will be of limited appeal to the veterans who have to excuse the author if they decide to at least glance through the text.

Such an awkward creation would not have ever seen the light of the day without the spontaneous and arduous help of three persons, Dr. Subodh Mahanti of Vigyan Prasar, New Delhi, my wife Dr. Maharani Chakravorty, Senior Scientist of Indian National Science Academy working at the National Institute of Cholera and Enteric Diseases, Kolkata and our daughter-in-law Dr. Shweta Saran, Reader, Department of Zoology, University of Delhi, Delhi. Finally thanks are due to Prof. Debi Sarkar of the Department of Biochemistry, Delhi University, South Campus for thorough checking of the text and the index. The permissions of a number of publishers (as listed earlier) for reproduction of the illustrations from their publications are gratefully acknowledged.

**Debi P. Burma**
Pather Sheshey
CF 186, Salt Lake
Kolkata – 700 064

# *Prologue*

The present composition 'Music of life', is a personal account of the development of molecular biology based on the central dogma DNA → RNA → Protein as enunciated by Francis Crick. Naturally, the major effort has been made to illustrate the nature and functions of DNA, RNA and protein. Some features like DNA *versus* RNA world; Cancer and Evolution have been brought into picture as adjunct materials as they are related to the central dogma. The history of development as narrated here has some personal touches in its mode of presentation. Through this introductory part, it is being planned to focus on two or three major events of which the author himself was a participant. Some minor incidents that played important roles in the development of molecular biology in the country and abroad where the author himself was either an eyewitness or a direct or indirect participant would be dealt within the individual chapters to justify personal accounting.

There are no compelling reasons for me to remember a date almost half a century later, especially when it was a minor incident as far as I was concerned. There was nothing unusual for Dr. Debendra Mohan Bose to call me in his office as he would do that quite often. That particular date was after the publication of the seminal paper of the century in Life Sciences in *Nature*[1]. Anyone interested in genes and their functions know that it was in the month of April, 1953. Before I elaborate this particular incident it is naturally my duty to explain who D.M.Bose was. He was the then director of the Bose Institute at Calcutta founded by his maternal uncle Sir Jagdish Chandra Bose. Just like his uncle, D.M. Bose was also a physicist. Many Indians and some Westerners know J.C. Bose as one of the pioneers in

demonstrating the wireless transmission of message over a short distance through microwaves before Guglielmo Marconi perfected the technique to send messages across the Atlantic. J.C. Bose who had passed 'Tripos' (an honors course) from Cambridge University and obtained D.Sc. degree in Physics from London University, was also a Fellow of the Royal Society. He was well recognized for developing simple but ingenious devices to measure the movement in plants, record the rate of photosynthesis etc. Actually, he demonstrated that plants are very sensitive to environment but to a lesser extent when compared to animals although they do not possess a nervous system. In a nutshell, he was a pioneer in researches in physical as well as biological sciences in India. The Bose Institute (*Basu Vigyan Mandir*, in vernacular) was founded on November 30, 1917 at 93, Upper Circular Road (at Calcutta), which is now named after Acharya Prafulla Chandra Ray another Indian scientist, a chemist by profession who tried his best to temper Indian society with scientific attitude. The southern portion of this unusually long road in Calcutta is named Acharya Jagdish Chandra Bose Road. The two were contemporaries and good friends. P.C. Ray, a life long bachelor lived next door in one of the rooms of the University College of Science. It is befitting that Calcutta Corporation has renamed a single road in that fashion in memory of both. Bose Institute was founded under the blessings of Rabindra Nath Tagore, the only Nobel Laureate in literature from India. He was an intimate friend of J.C. Bose. Those were the golden days of Indian science, especially in the Eastern zone. The nephew Bose who became the Director of the Bose Institute after the demise of his uncle was also a physicist by training, specializing in magnetism and did some work on cosmic rays (practically missed the credit of being co-discoverer of cosmic rays). But later he mainly became interested in life sciences and especially in plant physiology. The reason was his blind faith in the achievements of his maternal uncle. There was a wrong notion in those days about J.C. Bose's discovery that plants have life. That the plants are living entities was,

however, an age-old belief beyond any controversy. Nephew Bose would encourage everyone in the Institute to follow the footsteps of his uncle and perform research to prove that he was right in all his predictions. I was also a victim of this notion. To avoid confusion prevalent among the scientists of the Western countries neither of the two Boses were related to S.N. Bose. Bose is a very common surname in Bengal, situated in the Eastern part of the country. S.N. is known all over the world for Bose-Einstein statistics and 'Bosons' are named after him.

But why did D.M. Bose call me to his office on that day? He showed me a short two-paged communication in an issue of Nature[1], which had just arrived. By the way, all the journals of the Institute Library had to reach his desk first. After they were scanned by the Director they would be sent to the Library. He was a voracious reader just as he was a benevolent dictator. He looked agitated and said 'It looks like that structure of the gene has been solved.' I was familiar with the word 'gene' but had hardly any idea about its nature as I had no exposure to biology either in my school days, or in college years or even during the University education. But I had to say something to the excited old man who thought that as a budding scientist I was supposed to know every thing, like him. So I said in a false tone of excitement 'Is that so?' The next dictum was rather alarming for me. 'Would you go through the paper and explain to me how did the authors solve the structure of the gene?' I had no other alternative but to borrow the issue of Nature and start reading it over and over again. The names of the authors, who were destined to be Nobel Laureates, were not familiar to me. I consulted my biologist friends to learn something about the gene but the borrowed knowledge was not enough to explain to the physicist the physical and chemical evidences for the structure of the gene and their biological implications. As an urchin I gathered all my courage to meet the Director the next day in his office and applied all my brain power to throw light on the miraculous achievement of the authors, without losing any nerve. He listened

very attentively and finally said nodding 'No *Bapu* (in *Bengali*, affectionate word for son), I don't understand a single word of yours'. The best recourse for me was to leave the journal on his desk and quit the giant's office and breathe fresh air outside.

After almost a decade, I had the opportunity of meeting Francis Crick and that too in India. Before that I saw (*not met*) Watson in U.S.A. when I was doing postdoctoral research in the laboratory of Bernie Horecker, a world-renowned enzyme chemist. At that time Watson visited the National Institutes of Health in Bethesda. I was not at all impressed by his looks. Later when the 'Double Helix'[2], one of the best sellers of the century was published, I realized that the picture looked better than him in person. The historical record, in a sense a novel, was a fantastic writing in the most scientific and popular way about the seminal discovery of biology in the twentieth century. It is interesting to scan the picture of the duo (Crick and Watson) looking at the model of double helical DNA built by them. Crick appeared to be somewhat more helical than Watson in the picture but not in reality and this was not due to my defect in visual perception. I understand that people who knew him closely were not usually impressed by his appearance. Let us forget about Watson and go back to Crick for a while. My intimate friend Pushpa Bhargava, the Founder-Director of the Centre for Cellular and Molecular Biology at Hyderabad organized a symposium on 'Nucleic Acids' (Structure, biosynthesis and function) during January 16-22 in 1964)[3]. It was held at the Regional Research Laboratory (renamed later as Institute of Chemical Technology) of the Council of Scientific and Industrial Research and located at Hyderabad. Pushpa was heading Biochemistry Division of RRL at that time. It was the first ever International symposium on nucleic acids held in India. While in England Pushpa came in close contact with Crick. Other than Francis Crick, Seymour Benzer, Gerhard Schramm, A.E. Mirsky, H.G. Zachau, Uriel Littauer, H.G. Wittmann, Alexander Rich and many others came from abroad. We, the young ones from India tried to match with them in performance

but miserably failed. Nevertheless it was a highly educative interaction for us throughout the week. But what I enjoyed most was a solo drama enacted by Crick. Bhargava organized some popular lectures in science and so far I remember Crick's was the first one. All these lectures were delivered in the beautiful Rabindra Bharati auditorium. Such public halls were built in large cities of the country in memory of Rabindra Nath Tagore, the Nobel Laureate poet of India which are usually spacious and picturesque. The hall was overflowing with people. A Nobel Laureate was speaking on `gene' which is nothing but DNA and the backbone of life. When Crick appeared on the stage with his floating hair at the back and dazzling eyes like that of a dramatist he reminded me of Linus Pauling. Therefore, I am tempted to digress a little again.

The Duo (specially Watson) while building the double helical structure of DNA were worried that Pauling might beat them in the race. However, the tension was unfounded. By the way Pauling will be ever remembered for receiving the Nobel Prize in two different areas, one for Chemistry and another for Peace. The last award had a justification. The creation of the Prizes for scientific achievements by Alfred Nobel was based on his earnings due to the manufacture of the dynamite, a weapon used in the disturbance of world peace in spite of its numerous beneficial applications. Therefore, he created a separate prize for those who worked to create peace on earth. Pauling was one of them. There will be occasions later when Pauling's contributions in chemistry and medicine, which revolutionized biology would be described. I heard the lecture of this giant at the National Institutes of Health (NIH) auditorium in Bethesda. He also enacted a drama and as usual, I was enchanted. A senior scientist from NIH with quite a fame was sitting by my side. Before the lecture started he whispered in my ears 'There will be no new substance in his lecture. If you have read his 'Harvey Lecture' you will realize that he will repeat the same performance.' However, I did enjoy it as I was hearing

him for the first time. I went through the Harvey Lecture later and truly realized that the NIH scientist was right. That, however, did not make me lose my respect for the double Nobel Prize winner. Similarly, I had no chance to listen to Crick earlier but thoroughly enjoyed the drama. He acted like a hero in solo acting and delivered such a marvelous lecture that the large audience listened dumb founded. The lecture was naturally a popular one and meant for tempering the society with science. I forgot the actual title but distinctly remember, that the topic was DNA, the genetic material. By that time I had thoroughly learnt about genes and was fully knowledgeable that those were made of deoxyribonucleic acid (DNA in short). DNA and life are inseparable in the living world of today which was created 4 billion years ago. Crick's words still ring in my ears but I specially remember the concluding part of it. The DNA in a single cell of a human being is about 2 meters long. If all the DNA molecules contained in an individual are joined end to end to make a thread then the string will be long enough to cover the distance between the earth and the moon several thousand times. And if one tries to put that amount of DNA together to make a sphere that will be the size of a cricket ball. There could have been no better way to make people understand that DNA is the best example of a straight line, no breadth but only length. Such scientists are needed not only to unfold the mystery of nature but also to enthuse the society with science and the scientific spirit.

Bhargava organized a dinner in honour of Crick where the dinner tables were arranged in a double-helical fashion. Naturally, Crick appreciated that. However, his scientific lecture during the Symposium[3] was not very impressive. He talked on the 'Colinearity and interlacing of the genetic code'. Although the code was broken by that time the details were not yet worked out. He more or less ruled out the noncolinearity or interlacing of codons. Seymour Benzer, the famous *Drosophila* geneticist who chaired the session used two slides to introduce Crick. In a serious tone he declared emphatically,

'It is well known that Crick shuns publicity and avoids the company of women.' One of the slides showed cuttings from newspapers describing all about Crick's miraculous achievements and the other a photograph of a smiling Crick surrounded by a bunch of young and charming girls forming a double helix. They looked much better than the dinner tables.

Crick's lecture during the symposium ended with a verbal duel between Benzer and him. He admitted 'Benzer is always nasty towards me and he eventually wins.' Who is Benzer by the way? He initially worked with the bacterial viruses and mapped the fine structure of genes on the basis of their functions. Recons, cistrons etc. were the words coined in those days to specify functions of genes in the absence of complete knowledge of the organization of DNA. But the terminologies were quite helpful before the knowledge of structure-function relationship of DNA became clear and are used even in the present time. Later Benzer conducted extensive works on *Drosophila* (fruit fly) genetics.

Another hero of that Symposium was Gerhard Schramm who for the first time demonstrated that only RNA component of tobacco mosaic virus was responsible for infection of tobacco leaves contrary to the claim made by the expert virologist Heinz Fraenkel-Conrat. Incidentally, there are quite a few viruses of bacteria, plants and animals which have RNA as genetic material. Schramm was quite tall, handsome and humorous. At that time he was associated with the Institute for Virus Research in Tübingen, Germany, as its Director. Being the tallest person of the town he was known to all the people there. Throughout the sessions at Hyderabad Schramm was busy making sketches of the participants. Those expressed the artistic touch of the hand of the master mind. He drew a sketch of my wife (Maharani) who was also a participant and presented it to her. Unfortunately, we lost one of the prized souvenirs of our life. Schramm who died at a comparatively early age was interested in plant viruses

but talked on the 'Origin of nucleic acids' in the Hyderabad symposium.

My first encounter with DNA in the laboratory was rather uneventful. During the early 50's I was busy with the theoretical aspects of the techniques of paper chromatography and paper electrophoresis in connection with my doctoral thesis. Molecular biologists may laugh on hearing this as they routinely use sophisticated techniques like polyacrylamide gel electrophoresis, affinity chromatography etc. but may not be aware that column chromatography was the start and paper chromatography was the forerunner of all these techniques. At that time a friend of mine stepped in to my laboratory with a small vial having a partially damaged label which indicated that the material inside was nucleic acid. He planned to study the physical properties of the material and naturally wanted to know what type of nucleic acid it was. In those days only two types of nucleic acids were known, thymus nucleic acid and yeast nucleic acid. Thymus is rich in DNA, naturally the former was generally known as thymus nucleic acid. The latter contained ribonucleic acid (abbreviated as RNA), which we know today was mostly transfer RNA. Paper chromatography and paper electrophoresis were most powerful tools in those days for the separation of various cellular constituents. A.J.P. Martin and R.L.M. Synge got the Nobel Prize in chemistry in 1952 for inventing the technique of paper chromatography which helped to solve many problems, especially in combination with radioactive tracer technique (autoradiography). We, however, subjected the sample to paper electrophoresis after hydrolyzing the nucleic acid and ascertained that it was yeast nucleic acid or RNA. This is a minor incident but a major event for me as I started to switch over to biological sciences (specially molecular biology) from physical sciences with that sort of exposure.

Another symposium of no less importance in Molecular Biology was held in Delhi in 1969. G.P. Talwar who at that time was the Head

of the Department of Biochemistry of the All India Institute of Medical Sciences at Delhi organized a meeting on 'Control processes in multicellular organisms' under the auspices of the Ciba Foundation. G. E. W. Wolstenholme of the foundation was the main organizer and conducted the meeting very efficiently. Such meetings were held from time to time in England and other places as well and conducted in a different style than normally done. That particular meeting at Delhi held in 1969 was not in a lecture theater but around a center table in a seminar room at the India International Center. It was a closed-door conference. Some were direct participants who presented their own work and took part in the discussion following the presentations. There was plenty of opportunity for the participants to discuss at each and every stage, during and after the presentations. A limited number were also allowed to sit through but were not permitted to take part in the discussion. All the proceedings were recorded and published[4]. It was the first of its kind in India and my first exposure to such type of conference. Talwar not only organised a number of such symposia but did a lot for the development of Biochemistry and specially Immunology on modern lines in the country. He eventually founded the National Institute of Immunology, a premier research organization on Immunology at Delhi.

The star attraction of that Ciba symposium was Jacques Monod. He presented a paper entitled '*In vitro* studies of the *lac* operon regulatory systems', jointly authored with Suzanne Bourgeouis. There were plenty of opportunities to interact with him in the closed-door sessions. Most of the time he had a pointer with adjustable length in his hand, which he would lengthen and shorten during discussions. Apparently he was doing it unaware, especially when he was concentrating on a particular issue. I had the opportunity to meet him for the first time in 1960 when he visited the New York University School of Medicine and delivered a Saturday morning lecture. We were all aware of the famous 'Jacob and Monod' concept but the

seminal paper in *Journal of Molecular Biology* was not published at that time. I met the other one very briefly at the Pasteur Institute during my short trip to Paris. He looked entirely different from Monod but two names are glued together. Jacob's 'Logic of Life'[5] is a superb writing while, Monod's 'Chance and Necessity[6], a treatise on evolution is equally enchanting. The pair revolutionized the field of biology while others took sometime to prove their concept beyond any doubt but that took a firm ground as soon as it was conceived. In his lecture at New York University Monod mentioned that there is an intermediate message between DNA and protein synthesized in the cell. The message was transcribed from DNA and passed on to the translating apparatus (designated as ribosomes) to synthesize the protein according to the message. I do not remember whether the term messenger RNA was already coined or in the process of being coined. There was a lot of talk about DNA-like RNA. Elliot Volkin and Lazarus Astrachan had for the first time detected the presence of DNA-like RNA in T4-infected *E. coli* but they missed the implication of their own outstanding observation. I also had the opportunity to meet them personally when I visited their laboratory and delivered a talk there on the newly discovered DNA-dependent RNA polymerase, the final proof of the messenger RNA concept. At the Delhi meeting Monod elaborated the *in vitro* studies to prove this concept. Melvin Cohn who was associated with Monod in the experiment was also a participant. He addressed Monod as 'Guru'. Another Nobel laureate, G.M.Edelman of the Rockefeller University was the other star attraction. He had already determined the structure of the lectin (a sugar binding protein) concanavalin A and got the Nobel Prize for it. The topic of his lecture was, 'Antibody structure : a molecular basis for specificity and control of the immune system'. He gave a wonderful talk though he was not too serious about the symposium. He was mostly busy talking with Suzanne and Monod. The meeting was memorable for us as it was first of its kind in India. The third Nobel Laureate present there was Feodor Lynen. His work on acetyl CoA

earned the Nobel Prize for him. He talked on `Modulation of enzyme activities by metabolites and concentrated on the enzyme acetyl CoA carboxylase, a key enzyme in fatty acid metabolism. H. Harris of the Dunn School of Pathology, University of Oxford and an expert in cell culture spoke on `The expression of genetic information: a study with hybrid animal cells'. Another participant worth mentioning was J.F.A.P. Miller of the famous Walter Eliza Hall Institute of Medical Research, Melbourne, Australia. His contributions in Immunology known all over the world eventually earned him a Nobel Prize. He spoke on `Interaction between thymus cells and bone marrow cells in response to antigenic stimulation'. Each of these talks was a classic in their respective areas.

Late Zakir Hussain, the then President of India inaugurated the symposium and Monod delivered a general speech in the opening function held at the All India Institute of Medical Sciences. Both Monod and F.M. Young, Chairman of the Department of Biochemistry, University of Cambridge who was also the Chairman of the organising committee were overwhelmed by the gesture of the President who in spite of his busy schedule took his time off to meet the scientists.

Since I started out with D.M.Bose, Director of Bose Institute, let me conclude the Prologue with him again. Sometime in 1961 just after my return following three years' sojourn to Canada and U.S.A., Arthur Kornberg, Herman Kalckar, Joshua Lederberg and some other scientists visited the Bose Institute on my personal initiative, which was obviously approved by the Director who was overwhelmed by their visits. We were having dinner at his house, which was a rare occasion for me as I was a young research worker then. It was no doubt a great privilege to have dinner with the Director especially at his home. Arthur and Sara Kornbergs were the main invitees. While we were busy with the typical Indian food Arthur Kornberg suddenly looked attentively at D.M. Bose and shot a pointed question 'Why did Har Govind Khorana not get a job in the country? I knew, Khorana

was an intimate friend of Kornberg. At that time neither Khorana nor Kornberg had the Nobel Prize. So Khorana was hardly known in his own country. The Director in spite of his voracious scientific appetite had also not heard about him. He looked puzzled. To save the situation I whispered to him in my native language (*Bengali*) that Khorana, a brilliant chemist now settled in U.S.A. may eventually get the Nobel Prize. D.M. Bose was clever enough to fully understand the hint and explained to Kornberg, 'Difficult job situation in the country is responsible for that'. Kornberg, a hard nut to crack could not be pacified with his answer. However, he should have known that if he did get a job in the country after returning from England following his training in Alexander Todd's laboratory he would not have achieved what he did in Canada and subsequently in U.S.A. I saw Khorana from a distance for the first time at the National Research Council Laboratory at Ottawa, Canada surrounded by top-ranking chemists. He was already known worldwide for the synthesis of coenzyme A. A Canadian friend of mine was surprised to learn that I did not know him. Later I came in close contact with him when he joined the University of Wisconsin at Madison, U.S.A. He did not believe that proper research could be done in India. Whenever I met him he asked 'How can you work in India ?'

Kornberg delivered a brilliant lecture on the DNA synthesis in the Bose Institute Auditorium. During the informal discussion D.M.Bose asked his wife who also was a biochemist, 'When do you think your husband will get the Nobel Prize?' She replied, 'He would definitely get it but not so soon'. The next year (1959) he got it. It is also gratifying to note that Khorana (who received the Nobel Prize in 1968) delivered a lecture in the same auditorium in 1974 almost a decade and half later. Unfortunately, I had already left Bose Institute and moved out of Calcutta.

Drifting away from the main topic are deliberate attempts made by me in what follows. That was not just to trace the path of history.

The incidents, which have been cited, are not necessarily important historical events but throw light on certain aspects of molecular biology, which may not find a place in any historical record. A few who will have the curiosity to go through this book may get glimpses of the course of development of molecular biology, a very specialized area, through the eyes of a witness who was occasionally involved directly or indirectly in its development.

## References

1. *Molecular structure of nucleic acids (A structure for deoxyribose nucleic acid).* J.D.Watson and F.H.C.Crick, Nature, April 25, 1953, page 737.
2. *The double helix (A personal account of the discovery of structure of DNA).* James D.Watson, Atheneum, New York.1968.
3. Nucleic acids (Structure, biosynthesis and function). Organised by P.M.Bhargava at the Regional Research Laboratory at Hyderabad, Council of Scientific and Industrial Research, New Delhi.1965.
4. Control processes in multicellular organisms (A Ciba Foundation Symposium). Edited by G.E.W.Wolstenholme and J.Knight, J & A. Churchill, London.1970.
5. *The logic of life (A history of heredity).* Francois Jacob, Princeton University Press, Princeton, U.S.A.,1973.
6. *Chance and necessity (An essay on the rational philosophy of modern biology).* Jaques Monod, Collins, Fontana books , original in French,1970.

*****

# Chapter - 1
## Who creates the music?
### (Structure and Function of DNA, Master of the cell)

*The innumerable variety of organisms existing on the earth today seems to harp on a common theme 'reproduction'. More than fifty years have elapsed since the publication of "What is life?"[1] based on the provocative lectures of Erwin Schrödinger. In spite of the phenomenal development of molecular biology we still do not understand 'What is life?' But each one of us enjoys the "music of life" played by DNA (abbreviated form of deoxyribonucleic acid), the thread-like tape having all the information for the whole organism. The DNA molecules are composed of only 4 characters (A,T,G,C) but contain thousands or millions of those in a correct sequence. In the seminal paper published in Nature in 1953, James Watson and Francis Crick proposed the double helical structure of DNA. That paper revolutionized life sciences and put molecular biology on a firm footing. The structural features of DNA and their functional implications, specially duplication during reproduction, have been narrated in this 'introductory' chapter focusing on some personalities who brought the hidden facts of life into broad day light.*

There cannot be a better history of a discovery written but by the discoverer himself rather than someone else. 'Double helix'[2] is so far the sole example of that novel way. The main reason is not only the discoverer's thrilling experience but also the way those were put in fitting words and proper perspectives. The two words 'Honest Jim' were the starters. Jim Watson realized that his honesty might be questioned by the readers of *Double Helix*. Before reading the novel I had the impression that Crick was the main actor in the drama but

after going through the book my idea changed. I am sure, not only myself but also others would fall in the same line. No wonder that Crick wanted to sue Watson (may be it was a rumour only!). *The path to the double helix*[3] written by Robert Olby at the instance of Francis Crick was no match to *Double Helix* although its size is larger and the recordings are fully authenticated. However, we usually enjoy reading novels better than studying histories. The major share of the credit of the discovery should go to Erwin Chargaff but he was unfortunate. I hesitate to say that he was not intelligent enough, may be he was rather less imaginative. As a first rate chemist he extensively analyzed various DNA samples and observed that there was no exception to the rule, adenine invariably equals to thymine and guanine to cytosine. The genes, the carriers of heredity are deoxyribonucleic acid molecules abbreviated as DNA. The DNA is constituted of 4 alphabets of nature (bases) which are abbreviated as A (adenine), G (guanine), T (thymine) and C (cytosine). Unfortunately, Chargaff could not put his hand on a single stranded DNA (which was discovered later), otherwise he would have been puzzled but could have got the clue. Nature would have opened its signature to him for the first time but someone else read it. Else means 'Duo' who had their eyes and imaginative minds open. Cardboard structures of adenine, guanine, cytosine and thymine acted as their basic ingredients, which they started to fit in the wire-built double helix. After a lot of trial and error, the pairs, which fitted with each other, were adenine with thymine and guanine with cytosine. The space between the helically wound wires was just enough to accommodate the specific pairs only. Adenine and guanine, the two purines (as those are called, in general) are larger in size than the two pyrimidines, thymine and cytosine as shown in the (Fig. 1.1). Only one purine and one pyrimidine could fit in the space between the two helical strands. Linus Pauling's concept of hydrogen bond indirectly helped them. Only two hydrogen bonds could be formed for the specific pairing of adenine and thymine and three between guanine and cytosine. So everything matched perfectly.

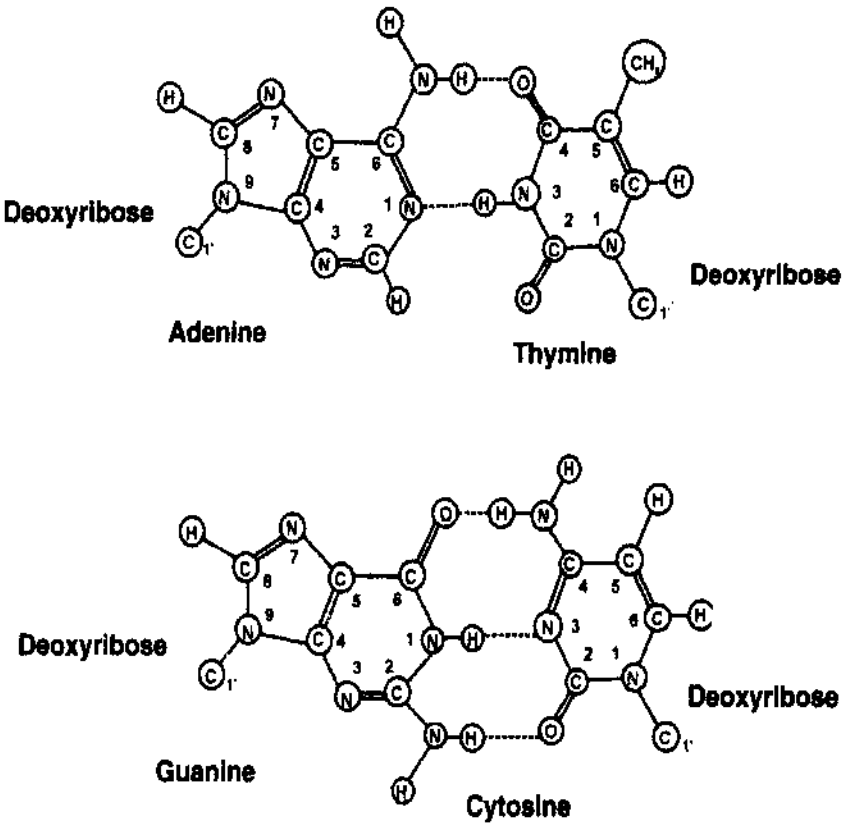



Fig. 1.1 Base pairing in DNA. Adenine (a purine) pairs with thymine (a pyrimidine) with two hydrogen bonds while guanine (a purine) pairs with cytosine (a pyrimidine) but with three hydrogen bonds. The space is not enough to accommodate the paired purines (both being of large size) while pyrimidine pairs are smaller in size and therefore not able to make contact to form hydrogen bonds.

That was the magic touch and the real dramatic one no doubt but how did they scent that DNA was helical? That's another story.

Chargaff was no doubt frustrated; it was quite obvious from the sarcastic writings, which emerged from his powerful pen[4]. But he himself was responsible for his misfortune due to lack of imaginative power. He most probably deserved a share in the Nobel Prize but 'most probably' was a big question between him and the prize. The story of Rosalind Franklin was equally, if not more, sad. She had her opponents; she was quarrelsome, fought with her own supervisor (Maurice Wilkins) who was also heading for the Nobel Prize. She had the clear X-ray picture of DNA with the signs of cross, a distinct signature of helical structure (Fig 1.2). Watson and Crick stole her data not in the true sense, as some people may like to think. Actually, the concept, which could have belonged to Rosalind, was stolen. Like Chargaff she also missed another signature. The hint was from Pauling's helical structure of proteins. Again the 'duo' drew a lesson from the giant imaginative rival (Pauling) who indirectly was the grandiose source of inspiration for the miraculous achievement. A veiled threat is

Fig 1.2 X-ray picture of DNA. The cross sign is a signature of helical structure in DNA. The picture was first taken by Rosalind Franklin and later confirmed by Morris Wilkins.

sometimes more powerful than the real one. So there were two big thrusts from two directions, the rest was imagination. When they coined the word 'double', they did not realize at that time that the two strands would be named after them. Crick strand represented the coded information which was more important than the other, Watson strand. As we came to know later, both are equally important; the coded information is not confined to one strand only. The knowledge of both chemistry and heredity proved to be of immense help to their fertile mind.

The question is why double? The answer is in the seminal paper in Nature '*It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material*'. The most exciting aspect of the paper was that particular statement. The structure explains the mechanism of duplication of the genetic material and the passing on of the genetic information from the parent cell to the daughter cells when it divides. There were at least three probable mechanisms. In the first one the DNA in the dividing cell would degrade and the degraded materials would lead to the synthesis of the new DNA. This was more or less ruled out conceptually from the hereditary point of view. The genetic material should never be degraded. The second one (known as conservative mechanism) assumed that DNA duplicates in the dividing cell and one of the two daughter cells gets the original DNA and the other, the duplicated DNA. As per the third possibility (semi-conservative mechanism) each of the two daughter cells gets one original strand and one of the newly synthesized one. Watson and Crick probably had no idea about the mode of replication as well as the mode of passing of the hereditary material (DNA) to the daughter cells after duplication though they made some hints at it in their seminal paper. The answer was provided by the famous experiment of Mathew Meselson and Franklin Stahl who could clearly demonstrate the semi-conservative mechanism of DNA replication.

The two helical strands have the probability of running parallel or

antiparallel to each other, which would be made clear when the primary structure of DNA would be discussed. Although the two strands are antiparallel it was difficult to establish this fact with only model building tools. It was, however, proved experimentally eight years later (in 1961) by Arthur Kornberg who synthesized radioactive DNA from its ingredients and analysed them. John Josse, a medical student in Kornberg's laboratory did this painstaking work. I met him at Kornberg's house where he was enjoying swimming in the Nobel Pool. Kornberg had built the pool in his house located on a hilltop at Stanford with his Nobel Prize money, so his students referred to that as the Nobel Pool. When I asked Josse how he managed to do that laborious business his reply was "I did it at the suggestion of Arthur (Kornberg) but I won't do it any more even if he wants me to do it." But that was really a very important contribution in establishing the final structure of DNA. Eventually, that was confirmed by X-ray crystallographic studies. Double helix with the beautiful helical turns surpasses the beauty of a maiden but we usually forget that the beauty lies in its grooves, the major and the minor one (Plate1 Fig 1.3). The bases are so arranged that the rise per base pair is 3.4Å (1Å = $10^{-8}$cm). Since each helical turn of DNA double helix has 10.5 base pairs, its pitch will be 36 Å. When we intend to measure the length of materials which are visible only under a microscope (specially electron microscope, the powerful one) Angstrom (Å) is used as the basic unit of length. 1Å= $10^{-8}$ cm means if one divides 1 cm in 100,000,000 equal divisions each one division would be 1Å. The unit is christened after the name of the famous Swedish physicist Anders Jonas Ångström.

The two helices embrace each other in such a way that the two grooves (like in a screw where there is only one groove) continue throughout the entire length. The wider groove is referred to as the major groove and the other one as the minor groove. Have these grooves any special function? It took quite sometime to understand

that. Although DNA stays in a folded form in association with some special proteins called histones (sometimes spermin) in the safe environment of the nucleus yet it has to replicate itself at the time of cell division and also unwind simultaneously. This process is rather complicated. An enzyme, named as DNA-dependent DNA-polymerase, discovered by Arthur Kornberg in 1954 does this job in the test tube but not in the cell. It adds the bases (rather nucleotides) one after another using both the strands as the template. The two strands have to partially unwind for that. We have to postpone further discussion of this till the DNA structure is elaborated. The main point to emphasise here is that the enzyme, which is responsible for its duplication, is a protein and has to recognize and bind to the helix. It recognises both the major and minor grooves before proceeding along the DNA chain. Replication is not the only duty of DNA. One of the two strands has to produce the requisite message, which is done by another enzyme, DNA-dependent RNA- polymerase. The process is known as transcription. That enzyme has to also move along the major groove. Sometimes the replication and transcription have to go on simultaneously, as in the bacteria. So both the enzymes should move simultaneously along the DNA. Apart from those two, a number of other proteins known as regulatory proteins, required to regulate the cell metabolism, also have to interact with DNA. They mostly interact with the bases and to some extent to the linker phosphate groups. The grooves are the best ways of their approach. Though we have already learnt a lot about the regulation of replication and transcription, we still need to learn more. We shall understand these complexities as we shall understand the details of the structure of DNA. So let us wait for that. The form in which DNA occurs in state of solution is very close to that described by Watson and Crick and is known as form B (Plate1 Fig 1.3b). There is another form known as 'A' which is in a less hydrated state. It has 11 bases per turn and its diameter is slightly larger than the B form (Plate1 Fig 1.3).

Chromosomes are the hereditary elements, which we inherit from our parents, 23 from the father and 23 from the mother (Fig.1.4) thereby, we inherit 23 pairs of chromosomes. All our cells (except red blood cells which have no nucleus) contain 23 pairs of chromosomes. However, germ cells of both male and female (sperm in the case of male and ovum in the case of female) contain only one set of 23, each ready to be transmitted to the offspring. Chromosomes are constituted of DNA and proteins, mostly histones (some non-histone proteins as well in few cases). The hereditary information is contained in DNA. For a long time it was believed that proteins are the genetic elements. The famous experiments of Avery on the one hand and Hershey and Chase on the other proved beyond doubt that the hereditary information is contained in DNA and not in the protein. What are genes anyway? Those can be looked upon as specific transmitters of hereditary characters and usually regarded as complex self-perpetuating molecules and the constituent of the chromosomes. The general mode of explanation in the textbooks is that gene has

Fig. 1.4. Human chromosomes (23 pairs) arranged according to their sizes (karyotyping). One of each pair is inherited from father and the other one from mother. Except chromosome 23, the sex-determining chromosome all others are known as autosomes. Sex chromosome of female contains two identical chromosomes (XX) whereas that of male contains two different ones (XY).

something to do with heredity e.g., "*Heredity, the conservative factor in nature results from the transmission of similar physical elements (genes) from parent to offspring.*"[5] Gardner has tried to explain it further: "*Tigers beget little tigers and not representative of other species because specific physical elements are transmitted from parents to their offspring through the gametes, that is eggs and sperms*". We all know that animals including humans originate from the fusion of ovum and sperm, so they must contain the physical elements (genes). Genetics is the study of genes, which are involved in heredity. The term gene means '*to generate*' (in Greek) and was coined by William Bateson in 1906. Gregor Mendel is appropriately called the 'Father of Genetics'. The celebrated Austrian monk carried out the cross breeding experiments with garden peas and discovered the basic principles of genetics. For analysis he used a variety of characters of garden peas, for example, outer surface of seed, colour of flower, height of plants etc. and could unequivocally demonstrate the flow of 'hereditary' characters from the parents to the offsprings in a predictable fashion. This work carried out during the mid-nineteenth century remained practically unknown for more than 30 years but was fortunately rediscovered in 1900 independently by three scientists. Subsequently, the science of genetics developed rapidly, specially in the hand of William Bateson who confirmed and crystallized Mendel's principles and introduced the term 'Genetics.' It did not take a long time to recognize that the 'chromosomes' present in the nucleus of the cell contain the hereditary elements known as genes. Early cytologists could demonstrate by cytological techniques that the nucleus of the cell and not the surrounding cytoplasm houses the hereditary elements. During cell division certain cell structures become distinctly visible on staining with basic dyes. Those structures were therefore called chromosomes ('chrome' means colour) in 1888 i.e., forty years after their discovery. It did not take long time to understand that 'chromosomes' contain the 'genes'. Every species has its characteristic number of chromosomes in a cell (usually they occur

in pairs, one set from the egg cell and one from the sperm cell). The number of chromosomes varies from species to species for example, humans have 46 chromosomes, mice have 22, certain species of lizards have as many as 140. The size of the animal or plant does not commensurate with the number of chromosomes.

As already mentioned, for a long time it was believed that proteins are the hereditary elements. Oswald T. Avery and his coworkers carried out a classical experiment in 1944 whose results established that DNA and not protein is the genetic material. The phenomenon of transformation in case of microorganisms was known for quite sometime from the experiments of Frederick Griffith, done in 1928. Pneumonia is caused by the bacteria *Pneumococci (cocci* is pleural of *coccus* meaning spherical or globular). There are two varieties of *Pneumococci*, the smooth type and the rough type, named on the basis of the type of colonies they form on agar plates. 'Smooth' type cells are capsulated with a mucous layer and produce pneumonia and are therefore branded as 'virulent'. The 'rough' type have different type of capsule and are avirulent. If an animal, say a mouse, is infected with the 'rough' type or heat-killed 'smooth' type of *pneumococci* it won't suffer from pneumonia. Surprisingly enough, Griffith found that if a mixture of 'rough' type and heat-killed 'smooth' type were injected together, pneumonia was produced. He also could find a large number of living 'smooth' type (virulent) cells in the mice after killing it. He unhesitatingly concluded that the genetic material of the virulent cells (from heat-killed 'smooth' cells) was passed on to 'rough' type of cells (living but nonvirulent) and transformed the latter to virulent ones (Fig 1.5). The marvelous achievement of A.T. Avery and his colleagues C.M. Macleod and M.McCarty was that they could demonstrate such transformation in a test tube by isolating DNA from the 'smooth' cells and adding it to the culture of the 'rough' cells when the latter were transformed to 'smooth' (virulent) cells. This remarkable paper was published in 1944 in the *Journal of*

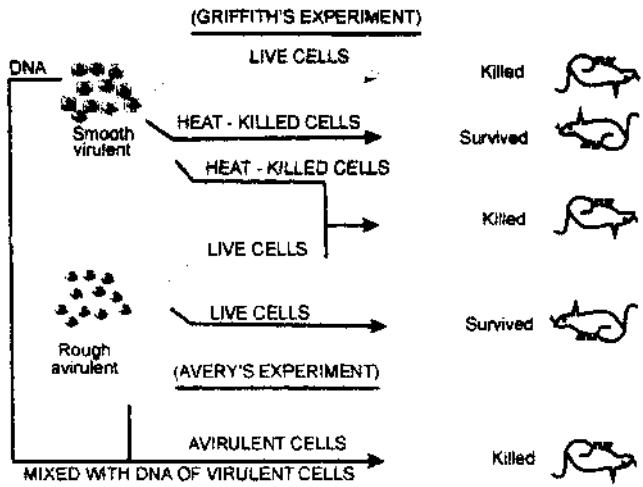



Fig.1.5 Transformation of non-virulent *pneumococci* into virulent ones. Griffith observed that on injecting heat-killed virulent cells (noninfective) along with nonvirulent live cells into mice produced pneumonia and lots of virulent cells were found in the dead mice. Avery and his coworkers demonstrated that DNA isolated from heat-killed virulent cells could transform avirulent cells into virulent type thus establishing for the first time that DNA is the genetic material.

*Experimental Medicine*. Although his contemporaries were somewhat skeptic, Avery was almost certain that the transforming principle was essentially pure DNA free from protein. That was a sensational discovery but being a person of quiet temperament he did not create any 'noise', so it took quite sometime for others to understand the implication of his results and design further experiments to establish it firmly. His biographer has properly depicted the person[6].

Another confirmatory experiment to prove that DNA is the hereditary material was done by A.D. Hershey and M.Chase (Fig.1.6) almost eight years later. They used a host-virus system (*Escherichia coli*, a bacterium and its virus bacteriophage T2) to show that DNA

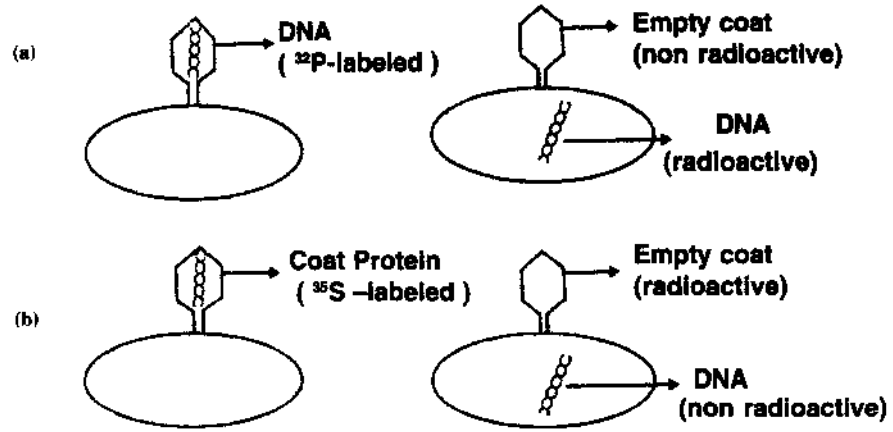



Fig. 1.6. Hershey-Chase experiment. a) *E.coli* cells were infected with bacteriophage T2 containing $^{32}$P-labeled DNA. The phage injected $^{32}$P-labeled DNA within the host. b) *E.coli* cells were infected with phage containing $^{35}$S-labeled coat protein. In this case the radioactivity remained outside. This proved that DNA and not protein contains the genetic information required for the production of phage particles within the host.

encapsulated within the protein coat acts as the genetic material. Viruses cannot be treated as truly living system as they don't have all the components of the living system. However, they can multiply within their specific hosts with the help of their genetic material. They themselves also have partial genetic information in the form of DNA or RNA. The fact that RNA can also act as genetic material became evident through the life cycle of RNA-containing viruses. Bacteria, plants, animals etc have their specific viruses. Bacterial viruses are usually referred to as bacteriophages. Hershey and Chase did the well-planned experiment using radioactive phosphorous ($^{32}$P-labeled DNA) containing bacteriophageT2 in one case and radioactive sulfur ($^{35}$S-labeled protein coat) in the other case. DNA has no sulfur; protein coat of T2 has no phosphorous. Only $^{32}$P-labeled DNA entered into the *E.coli* cells while $^{35}$S-labeled coat was left outside giving a clear

demonstration of DNA being the genetic material of the virus.

So we accept DNA as the genetic material but what is a 'gene' then? By classical concept 'gene(s)' carry the information for some hereditary characters like colour of the eye, height of a plant, shape of a seed etc., but by present day knowledge we know that a particular colour of an eye is the net effect of the action of several genes, a plant is tall as the ultimate result of expression of a large number of genetic elements and so on. Luckily by now, we understand the structure of DNA thoroughly. There would be an occasion later to discuss that the hereditary or genetic messages contained in DNA being finally expressed as a variety of proteins formed in the cell which carry out a large number of functions of the cell, starting from its duplication and ending in death. In terms of 'molecular biology', which deals with the living system at the molecular level, a gene may be defined as the basic element of the genetic material (genome composed of DNA), which has information for a particular protein. In fact, we are indebted to George W. Beadle and Edward L. Tatum for the one-gene-one-enzyme hypothesis. Enzymes are invariably protein molecules, therefore the number of genes is equal to the number proteins. This statement may not be absolutely right but so long as this conveys the meaning let us be happy with that. In the early 1950's two types of nucleic acids, yeast nucleic acid and thymus nucleic acid were known. The so-called yeast nucleic acid turned out to be transfer RNA (in modern nomenclature), which is one of the various types of ribonucleic acids or RNAs. The thymus nucleic acid was established as deoxyribonucleic acid or DNA. The discovery of nucleic acids, particularly DNA, dates back to 1858 almost hundred years before the structure of DNA was proposed by Watson and Crick in 1953 and also much before DNA was established as the genetic material by Avery and his colleagues and later confirmed by Hershey and Chase.

Frederich Miescher, while a medical student was advised by his physician uncle William Hess to work on the constituents of a cell.

Hess was a well-known anatomist of Swiss origin. After passing his M.D. examination Miescher joined the laboratory of famous organic chemist Hoppeseyeler (Hoppeseyeler's journal is well known). It was located in an abandoned palace on the river Necker at Tübingen. This was the first biochemistry laboratory in the world. In the hope of finding something new in the cells, Miescher used to collect rejected bandages from a nearby clinic and analyzed them. Working alone at very late hours he sometimes felt haunted in the huge palace. When I had the unique opportunity of looking at the palace from the other side of Necker the question of feeling haunted did not arise but I was overwhelmed with its majestic beauty, with the built-in inherent 'history of life' created in the nineteenth century. No one can escape that feeling. When Miescher got the thread-like material and found that it contained phosphorus he convinced himself that he had isolated some new biological material. Proteins (`protein' or proteus in Greek means `of first rank') were well characterized by then and thought to be of prime importance in living systems. It was known at that time that proteins do not contain phosphorus although we know now that some proteins may be phosphorylated. However, phosphrous is not a normal constituent of protein. Somehow or the other, he believed that the material came from the nucleus of the cell and called it `nuclein'. Much later it was realized that what he isolated was DNA and it really came from the nucleus. That was a magical feat! The other type of nucleic acid (yeast nucleic acid) was called RNA and the basic difference between the two was in the sugar component. RNA contains ribose sugar (hence called ribonucleic acid) whereas DNA contains deoxyribose sugar instead. There are some differences in the base composition as well. Both, however, contain phosphate. As already mentioned, DNA contains four bases, adenine, guanine, cytosine and thymine whereas RNA contains adenine, guanine, cytosine and uracil instead of thymine (Fig. 1.7). It should be pointed out at this stage that sometimes RNA may contain thymine but DNA never contains uracil. Further, the bases can be modified in DNA as well as RNA

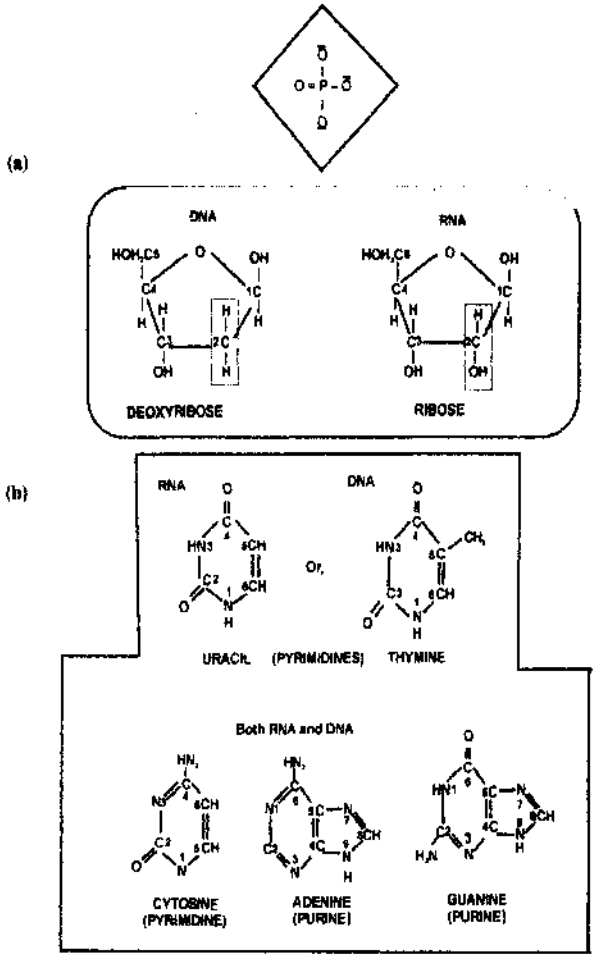



Fig. 1.7 Structures of nucleic acid constituents. a) Phosphate is the common constituent which bridges the pentose sugars (deoxyribose in DNA and ribose in RNA) in the polynucleotide chain. The difference between the two sugars is in the presence of 2′OH group in RNA and only H (no O) in DNA Numbering of carbon atoms are indicated. b) Structure of bases (adenine, guanine and cytosine which are common between the two nucleic acids).Thymine occurs in DNA and uracil in RNA (numbering of the atoms are indicated).

after their incorporation into nucleic acids. Phosphate, however, is present in both DNA and RNA and acts as linkers between riboses or deoxyriboses as the case may be. Both ribose and deoxyribose are pentose (5-carbon) sugars. The difference lies in one oxygen atom at position 2. This simple oxygen atom is responsible for a number of differences between the two nucleic acids. RNA can sometimes act as genetic material as in RNA viruses but it carries out lots of other functions, which DNA does not.

The backbone of the structure of nucleic acid is built with alternate residues of sugar and phosphate in a very simple fashion as shown in (Fig.1.8a). Alternate residues of sugar and phosphate make the monotonous chain, sugar-phosphate-sugar-phosphate-sugar-phosphate-sugar-phosphate- sugar-phosphate. Phosphate actually bridges two sugar residues (ribose or deoxyribose). Since there are two ester bridges on the two sides of the linker phosphate it is referred to as a Diester Bridge. Phosphate links 3′ position of one sugar to the 5′ position of the other sugar (hydroxyl groups being present in both the positions). In DNA there are only two hydroxyl groups free (3′ and 5′), therefore, the connection is invariably between 3′ position of one to the 5′ position of the other. In RNA similar bridges are formed but 2′-5′ bridges can also be found in rare cases. The 2′ -OH group remains free in RNA, which makes it labile to alkali. DNA without 2′-OH is quite stable under that condition. The bases A, T, G, C, etc. are attached to the 1′ position of sugar residues as shown in (Fig.1.8b). The position is marked as 1′ as the positions of the atoms of bases are marked as 1, 2, 3 etc. in order. Hence, the atoms of sugar residues are marked as 1′, 2′, 3′ etc. to avoid the confusion. The pyrimidines are connected to 1′ position of sugar through N in position 3 and purines through N in position 9. This will be more clear from the following discussion.

When DNA and RNA are partially hydrolyzed their phosphate bridges are broken and nucleotides are obtained. The nucleotides are

the combination of one unit each of base, sugar and phosphate and constitute the building blocks of nucleic acids. Adenine-ribose-phosphate is specifically called adenosine monophosphate (adenosine-5′- phosphate) or adenylic acid. Thus the four types of nucleotides (adenylic, guanylic, cytidylic and thymidylic acids) constitute DNA. RNA is usually constituted of the first three and uridylic acid. To distinguish between DNA and RNA ingredients deoxyadenylic, deoxyguanylic and deoxycytidylic acids are, in general, called deoxyribonucleotides. Without the prefix deoxy those mean ribonucleotides. Thymidylic and uridylic acids are used without a prefix and mean deoxynucleotide and ribonucleotide respectively. If the nucleotides are further broken down by removal of phosphate, nucleosides (combination of base and sugar only) are obtained. If the base is adenine, adenosine (adenine-sugar) is released. This would be called adenosine or deoxyadenosine depending on whether it is obtained from RNA or DNA or, in other words, contains ribose or deoxyribose as constituent sugar. With this understanding we can easily write the structure of DNA or RNA as shown in Fig. 1.8b.

For all practical purposes there is no necessity of writing the structural details of DNA or RNA. By convention we may not write the backbone structure, which is obvious and thus can present the structure of the nucleic acid in a simpler way as shown below.

5′ pApGpApApGpTpCpC.....pT.3′,

or in a still simpler way

5′ AGAAGTCC.....T. 3′

One has to remember that bases are not directly connected to each other but connected to the sugar residues of the monotonous chain of sugar-phosphate backbone. However, these are present in a perfect sequence, which leads to the identity of the individual nucleic acid. Thousands of them may be present but in correct sequence. One more convention should also be clarified. ApG and GpA are structurally different. ApG means A-3′p5′-G whereas GpA as G-3′p5′-A. In the

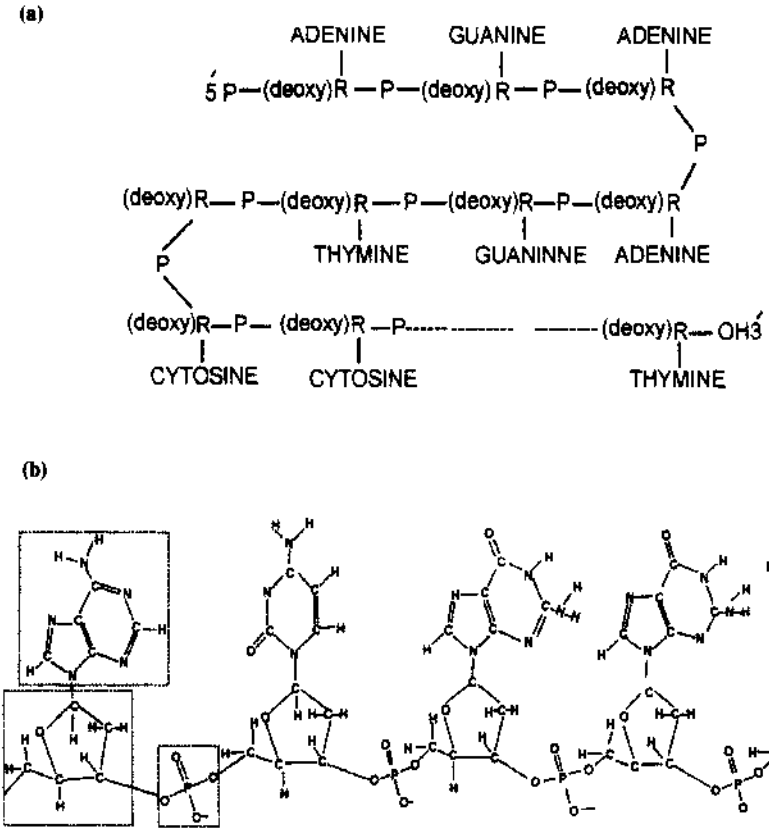



Fig.1.8 Primary structure of deoxyribanucleic acid (DNA). The backbone chain is monotonous, constituted of alternate residues of sugar and phosphate. Phosphate makes the diester bridge between the two consecutive sugar units connecting 3´OH group of one to the 5´OH of the other. Bases are the only variable units determining identity of the nucleic acid. a) Structure written without details. (deoxy)R indicates deoxyribose and P stands for phosphate. b) Detailed primary structure of a DNA is shown in the diagram.

same way all phosphate bridges in nucleic acids are normally written with the 3′ connection on the left hand side and 5′ connection to the right hand side. There is a free phosphate at the 5′ end and free OH at the 3′ end. So the two ends are designated as 5′ and 3′ ends.

Since A pairs with T and G pairs with C it is easy to write the structure of double stranded DNA. There are two ways (parallel and antiparallel) of writing the structure of double stranded DNA as shown below.

| Antiparallel | Parallel |
|---|---|
| 5′ AGAAGTCC.....T.3′ | 5′ AGAAGTCC.....T.3′ |
| 3′ TCTTCAGG.....A5′ | 5′ TC TTCAGG.....A3′ |

The two types of structures have been presented schematically in (Fig.1.9). From the above discussion on the structure of DNA one should keep in mind that the molecules (parallel or antiparallel) presented as above are not identical. Another point to remember is that while writing the antiparallel structure the convention of writing the opposite strand from 5′ to 3′ direction cannot be followed for obvious reason. It has already been mentioned that the two strands of DNA are antiparallel.

As realized for the first time by Pauling, H atom itself can form a weak bond between two electronegative atoms like N,O etc. Watson and Crick argued for the first time that adenine can pair with thymine through two hydrogen bonds (A=T) and guanine can pair with cytosine with three hydrogen bonds (G≡C) as shown in Fig. 1.1. The space between the two strands can accommodate one purine (A or G) and one pyrimidine (T or C). Purine-purine pair can not be accommodated as both are of large size while pyrimidine-pyrimidine pair both being small in size can not come close enough to form the hydrogen bonds. Thus A=T and G≡C are the only specific base pairs in DNA. Sometimes there are unusual pairings but let us not worry about that.

With the basic knowledge acquired, let us pause for a while to

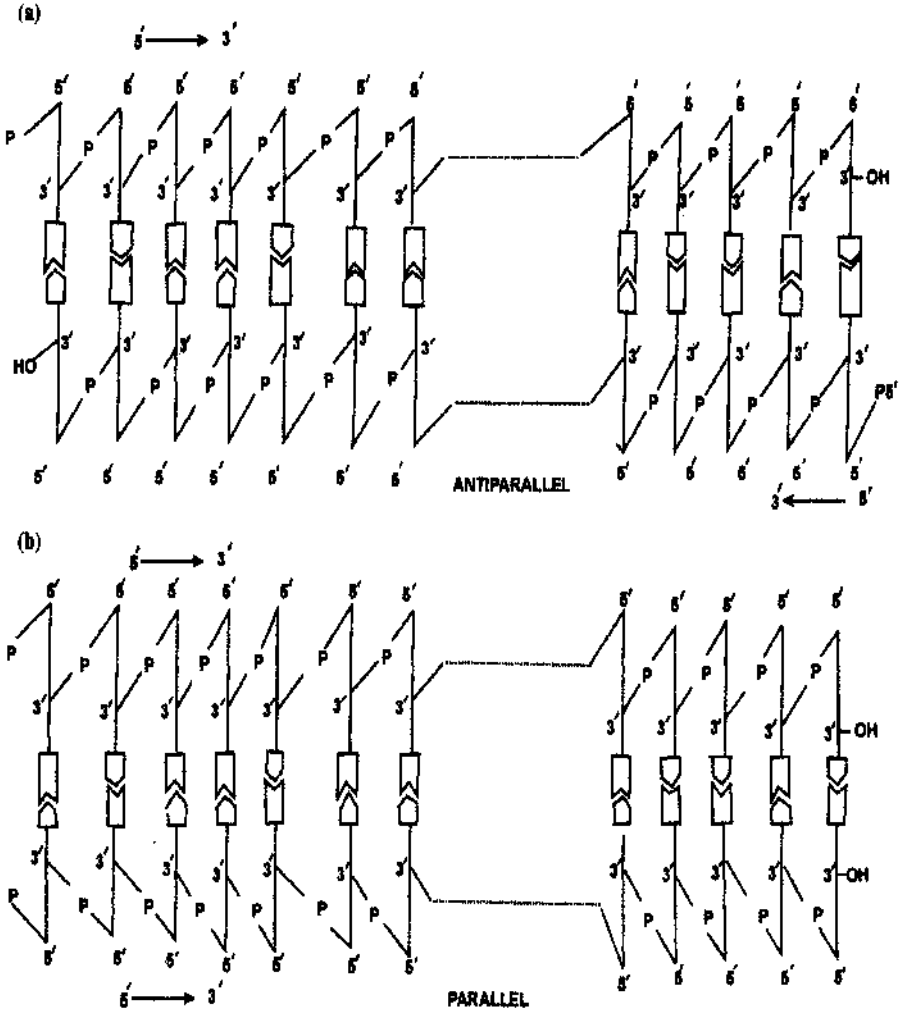



Fig.1.9 Two possible ways (parallel and antiparallel) of alignment of DNA strands. (a) antiparallel (b) parallel. In nature, the two strands are antiparallel.

find out how Crick arrived at the enormous length of human DNA which he mentioned in his popular lecture at Hyderabad (as narrated in Prologue) Let us try to do some calculations to find out the total length of DNA present in a human body. The human DNA has approximately 3 billion ($3 \times 10^9$) base pairs. Assuming that its total length is approximately 2 meters and the human body consists of approximately $10^{14}$ cells and if all the DNAs of an individual are joined end to end the total length of DNA will be $2 \times 10^{14}$m or $2 \times 10^{11}$Km. This length is enormous as compared to the following: circumference of the earth ($4 \times 10^4$Km), distance between earth and moon ($4 \times 10^5$Km) and distance between earth and sun ($5 \times 10^8$Km). Since the diameter of a single DNA strand is 1nm ($1 \times 10^{-9}$meter) its area of cross section will be $\pi r^2$, where $\pi$ is 22/7 and r is radius ($0.5 \times 10^{-9}$m). So the cross section becomes $0.75 \times 10^{-18}$m$^2$. Volume of the total DNA chain will thus become $0.75 \times 10^{-18} \times 2 \times 10^{14}$ m$^3$ = $1.5 \times 10^{-4}$ m$^3$ = $1.5 \times 10^2$cm$^3$ = ~150cm$^3$ which is the approximate volume of a cricket ball.

Let us talk about another DNA expert, Arthur Kornberg. I came to know him through his close friend Bernie Horecker, the famous enzyme chemist with whom I worked at the National Institutes of Health, in U.S.A. Kornberg was deeply involved at that time in discovering enzymes responsible for the synthesis of ingredients of DNA, deoxyribonucleotides and eventually DNA itself. The term enzyme was coined for the catalysts in biological systems. Usually these are proteinaceous in nature. The Greek term enzyme means in yeast (*en* means 'in' and *zyme* means 'yeast') since such catalytic activity was first discovered in yeast. After passing his medical examination Kornberg decided to work with enzymes to learn enzyme chemistry, which was essential to understand the living system and hence joined Severo Ochoa's laboratory in New York University School of Medicine. This specialty, then a new area, was developing very fast. Alan Mehler whom I also met at NIH and developed intimate friendship with was in Ochoa's laboratory at that time and was of

great help to Kornberg. The latter freely admitted that whenever the necessity arose he fully utilized this training. Later, I had the good fortune of joining Ochoa's laboratory just after Ochoa, the *Guru* or Master and Kornberg the *Sishya* or disciple shared the Nobel Prize in Medicine in 1960. When I joined NIH in 1956, Kornberg had already moved to Washington University at St. Louis. He was also an intimate friend of Har Govind Khorana, the Nobel Prize winner of Indian Origin. Kornberg, the wizard of enzymes had his eyes on the synthesis of gene almost from the very beginning of his career as Watson had on its structure. What Watson could achieve within a short period with the help of painstaking collection of data from others (specially, Chargaff and Franklin) Kornberg had to collect all the data himself from the test tubes. He started in a rationalistic and simplistic way as Pauling did for determining the structure of proteins. The latter started with peptides (components of proteins) without jumping into proteins. When Kornberg started, enzyme chemistry had developed reasonably well and new enzymes were being discovered every now and then but enzymes responsible for the synthesis of macromolecules like glycogen were being just touched upon. Those days the discovery of enzymes responsible for the synthesis of proteins, nucleic acids etc. were beyond imagination. He started to discover enzymes after enzymes responsible for the synthesis of nucleotides, specially deoxyribonucleotides, the building blocks of DNA. It was well known at that time that radioactive thymidine (one of the bases of DNA used as marker) could be incorporated into DNA of intact cells but its pathway was not known. He worked out the pathway from thymine to thymidylic acid, which is known as the salvage pathway. He also worked out the alternate pathway for the synthesis of oligodeoxynucleotides, which as a whole is known as the *de novo* pathway.

These painstaking researches were discussed by him in the symposium on 'Chemical basis of heredity' held at Baltimore in 1956.

His seminal paper on DNA synthesis was first published as short communication in the journal *Biochimica Biophysica Acta* and the details were provided later in Science. From the very beginning he realized that there would be involvement of energy in DNA synthesis. dATP, dGTP etc. which are thought to be more energy rich than the di- or mono-phosphates would be the actual precursors of DNA. Their structures are shown in (Fig 1.10). It has already been discussed that the basic units of nucleic acids (both DNA and RNA) are nucleoside monophosphates (base-sugar-phosphate). If an extra phosphate is joined to the monophosphates, nucleoside diphosphates ADP, GDP etc. are formed. Addition of one more phosphate in the same way leads to the nucleoside triphosphates like ATP, GTP etc which are ribonucleoside triphosphates. Similarly, dATP, dGTP etc. are called deoxyribonucleoside triphosphates. These triphosphates were known to be quite energy-rich. So Kornberg used the triphosphates for DNA synthesis in the test tube. Another requirement for such synthesis did not escape his sharp eyes. As a genetic material DNA has to replicate itself, naturally by copying the two strands. So he added DNA to provide the template for DNA synthesis. Metal ion (like magnesium)

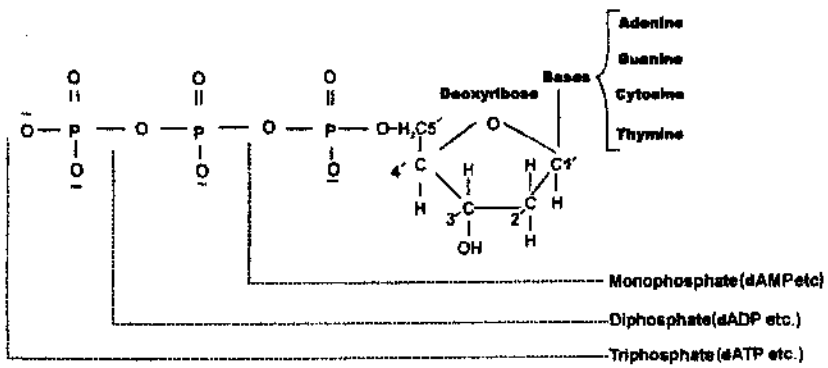



Fig. 1.10 Structure of mono-, di- and triphosphates of DNA bases. Mono-, di- and triphosphates of deoxyadenosine, deoxyguanosine etc.

was known by that time as an essential requirement for many reactions involving nucleoside phosphates. So he added $Mg^{++}$ in the incubation mixture. DNA polymerase was the name justifiably coined by him for the enzyme isolated from a bacterium *Escherichia coli.* Many of the enzymes have the suffix 'ase'. Since the enzyme which catalyses the synthesis of DNA, a polymer of deoxnucleoside phosphates, it was designated as DNA polymerase. Thus, the synthesis of DNA in a test tube, mimicking nature was achieved. The first enzyme isolated by Kornberg was later found not to be responsible for DNA synthesis under physiological conditions. This enzyme is, however, used extensively as a tool in genetic engineering. Later, he isolated the specific enzyme responsible for DNA synthesis in the cell.

The reaction for the synthesis of DNA can be written in simple chemical terms as shown below and illustrated in (Fig. 1.11). This was a miraculous achievement, synthesis of the genetic material in a test tube by Kornberg.

$$ndATP + ndGTP + ndCTP + nTTP + DNA(template) + Mg^{++} \rightarrow d(AGCT)_n$$

It has already been mentioned that the two strands of DNA can be either parallel or antiparallel. Josse working in Kornberg's laboratory and using his synthetic method demonstrated for the first time that the strands are actually antiparallel as shown in Fig.1.9. No wonder that Kornberg got the Nobel Prize for the synthesis of DNA in 1960. However, shortly after isolation of DNA polymerase it was realized that the enzyme cannot carry out DNA synthesis within the cell as the rate is very slow in comparison to the rate at which DNA is actually synthesized. Further, it can copy only short stretches of DNA and fails to do that for long chains. Genetic studies also indicated that there are several genes involved in DNA synthesis, which is rather quite complicated. A breakthrough came through the isolation of a mutant (designated as $pol^-$) of *E.coli.* which produces defective polymerase but carries out DNA synthesis normally. This was achieved by John Cairns, then working at Cold Spring Harbor, for which he

had to analyse thousands of mutants. A search for the proper enzyme for DNA synthesis led to the isolation in 1970 of two more enzymes (Polymerases II and III ) by Kornberg himself. Accordingly, the first enzyme was designated as Polymerase I, which constitutes 90% of the polymerase activity of the cell. Pol II acts mostly as a repair enzyme correcting the wrong substitutions of bases in DNA. Pol III which is much more complex, is the enzyme directly responsible for DNA synthesis but requires a number of factors to achieve it. Pol I is indirectly involved in DNA synthesis and has other important functions. The global problem of DNA synthesis appeared to be solved but the problem of both the strands to be copied in this fashion did not seem feasible. A phenomenon discovered by Okazaki, one of the Japanese students of Kornberg, set the stage to solve this problem. Incidentally, Okazaki died at an early age as a victim of 'atomic bomb' dropped at Hiroshima (and Nagasaki). He did not die as a direct target of explosion but indirectly, suffering much more as a victim of radiations. He discovered that the newly synthesised DNA was in fragmented form, or, in other words, the synthesis of DNA was discontinuous. This solved a very vexed problem of replication of double stranded DNA, the strands of which are antiparallel, as mentioned earlier. The synthesis of the strands starts from the 5′ end as expected as the nucleosides with 5′ phosphate are to be added to the 3′ end (OH group of deoxyribose) of the preceding nucleotide and the chain would thus elongate in 5′ to 3′ direction Fig 1.11. Thus the two strands have to be replicated in the same way from both the ends. That requires unwinding of the two strands (which are unusually long) from both the directions. There won't be enough space in the cell to do the job. The much simpler situation will be to unwind the two strands from one end only and copy the two strands from opposite directions from the fork as depicted in Fig. 1.12. Subsequently, it was shown that the replication actually proceeds from one end only in both the directions, as illustrated in the figure. The formation of the fork during the synthesis could be visualized by electron microscopy.

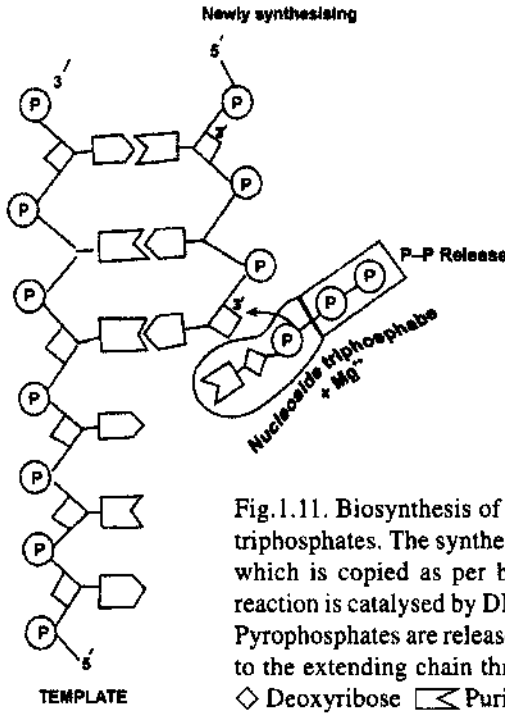



Fig.1.11. Biosynthesis of DNA from dioxyribonucleoside triphosphates. The synthesis takes place on DNA template which is copied as per base complementarity rule. The reaction is catalysed by DNA dependent DNA polymarase. Pyrophosphates are released and nucleotides are connected to the extending chain through phosphodiester bridge.
◇ Deoxyribose ⊏≺ Purine ⊏⊐ Pyrimidine

With the discovery of the Okazaki fragments the problem of replication was no doubt solved but an enzyme had to be found which could ligate the fragments to make the chain continuous. The DNA ligase eventually discovered by Kornberg himself and a few others was shown to do the job. This enzyme also became a valuable tool in genetic engineering. Another intricacy about the mechanism of synthesis of DNA was learnt a little later. An RNA primer (a short stretch of oligoribonucleotide) was found essential to start the DNA synthesis. Before the start of DNA synthesis, a number of ribonucleotides complementary to the 5′ end of the DNA chain are added. Once this short RNA primer is synthesized on the DNA template (by an enzyme called primase) the first deoxynucleotide is added to the so-called

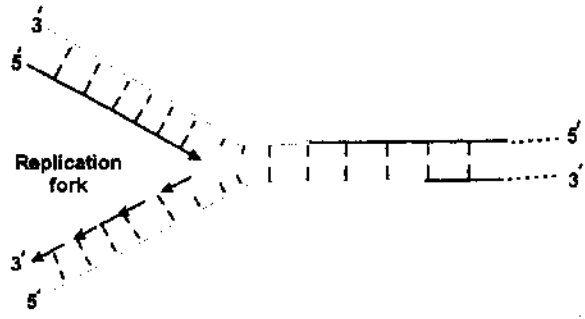



Fig. 1.12. The mode of replication of the double-stranded DNA. The strands are partially opened up at one end and both the strands start replicating but in opposite directions (5′→3′) as shown above. Thus, a fork type of structure is created at the starting point of replication. The synthesis, however, is not continuous and takes place in discrete fashion (fragment wise). *Okazaki fragments* are thus formed and ligated by DNA ligase.

primer RNA and then the elongation of DNA chain continues (Fig. 1.13). The enzyme, primase, necessary for synthesis of RNA on DNA strand was discovered later. Subsequently, this RNA primer has to be removed by hydrolysis and replaced with DNA. As expected, it was found that Okazaki fragments are preceded by short RNA chains and joined together to form the full length DNA chain after RNA primers are removed. The mechanism of DNA synthesis however, is not so simple as described here. For example, the situation is somewhat more complicated in case of synthesis of the bacterial genomes, which are circular, without any ends. There is a position known as origin of replication from which replication starts in both the directions (Fig. 1.14) and finishes at almost vertically opposite position on the circular genome. But it should be remembered at this stage that Kornberg detected the DNA polymarase in a bacterium *Escherichia coli* but it was established later that similar enzymes occur in all living systems, plants, animals etc., and the basic mechanism of DNA synthesis is the same through out. The mechanism of DNA synthesis within the cell

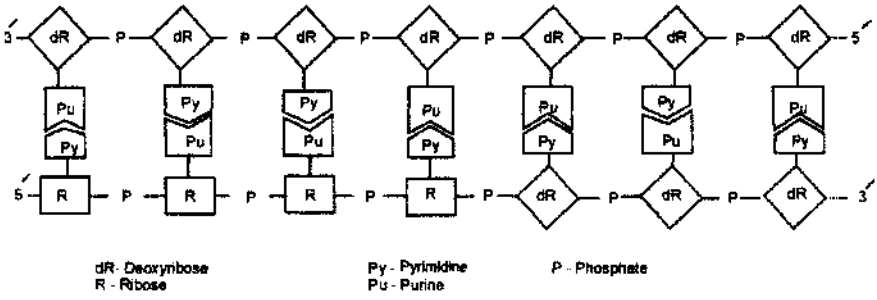



Fig. 1.13 Synthesis of DNA by the extension of RNA primer synthesised on the DNA template. Okazaki fragments are preceded by short RNA chains which are eventually hydrolysed and replaced with deoxyribonucleotide chains.

is much more complicated than described here. A number of factors and steps are involved in this process. Kornberg himself has described the intricacies of the process in his well-known book[7]. The situation is much more complicated in animals and plants.

As already mentioned, RNAs like DNA are composed of four bases, adenine, guanine and cytosine are common to both, whereas RNA contains uracil instead of thymine, which is, methylated uracil. RNA may sometimes contain other unusual bases but those are not very common. RNA chain like DNA is formed by phosphodiester bridges between ribose (and not deoxyribose as in DNA) and the bases are attached to 1′ carbon of ribose. Nomenclature is the same as in DNA (already discussed). The primary structure of the two are similar and can be written in the same way as shown below (U in place of T in DNA)

5′ AGUCCGAA.....CGAUUGCA 3′

The two ends are designated 5′ and 3′ as in case of DNA. The linkage between the two phosphates is through the 3′-5′ Phosphodiester Bridge as in DNA. However, in RNA there is an additional OH group in the 2′ position whereas in DNA there is only H at that position, as already mentioned.

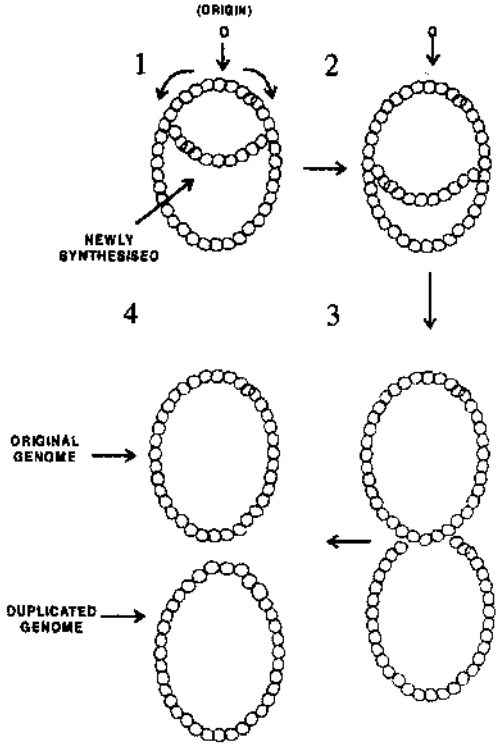



Fig.1.14 .Mode of replication of circular bacterial genome from the origin of replication. The replication starts in both the directions simultaneously from the origin and the replication ends at the position vertically opposite to the start point.

However, there are quite a few differences between the two in higher order structures. DNA is usually double-stranded although some bacterial viruses (bacteriophages) may contain single-stranded DNA. RNA normally has single-stranded structure although double-stranded RNA is found as the genetic material in some viruses, specially plant viruses. The double-stranded RNAs are formed when necessary by the same Watson-Crick base pairing rule as in case of DNA, the only difference being that adenine pairs with uracil (instead of thymine). However, single-stranded RNAs may also form double-stranded structure by flipping over itself (Fig. 1.15a) and the structure may be very similar to the double-stranded helix of DNA. To form such double stranded structure some bases have to loop out (there

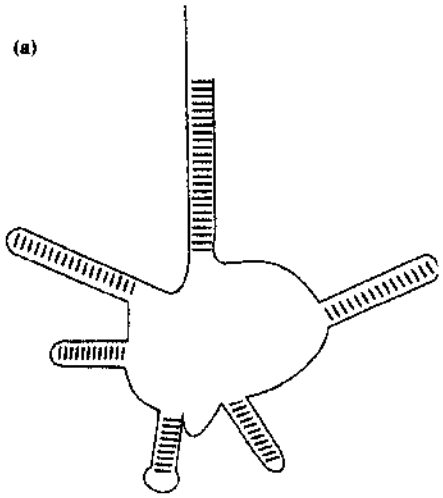



Fig 1.15 (a) Stem-loop structure present in RNA. Generally RNA is single-stranded but it can fold on itself due to the base complementarity of stretches of oligoribonucleotides on the same chain. For this, the chain between the two complementary regions has to loop out. The double-stranded region(stem) has DNA– like structure but in A form as shown in (b).

(b)

(b) Space-filled model of A type of structure of RNA. There are other types of structures as well.

may be small loops, and sometimes large loops as well). This type of organisation is known as stem-loop structure. Such stem loop structures are the basis of the secondary structure of RNAs. The stem part assumes more or less the A form of DNA structure and is known as A-RNA configuration (Fig. 1.15b).

As we progress we shall realise that there are varieties of RNAs of small and large sizes occurring in the cell and those naturally carry out various functions (one function which we have already mentioned, is to act as the genetic material like DNA). Their three-dimensional structures are rather complicated. X-ray crystallography has, however, helped to determine these structures, not only of small size RNAs but also large size RNAs as well. One point is absolutely clear that all types of RNAs are synthesised in the cell on DNA template. One of the two strands acts as a template and RNA is synthesised on that by copying the strand according to the base-pairing rule.

## References

1. *What is Life? (The physical aspect of the living cell) and Mind and Matter.* Erwin Schröndinger, Cambridge University Press, 1967
2. *The double helix (A personal account of the discovery of structure of DNA).* James D.Watson, Atheneum, New York.1968.
3. *The path to the double helix.* Robert Olby, University of Washington Press, Seattle, 1974.
4. *Heraclitean fire (Sketches from a life before nature).* Erwin Chargaff, The Rockefeller University Press, New York, 1978
5. *Principles of genetics.* E. J. Gardner, John Wiley and Sons, New York, 1960
6. *The Professor, the Institute and DNA.* Rene J. Dubos, The Rockefeller University Press, New York, 1976.
7. *DNA replication.* Arthur Kornberg and T. A. Baker, W.H. Freeman & Company, New York 1991.

*****

# Chapter - 2
# How is the music created?
## (Structure and Function of Proteins and the Protein Synthesing Machinery)

*If we accept that DNA plays the music of life (Chapter 1) the question arises, how is the music created? The music is actually the coordinated expression of the innumerable functions of the living entities, responsible for their maintenance, reproduction and genetically programmed death. Proteins may be looked upon as the building engineers of the cell, primarily responsible for a variety of functions. Amino acids are assembled in correct sequence on the ribosomes, the protein manufacturing machinery of the cell, composed of both RNA and protein, and linked by peptide bonds to give rise to proteins. Another class of RNAs acts as handles to bring the free amino acids present in the cytoplasm of the cell to the ribosomes, the site of protein synthesis. The synthesis and coordinated functions of proteins with the help of nucleic acids (other than DNA) of the living cell create the music of life. Therefore, the history of understanding of the complicated machinery of the living cell has been narrated in a tune somewhat different from what is normally done.*

There was a small group seminar in the Department of Microbiology of the New York University School of Medicine and the speaker was none other than Fritz Lipmann. Frankly speaking, I don't remember the topic of the seminar as it was held sometime in 1960 or '61. As soon as he started, Aaron Novick, a well-known microbiologist, asked something which the speaker clarified. Before Lipmann could proceed further Novick asked another question. Lipmann though visibly

irritated, replied to that as well. Before Novick tried to interrupt for the third time a little later by raising an issue with the words "May I raise _____'. Lipmann angrily said 'No'. After that Novick did not raise any more question throughout the lecture. I blame none of them. As such Lipmann was pleasant, especially with those who wanted to discuss science with him. But he could get easily irritated in simple matters like the incident cited above. Although he was not a good speaker he conveyed the message diligently. He was highly active in science even during his last days of life.

Fritz Lipmann was a great biochemist. If Otto Warburg could be the Father of modern biochemistry, Lipmann should be called its 'Foster Father'. His conceptual approach was as good as his experimental approaches. He was one of the first to visualize the involvement of energy in protein synthesis. The term "protein" is derived from the Greek word 'Proteus' indicating, 'of the first rank'. They are the important functional components of the cell. Proteins were well characterized long before the genetic material (DNA) was isolated. Biochemistry actually started with proteins, specially because some of those act as catalysts of the numerous chemical reactions occurring in the living systems and are known as enzymes. Historically speaking when J.B. Sumner declared in 1926 that urease, an enzyme

**Table 2.1**

**Enzymes as powerful catalysts**

| Enzymes | Rate enhancements* |
|---|---|
| Carbonic anhydrase | $10^7$ |
| Phosphoglucomutase | $10^{12}$ |
| Succinyl CoA transferase | $10^{13}$ |
| Urease | $10^{14}$ |

* In comparison to rates in absence of catalyst

which degrades urea (hence the name), was a protein, no one believed it. The mystery of enzymes was hidden behind its existence in trace amounts and their tremendous catalytic power (Table 2.1). Liebig, the great chemist, ridiculed Sumner for his fantasy.

Proteins are composed of 20 different amino acids. These amino acids were originally abbreviated by three letter symbols as shown in Table 2.2. Later a one-letter code (shown in the same table) was devised for the sake of convenience. There may be a chain of hundreds to several thousands of amino acids in a protein which are connected by peptide bonds. The primary structure of a small protein, the hormone insulin which controls blood sugar level is shown in Fig.2.1. Insulin has two chains (A and B) connected through sulphur bridges. This is the first protein sequenced by Frederick Sanger, for which he got the Nobel Prize in 1958. Sanger took almost a decade to sequence insulin. Now-a- days it will take a day or two to sequence a protein with the help of automatic sequencer.

Let us discuss a little bit more about proteins[1,2] at this stage, which

**Table 2.2**

**List of amino acids, their abbreviatons and one letter codes**

| Amino acid | Abbreviation | Code | Amino acid | Abbreviation | Code |
|---|---|---|---|---|---|
| Alanine | Ala | A | Leucine | Leu | L |
| Arginine | Arg | R | Lysine | Lys | K |
| Asparagine | Asn | N | Methionine | Met | M |
| Aspartic acid | Asp | D | Phenylalanine | Phe | F |
| Cysteine | Cys | C | Proline | Pro | P |
| Glutamic acid | Glu | E | Serine | Ser | S |
| Glutamine | Gln | Q | Threonine | Thr | T |
| Glycine | Gly | G | Tryptophan | Trp | W |
| Histidine | His | H | Tyrosine | Tyr | Y |
| Isoleucine | Ile | I | Valine | Val | V |

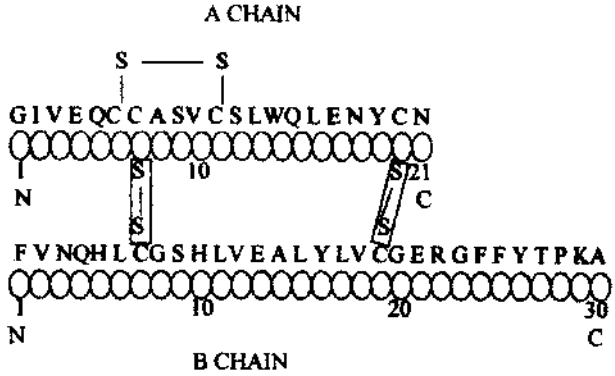



Fig. 2.1. Sequence of amino acids in insulin. The protein hormone is composed of two chains, chain A (21 amino acids) and chain B (30 amino acids). The two are connected through S-S bridges. There is an internal S-S bridge in A chain. Below the chains N and C terminals are indicated.

are the vital constituents of the cell. As mentioned above, proteins are composed of 20 odd different amino acids. As the name indicates, an amino acid has an amino ($NH_2$) group (proline is an exception, let us ignore that at this stage) and an acidic carboxyl (COOH) group. The basic structure of each and every amino acid occurring in protein is as follows. One tetrahedral carbon atom (marked '$\alpha$' due to its position next to COOH) is attached to 4 different entities (H, NH2, COOH and R) as shown in Fig 2.2 a. Since $\alpha$ carbon is attached to 4 different moieties each amino acid (except glycine where R is H) can occur in two different stereochemical forms, L and D varieties depending on the positions of the groups as shown in Fig 2.2 b.

R is the variable group (20 different R's for 20 amino acids) starting from H (as in glycine) and ending in a fused ring structure as in proline which is an imino acid and not an amino acid in the true sense (Fig.2.3). Let us not worry about details of that but should remember that amino acids occurring in proteins are invariably of L variety

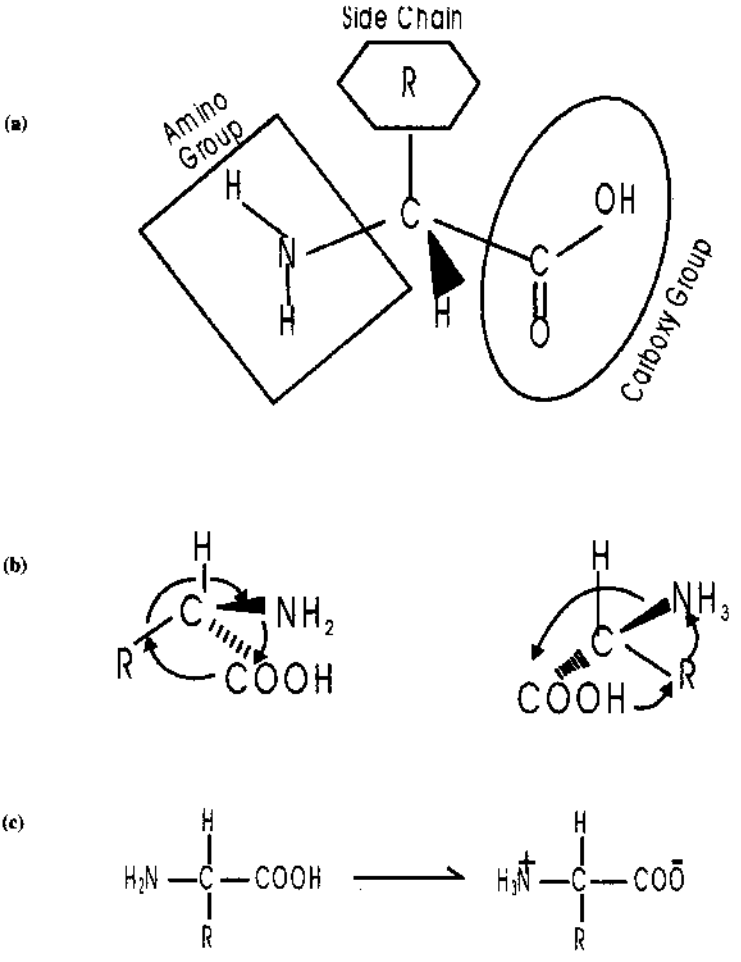



Fig. 2.2 Generalized structures of amino acids. (a) R could be H, $CH_3$, $CH_2OH$ etc. (there are 20 different R groups in 20 different amino acids occurring in protein. Since α carbon (carbon atom next to COOH group) is attached to four different groups (H, $NH_2$, COOH and R) each amino acid (except glycine) occurs in two stereoisomeric forms (b) L- and D-Amino acid (c) Zwitter ion form of an amino acid (on the right).

although D - amino acids do occur in nature.

An amino acid is expected to be electrically neutral as it has an acidic group and a basic group. However, an amino acid may have an additional amino group (or other basic groups) other than $\alpha$-$NH_2$ group (these are called basic amino acids like lysine, arginine see Fig 2.3). Similarly there could be additional COOH group as in aspartic acid and glutamic acid. These are known as acidic amino acids. Those, which do not have additional basic or acidic group like alanine, are known as neutral amino acids. Some amino acids like serine, threonine

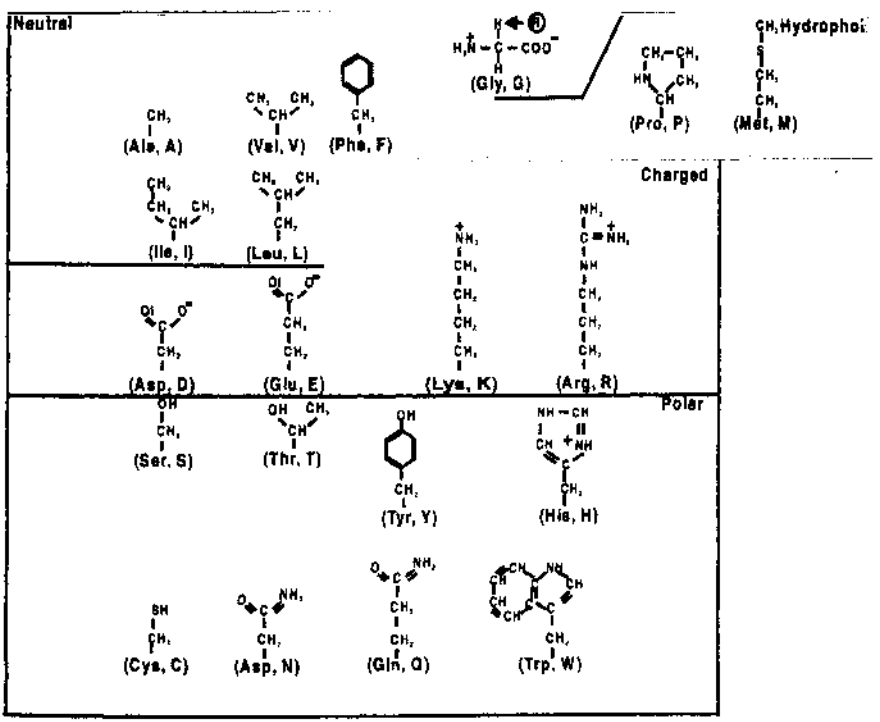



Fig. 2.3 Different R groups in 20 different amino acids occurring in protein. Each amino acid can be identified by its R group.

etc though apparently neutral (without any net change) are polar molecules where both positive and negative charges occur on the same molecule without charge neutralisation. Other than that amino acids can also occur in Zwitterions form (Fig.2.2c) depending on the pH of the medium. $NH_2$ will acquire a proton to become $NH_3^+$ and COOH will lose one to become $COO^-$. Those amino acids, which have an aromatic ring in the side chain, are sometimes grouped separately.

The molecular weights of all 20 amino acids grouped accordingly have been plotted in (Fig 2.4) using one-letter code. It is interesting to observe that nonpolar amino acids are usually small in size and those with the aromatic ring containing side chains are of larger size, as

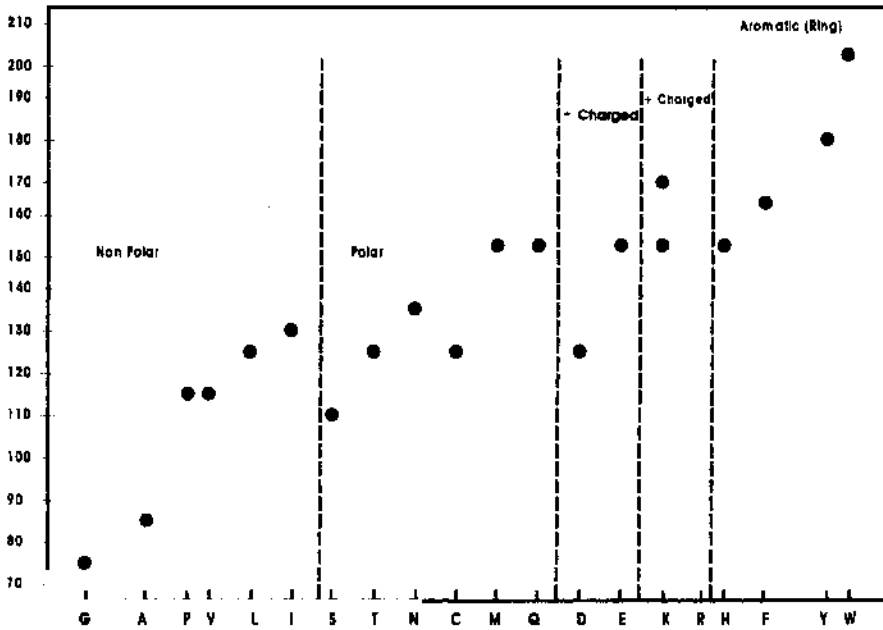



Fig. 2.4 Molecular weights of amino acids and the characteristics of the side chains (R groups). Different amino acids in one letter codes (Table 2.1) are written on the X-axis. The Y-axis represents molecular weights in Daltons.

expected. Polar and charged (positively as well as negatively) amino acids are in the middle range. This may be an indication of the order in which the amino acids emerged from simple to a variety of structures during the chemical evolution on Earth and also of natural selection.

Two amino acids are connected together by the interaction of NH2 group of one with the COOH group of the other with the elimination of water. The combined product of two amino acids is known as dipeptide, the linkage between the two being designated as peptide bond. Similarly tripeptide, tetra peptide etc., are formed by the joining of three, four or more amino acids. Proteins in this way can be looked upon as polypeptides. The amazing fact about proteins is the perfect sequence in which all the amino acids are arranged (just like the bases in case of DNA and RNA) discussed in earlier chapter. Even the change of a single amino acid may create a havoc as in sickle cell anemia.

To avoid writing the detailed structure of protein (as shown below)

Glycine-Alanine-Serine.....Tryptophan-Lysine

It can be simply written by naming the amino acids in sequence using the three-letter code (Table 2.2).

Gly-Ala-Ser.....Try-Lys

However, to simplify the writing a one letter code has been derived (Table 2.2). According to that the same protein chain can be written as follows

G-A-S.....W-K

In any mode of writing the convention of marking the two ends as in DNA and RNA is to be followed to avoid confusion. The first amino acid with the $NH_2$ group free should be written on the left marked as the N-terminal and the last amino acid with the COOH group free should be written on the right (C-terminal). In the above example glycine is at the N-terminal end and lysine is the C-terminal amino acid. If it has to be written in the reverse way N- and C- ends have

to be indicated. A similar convention in the case of DNA (based on free phosphate and hydroxyl groups) has been discussed in the earlier chapter.

Just like DNA, proteins do not remain as extended chains under physiological condition. Each protein folds in its own characteristic fashion in three-dimensional space. Some of them take more or less a globular shape (known as globular proteins), some may have an elongated form (ellipsoidal protein) and some others may be extended as thread-like or fibrous structure (known as fibrous proteins). These shapes arise due to the nature of the individual amino acids and interactions between them. The bonds normally present in proteins are shown in Fig 2.5. The most common interaction is van der Waals' forces (Fig.2.5a), which is a non-specific interaction between atoms, and molecules, which come close together within some specific distance. van der Waals interaction between atoms is a common feature of all molecules. There can be hydrogen bond (Fig.2.5b) as originally conceived by Linus Pauling. H itself acts as a bond if it is attached to an electronegative atom (N or O) and when there is another electronegative atom nearby. There may be other interactions as well, like hydrophobic interaction (Fig.2.5c). When the water-repellant groups in amino acids cluster together to give rise to a kind of bondage but not a true bond it is known as hydrophobic interaction. There can be charge interactions or salt bridges between free acidic and basic groups, as shown in Fig.2.5d. Sometimes two or more protein chains can be connected to each other through S-S bridge between two cysteine residues (Fig.2.5e) as in the case of insulin (see Fig. 2.1).

When some crystallographers like William Astbury were busy finding out the three-dimensional structures of proteins, Linus Pauling followed a simple strategy. At first he studied the structure of the simple di- and tripeptides by X-ray crystallography and found that atoms constituting the peptide bond C=O and NH lie in one plane and O and H are in trans-position (Fig.2.6). From this observation based

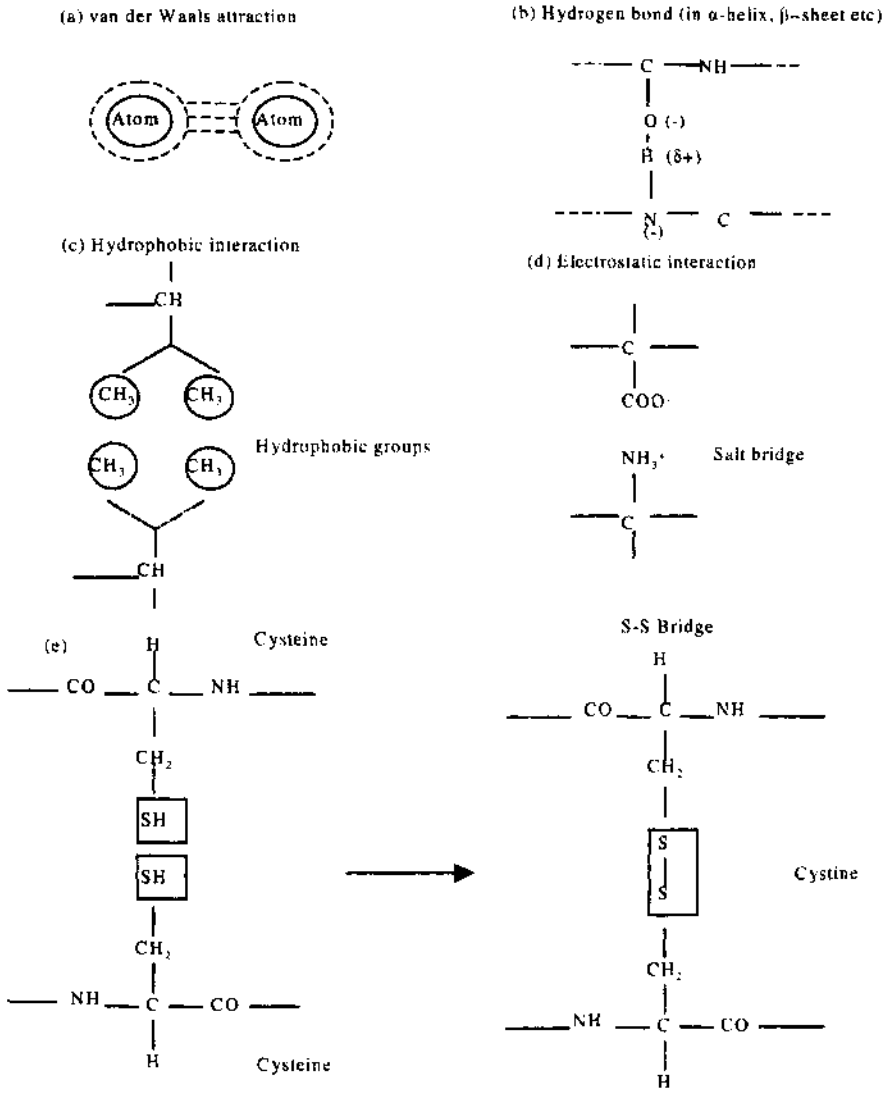



Fig. 2.5 Different types of interactions between the protein chains or parts of the same protein which lead to the three-dimensional structure of proteins. Details have been described in the text.

on the studies of peptides he proposed the α–helical structure of protein. According to his proposition (which was found to be absolutely correct), the protein chain takes a α–helical form in which C=O of the fourth amino acid forms the hydrogen bond with the first amino acid (Plate2, Fig.2.7) The angle between the carbon of the two peptides can be somewhat twisted to give rise to the alpha (α) helical conformation. One should remember that this helix is quite different from the double helical structure of DNA (Chapter 1). Other types of helices also occur in proteins but are quite rare. Another type of structure known as beta (β) sheet structure was also proposed by Pauling, which was due to hydrogen bonding as in helix but in a different way. The chain remains more or less extended but is hydrogen-bonded (in multi chain protein) as shown in (Plate 3, Fig.2.8). Two hydrogen-bonded chains may be (a) antiparallel (Plate 3, Fig.2.8a) or (b) parallel (Plate 3, Fig.2.8b) to each other. Most of the fibrous proteins have antiparallel β–sheet structure. There is another type of structure present in protein, which is known as beta (β) bend (Fig 2.8c). As the name indicates the chain at β–bend abruptly takes a turn (U turn) and helps to form the anti parallel sheet structure

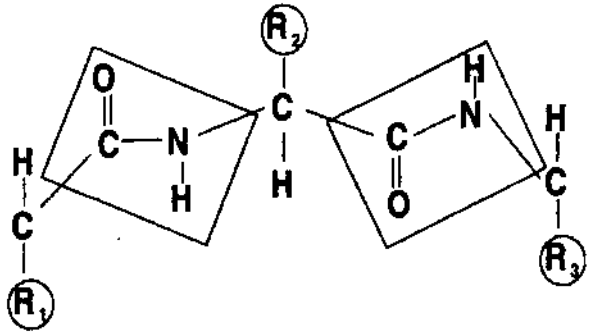



Fig. 2.6 Planar nature of a peptide bond. All the four constituent atoms of the peptide linkage (C, N, O and H) lie on one plane. Oxygen (O) and hydrogen (H) atoms are in trans-position, those being on opposite sides of C-N linkage.

(Fig. 2.8b). Usually proline (a ring structured amino acid without free NH2 group (as shown in Fig. 2.3) is found at the bend which helps to stabilize the structure. Originally it was believed that globular proteins like myoglobin and haemoglobin contain α–helical structure but as many more X-ray structures were worked out it was observed that β–sheet structures predominate in many cases and the helical structure is not so common. Usually a protein has a combination of different types of structures including extended structures without having any ordered structure.

The three-dimensional folding of a protein molecule, which is the resultant of the various regional structures, is known as the tertiary structure of protein. The sequence in which amino acids are present is known as primary structure whereas α–helical, β–sheet, β–bend etc. lead to the secondary structure (Plate 2, Fig. 2.7), (Plate 3, Fig 2.8). When a protein has more than one chain (referred to as subunits), like hemoglobin, which has four chains, (two α and two β) the mutual interaction (again of various types, salt bridges, hydrogen bonding, hydrophobic interaction etc.) lead to the overall structure of the protein known as quaternary structure. Myoglobin is a single chain protein present in muscles and is responsible for the supply of oxygen to the muscle cells. It consists of a single chain similar to the α–chain of hemoglobin. The tertiary structure of myoglobin and that of α–chain of hemoglobin are similar (Plate 4, Fig. 2.9a). As already mentioned, the quarternary structure of hemoglobin consists of four chains as shown in Plate 4, Fig. 2.9b. Ferrous iron ($Fe^{++}$) present at the centre of each chain is responsible for binding oxygen and its subsequent delivery to the tissues. It is obvious that the varieties of structures in proteins indicate a variety of performances. Therefore it was very apt to coin the word protein (which indicates "of the first rank" in Greek) even in early times. The structure of an enzyme pyruvate kinase which is a protein like most of the enzymes, consists of both α–helical and β–sheet structures (Plate 4, Fig.2.9c).

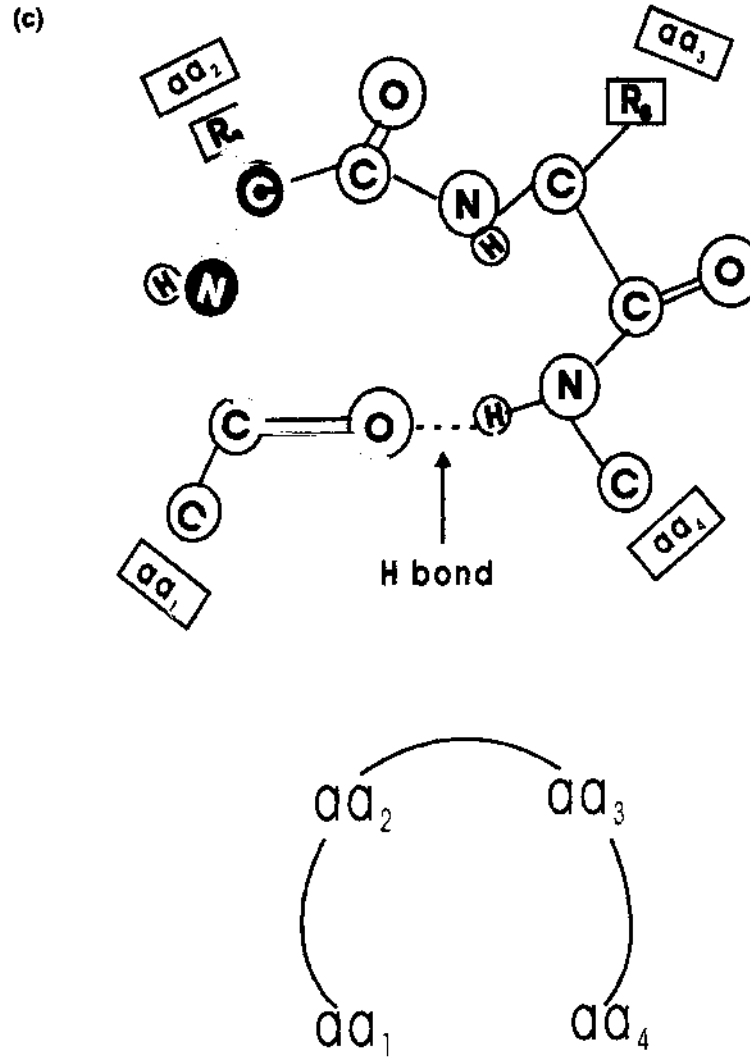



Fig. 7.8 *continued* from Plate 3 (Fig 2.8 a and b) (c) β–bend structure present in protein helps to form the antiparallel β–sheet structure. When a protein chain takes a sharp turn (β–turn) it makes a β–bend. Usually four amino acid residues are present in a β–bend. Proline also helps to make the β–turn.

Max Perutz, who died recently, was the father of modern X-ray crystallography. Following Pauling's classical demonstration of a helical structure in proteins Perutz's solution of phase problem in X-ray crystallography by isomorphous replacement with heavy atoms was the next outstanding achievement, which helped the X-ray crystallographers to solve the structure of proteins readily. When he undertook the challenging job of showing the X-ray structure of haemoglobin, (the oxygen-binding protein present in red blood cells having two α–chains and two β–chains, each of molecular weight of 17,000 daltons) it was only possible to solve structures of molecules of molecular weight of ~100 daltons or so. Sir William Lawrence Bragg (son of famous William Henry Bragg), then the Head of the Medical Research Council Unit for the study of molecular structure induced him to this challenging problem. It took painstaking work of almost two decades to finish this Herculean job. Pertuz was joined later by John Kendrew to solve the structure of myoglobin (a single chain oxygen-binding protein in muscle) and shared the Nobel Prize with him in 1962. Following their tremendous achievement, X-ray crystallography came to the doorstep of X-ray crystallographers and the structures of numerous proteins were solved one after another. The main idea of Perutz behind this work was to understand the functions of the protein, specially the mode of binding oxygen to hemoglobin. This opened a new chapter in the understanding of structure-function relationship of macromolecules, in general.

Proteins carry most of the functions of the cell. On one hand, they are constituents of the variety of structural elements of the cell. While on the other hand, they carry out numerous biological functions of the cell, which are essential for cell growth, cell division as well as programmed cell death. Regarding the structural elements one may think of transport vesicles like endoplasmic reticulum, Golgi apparatus, nuclear envelope and a variety of small-interfolded vesicles which are in flux with the plasma membrane constantly forming and reforming. A cross section of animal cell structure is depicted in

Fig.2.10. All these organelles have proteins as constituents. Cellular membranes, nuclear membranes and other types of membranes also invariably have some protein as structural and some as functional elements. Fibrous proteins like fibrin, keratin, etc. are distributed all over the biological systems as structural elements. The most important biological function, catalysis of numerous chemical reactions occurring in living system is carried out by proteins classified as enzymes. Catalytic functions are usually very specific, so a large number of enzymes occur within the cell. Enzymes were originally thought to be invariably protein molecules. This idea has been recently changed as some RNA molecules were later shown to have catalytic functions for which they are called ribozymes (ribo indicates ribonucleic acid). Besides the catalytic function proteins have numerous other functions for e.g., both hemoglobin present in red

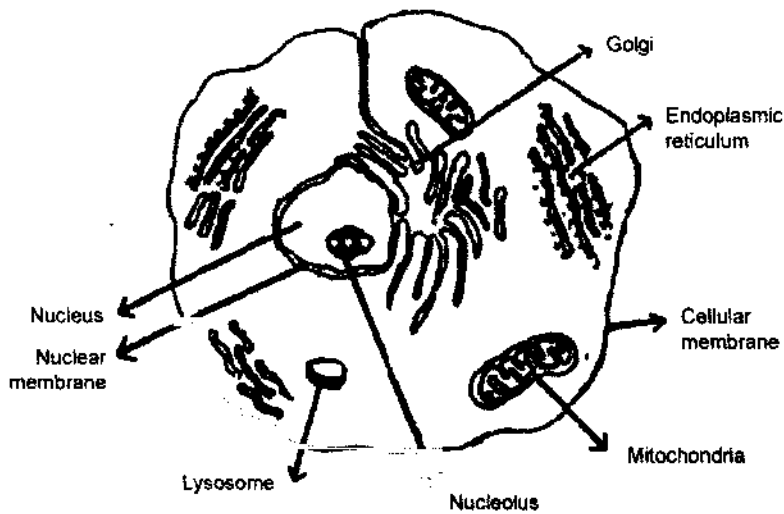



Fig 2.10. Structure of an animal cell. There are a number of organelles like nucleus, endoplasmic reticulum, Golgi apparatus, vacuole etc., of which some are in constant flux with the cellular membrane.

od cells and myoglobin present in muscle bind oxygen. Insulin reted by pancreas acts as hormone in glucose metabolism. rmones (not necessarily proteins) are regulatory factors of chemical ctions under physiological conditions. Both myosin and actin sent in the muscle are involved in muscular contraction. Proteins o act as receptors located in the membrane and help in the transport materials from the outside to the inside of the cells. munoglobulins known as antibodies which are part of the defence chanism against invading foreign materials (antigens) are protein olecules. The examples can be piled up. Therefore, from the very ginning of their discovery those were recognised as most important mponents of the cell.

It is obvious from the above discussion that the proteins are stined to carry out a variety of functions in the cell. That is rroborated by the presence of 20 odd different amino acids with a riety of side chains. It has already been discussed in Chapter 1 that e nucleic acids (both DNA and RNA) are made up of only four fferent types of basic units (nucleotides). Even that small number roduce a variety of nucleic acids starting with small size transfer NAs to DNAs of various sizes (containing a few to billions of ucleotides) depending on the chain length. Even then the functional ttributes of both are somewhat limited compared to those of proteins. here are also two basic differences between the organisation of the wo types of macromolecules. Proteins usually are multicomponent nd may have more than one chain in a molecule, which nucleic acids ormally do not possess. Further, molecular weights of the proteins nay vary over a wide range (a few thousands to a million daltons or o) and their chain lengths are short, when compared to the chain engths of DNA. The compromise on the size is achieved in the rganisation of the molecules, so far as the charge distribution, ydrophobicity, recognition of a variety of other molecules etc., are concerned. Nucleic acids usually have monotonous type of structural

organization (stem-loop type) while proteins have a variety of organizations (although the basic structures like α–helical, β–strand etc are somewhat limited). The three dimensional structure of a protein is perhaps its key to success as far as its function is concerned. Before going into details about the intricacies it may just be mentioned at this stage that the information written on nucleic acids (specially DNA) may be treated as one-dimensional whereas that on the proteins is actually three-dimensional. Thus the one-dimensional information of DNA is translated into the three-dimensional structure of proteins.

When I started my career as a biochemist in the 1950's some experts in biochemistry were busy proving that proteins are synthesized from the amino acids by the reversal of the degradation process. Several proteolytic enzymes, which catalysed the degradation of proteins by rupturing peptide bonds with the help of water were known by that time. It was assumed that the same enzymes catalyse the formation of peptide bond between the amino acids working in the reverse direction. This type of mistake was repeatedly done from time to time by many biochemists. Carl Cori and Getri Cori (Nobel Prize winning couple) thought for a long time that glycogen phosphorylase which catalyses the breakdown of glycogen (present in liver and muscle) to glucose-1-phosphate, synthesises glycogen from glucose-1-phosphate by catalysing the reaction in the reverse direction. He was eventually proved wrong by Louis F. Leloir of Argentina. The synthesis requires the involvement of energy. Uridine diphosphate glucose (UDPG) is formed from UTP and glucose-1-phosphate. The energy rich UDPG synthesises glycogen by donating glucose to the existing glycogen acting as a "primer". Let us not worry much about this reaction. However, it is difficult for me to check the temptation of not saying a few words about Leloir. He is the glaring example of a scientist from the third world who did a Nobel Prize winning work under the most adverse circumstances in a poor country. While he visited Bernie Horecker's laboratory at the National Institutes

of Health in 1956, he became quite impressed with the automatic fraction collector, which I was using to collect the fractions from a column used for separation of proteins. That was a kind of luxury for developing countries those days!

A similar mistake as in the case of glycogen synthesis, as discussed above, was made with fatty acid synthesis. Short chain or long chain fatty acids are degraded in a stepwise fashion, finally to acetyl coenzyme A. Coenzymes (coenzyme A being one of these) help the enzymes in their catalytic function. For a long time it was believed that fatty acids are synthesised by reversal of the degradative pathway. This was proved wrong when it was established that the synthetic pathway is entirely different. It involves energy and malonyl coenzyme A, an energy rich compound as an intermediate. Even synthesis of RNA is another glaring example. History repeated itself in case of protein synthesis. Lipmann was the first one to predict that protein synthesis cannot take place by the reversal of degradative pathway it required energy. The great biochemist contributed extensively in the area of chemical energy (metabolic) and mechanism of activation with the help of energy-rich compounds like adenosine triphosphate (abbreviated as ATP), guanosine triphosphate (GTP) etc. It may be remembered that DNA is synthesised from dATP, dCTP, dGTP and TTP, all energy-rich compounds (Chapter 1).

Paul Zamecnik, then at the Massachusetts General Hospital followed the *dictum* of Lipmann and started to work with cell-free extract (free from any intact cells) from rat liver keeping the energy picture in mind. Lipmann at that time was in the Harvard Medical School. He was later associated with the Rockefeller University. Earlier Borsook did a few preliminary experiments by feeding radioactive amino acids to rats. Zamecnik performed those experiments in a different way. He injected radioactive amino acids to the animals, which were killed at different intervals following injection. Cell-free extracts were prepared from the liver and then

fractionated by differential centrifugation. The idea was to trace the path of amino acids with a radioactivity tag. At the earliest time of the experiment radioactivity was found to be associated with the microsomal fraction. Microsomes are the fragmented material of the endoplasmic reticular (ER) structure, which is a part of the network in the cytoplasm as mentioned earlier. Those sediment at 100,000xg. So, it was assumed that the microsomes are the sites of protein synthesis. It was found later that some particles attached to the reticular structure are actually the sites where protein synthesis takes place. This will be discussed in greater detail in Chapter 4. I would like to introduce a personal touch at this stage from an article of mine.

"Prior to my serious involvement in `ribosomology', my association with ribosomes was just `on' and `off' and that also in an indirect way. This article documents my interactions with ribosomes and humble contributions in the area, which may be of some interest to the younger generation and mere curiosity to the elders. R. H. Burris and P.W. Wilson were the pioneers in the studies on the mechanism of nitrogen fixation (how do some plants and bacteria fix nitrogen directly from the atmosphere). They mainly chose *Azotobacter vinelandii*, the free-living nitrogen-fixing microorganism for the study and nitrogen enriched with $^{15}N$ (heavy isotope of nitrogen) as tracer. My interest in joining Burris's laboratory in 1955 was to learn mass-spectrometry through the use of $^{15}N$. I did learn the technique with a primitive equipment (very advanced in those days!) in his laboratory and could use nitrogen enriched with 60 atom percent $^{15}N$ but never had the opportunity of using it later as the radioactive $^{14}C$ (radioactive isotope of carbon) was readily available and served the purpose of most of my future studies. Radioactive tracer technique was very much in vogue at that time and even today. In Madison, I was associated with a project whose objective was to trace the path of ammonia in the form of $^{15}NH_3$ (the first detectable product of nitrogen fixation) in bacteria leading to the amino acids, bases, proteins, nucleic acids etc. The project was planned for a period of one year and sanctioned

accordingly by the National Institutes of Health, USA. Therefore, I could not afford to stay in his laboratory for more than that time period. When I finished the project within six months or so he allowed me to work for the remaining period in any area of my choice. Zamecnik and his coworkers had by that time standardized the method of measuring protein synthesis in the rat liver extract with $^{14}$C-amino acids as tracer. They separated the extract into two fractions at acid pH. Both the precipitated fraction and the supernatant were necessary for the incorporation of radioactive amino acids into trichloroacetic acid-insoluble material (containing mostly protein). The precipitate formed at acid pH (called pH 5 fraction) contained microsomes (fragmented endoplasmic reticular structure with attached ribosomes, a term coined in 1959). I was fascinated by his preliminary findings and chose to work with the cell-free extract of *Azotobacter vinelandii.* I believe that was the first report on protein synthesis with a microbial extract. Although Zamecnik was a pioneer in this area his unfortunate choice of rat, an eukaryote was the bottle-neck. The use of *E.coli* (a prokaryote) by others following his early observations with rat liver led to the quick understanding of the protein synthesising machinery. Even after my return to India I continued the work but using $^{14}$C-amino acid as tracer. This work carried out with the help of my would-be wife, Maharani Chakravorty was presented by me in one of the Federation meetings in a session on protein synthesis chaired by Zamecnik himself. By that time it was already realised that ribosomes attached to the microsomes are the sites of protein synthesis. He was quite surprised to learn that the work was done with a microorganism in India, a developing country. Let me now quote from the Master himself from a paper published in the Journal of *Biological Chemistry* (Vol 231 No.1 March, 1958 pp 533-549). "The cell-free rat liver system in which $^{14}$C-amino acids are incorporated irreversibly in peptide linkage into protein has been used in our laboratory for a number of years as a measure for protein synthesis. The essential components of this system are the microbial ribonucleoprotein particles, certain

enzymes derived from the soluble protein fraction, adenosine triphosphate, guanosine di- or tri-phosphate and a nucleoside triphosphate-generating system (references cited are all in 1954-1957 issues of *Journal of Biological Chemistry*). The ribonucleoprotein particles of the microsomes appear to be the actual site of peptide condensation. The soluble enzymes and ATP have been found to affect the initial carboxyl activation of the amino acids (1956 issue of the same journal). The role of GTP is not yet understood although the present paper sheds light on its probable locus of action."

So Zamecnik started with a big bang. But a more prophetic version with one unavoidable mistake was in the concluding section of the paper. The quotation continues as below.

"We have suggested elsewhere (reference to Current activities in molecular biology, in press) a hypothetical reaction sequence for protein synthesis which accounts for the findings presented in this paper. The central idea is that pH5 RNA molecules were each charged with amino acids in characteristic sequence, by polymerase in microsomes (*specific order determined by the complementary structure of microsomal RNA*) to higher molecular units with resultant changes in configuration which permit peptide condensation between contiguous amino acids. This working hypothesis will form the basis for further studies in these laboratories on the mechanism of protein synthesis." The italics are mine and their conjecture was entirely wrong. Zamecnik would have got the Nobel Prize eventually but he was unfortunate to choose the rat liver system. As already mentioned, rat being a eukaryote is much more complicated than a prokaryote like *E.coli*. He started on the right path and had all the requisite information to complete the story. Other scientists chose the simple microorganism *E.coli* and quickly completed the story by tying the loose ends. However, it took a number of years to work out the basics. The mistake (as evident from the concluding lines) was that amino acids are arranged in sequence on microsomal RNA (through pH5

RNA molecules). He is not to be blamed, as messenger RNA was unknown at that time. For a long time some veteran researchers in the area mistook ribosomal RNA as messenger RNA even when the structure of ribosomes became well known.

Let me elaborate a little of Zamecnik's early experiments. He started with the extract of rat liver cells and by lowering the pH of the extract to 5 by acetic acid he got a precipitate, which he called pH 5 fraction. He used radioactive amino acids to study their incorporation into proteins (which were known to be trichloroacetic acid-insoluble). So proteins could be separated from free amino acids, which were soluble in trichloroacetic acid. As mentioned earlier, there was a general belief at that time, that the enzymes which are responsible for the hydrolysis of proteins (proteases) are responsible for protein synthesis also but it was Lipmann, the Nobel Prize-winning biochemist who pointed out that energy would be required for the synthesis of proteins. One of the most well known energy sources at that time was adenosine triphosphate (abbreviated as ATP). So Zamecnik and his coworkers routinely added ATP to such system and later they showed that it required GTP (guanosine triphosphate) as well (as mentioned in their classical paper). It was Lipmann again who clarified the function of GTP. Very soon Zamecnik found that both the pH5 precipitate and the supernatant (soluble fraction) were required for protein synthesis. With the help of Mahon Hoagland and others, he demonstrated that the pH5 precipitate contained microsomes or fragmented endoplasmic reticulum. As it will be discussed later, actually ribosomes having both RNA and protein attached to the reticular structure are required for protein synthesis. RNAs present in the supernatant were called soluble RNA (transfer RNA in the present day nomenclature, as their function is to transfer amino acids from free pool to the site of protein synthesis). The proteins (rather enzymes) present therein were also required for protein synthesis. The latter became known as the amino acid activating enzymes. They clearly demonstrated that amino acids are activated with ATP as catalyzed by amino acid activating enzymes and

then transferred to soluble or transfer RNAs. Each specific amino acid has a specific activating enzyme, which also transfers the activated amino acid to a specific transfer RNA by the two-step reaction as shown below (as well as in Fig.2.11)

Amino Acid + ATP → Aminoacyl.AMP + PPi
(Pyrophosphate)

Aminoacyl.AMP+transfer RNA → Aminoacyl.transfer RNA +AMP

The structure of a transfer RNA was however, solved by Robert Holley working in Rockefeller Institute (later designated as University)

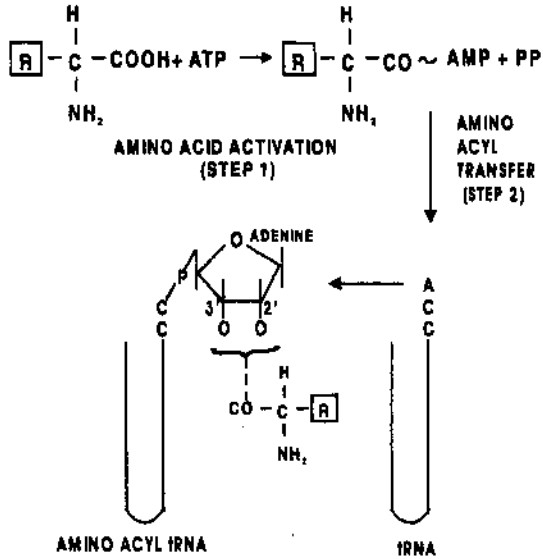



Fig. 2.11 Activation of an amino acid with ATP and transfer of activated amino acid (aminoacyl AMP) to the transfer RNA. Following activation (step 1) the amino acid is attached to the 2´ or 3´ OH group of adenosine at the CCA terminus of the tRNA (step 2). Both the reactions are carried out by the amino acid activating enzyme.

who shared the Nobel Prize for this feat along with Marshall Nirenberg and Har Govind Khorana (the two were responsible for breaking and elucidating the genetic code). Transfer RNA or tRNA molecules are comparatively small in size. These are also known as 4S RNA as they sediment at a rate of 4S. One Svedberg unit corresponds to $10^{-13}$sec. Those tRNAs are composed of 70-80 nucleotides having invariably the three bases CCA at the 3′ end. As already discussed, the primary structure of RNA is like that of DNA with similar 5′ and 3′ ends. Ribonucleotides are joined by phosphodiester bridges as in DNA. It is a single chain. The three constant bases (C, C, A) are at the 3′-end. The amino acids are attached to the ribose of adenosine at the 3′ end through their COOH group as shown in (Fig. 2.11). Both the activation (step1) and transfer to tRNA at the CCA end (step2) are catalysed by the same enzyme. The attachment is at the 3′ OH group of ribose. In some cases amino acids are added to the 2′OH group (instead of 3′ OH group) but eventually transferred to the 3′ OH group with the help of an enzyme (named as isomerase). Half of 20 amino acids are attached to 3′ OH and the rest 10 to the 2′OH. The amino acid activating enzymes are known as tRNA-aminoacyl synthetases. Each amino acid has a specific tRNA and a specific synthetase. This has a very important implication and will be discussed a little later.

Holley was determined to find out the structure of a tRNA but there were so many hurdles. Whatever isolation method was available at that time would end up in a mixture. So he developed a specific column chromatographic technique, using benzoylated DEAE (diethylaminoethyl)-cellulose which is a special type of cellulose with ion exchanging groups attached to it and isolated a single tRNA from yeast which was the carrier for alanine. So he went ahead to determine its structure by the painstaking chemical methods. It took several years to reveal the primary structure (the sequence in which the bases are arranged). One must not forget that not a single DNA or RNA primary structure was known during that time. Therefore, it was not surprising

that he was awarded the Nobel Prize for such achievement. I
(except a few viral RNAs) are mostly single-stranded. But the
double-stranded regions in RNAs where a part of RNA folds on
due to complementarity in the same RNA molecule. Thus part
RNA chain in between has to loop out. After completing the pr
sequence, Holley tried to build up the secondary structure by all
maximum hydrogen bond formation and ended up with the clo
model (Plate 5, Fig. 2.12). It became immediately clear that the
double-stranded regions (called stems) as well as single-str
regions (called loops) in RNAs. The double-stranded regions
helical structure as in DNA, which resembles A-type of str
Stem-loop structure is found to be very common in all types of
as already discussed in Chapter 1. The cloverleaf model (Plate
2.12) proposed for tRNA was on the basis of hydrogen bondin
case of DNA) between A and U (instead of T in case of DNA)
and C. The structure seems to have four arms in this way. Th
referred to as amino acid acceptor arm (mentioned alr
dihydrouridine loop as it contains dihydrouridine (DHU) ins
uracil in this loop, TψC loop so called as it contains thymi
actually ribothymidine which is usually not present in
pseudouridine (ψ), which is an unusual uridine containing
carbon bond between the base and pentose and cytosine (C)
sequence and the fourth loop which is the most crucial one
anticodon loop which recognises the message from DNA in t
of messenger RNA. Loops are marked in the figure accor
structure and function. Amino acyl moiety is attached to the a
stem. TψC loop is so called for having pseudouridine (ψ) and
is so called as it has dihydrouridine. The function of the an
loop is to put amino acid in proper sequence (to be discusse
The size of the variable loop varies among the different tRN

However, it took quite sometime to learn about the
dimentional structure of tRNAs. It was only possible when

could be crystallised and their X-ray diffraction patterns analysed. Alexander Rich and Aaron Klug independently achieved this job and showed that tRNA is L-shaped (Plate 6, Fig. 2.13), if L is written upside down, the lower part will contain the anticodon and the upper part which has CCA terminal (3′ end) will be the acceptor for amino acid. Basically the cloverleaf structure is folded into an L-shaped structure. So the critical first step of protein synthesis involving activations of amino acids and loading on tRNAs became further clear. We have to wait to understand the role of messenger RNA till we come to the next chapter. In some tRNAs, sometimes another additional loop which is variable in size is found. It has already been mentioned that Alexander Rich and Aaron Klug were the first ones to establish the three-dimensional structure of tRNA by X-ray crystallographic technique. Then it became evident that RNA does not exist in cloverleaf form in three dimensions but takes the form of an inverted L. However, the cloverleaf structure is basically correct. The two other arms are folded in such a way that the DHU and TψC arms do not spread out but become folded to give rise to the stem of the L-shaped structure.

Each amino acid has its specific tRNA. It is therefore expected that 20 amino acids should have 20 transfer RNAs to carry them to the site of protein synthesis. Actually there are 32 tRNAs (more than that in some systems) as some amino acids have more than one carrier tRNA. As already mentioned, these RNAs are comparatively small in size. In bacteria and higher organisms the size varies (73 to 93 bases) having molecular weights within the range 24 to 31 kilo daltons assuming that each nucleotide on the average has a molecular weight of 300 daltons. After Holley's miraculous achievement the primary structures of most of the tRNAs have been determined and have been found to have many interesting features. Quite a few tRNAs have unusual bases. Some of the bases are methylated. Some have thymine other than uracil. The newly synthesised RNAs have only A, U, G

and C. Later the modifications of the bases take place. Most of them have a guanylate residue at the 5′ end and the trinucleotide C-C-A at the 3′ end. This is known as the acceptor arm for amino acid, which is attached to its specific tRNA through adenosine at the 3′ CCA end. The amino acid is attached through its carboxyl group to the 3′-OH group of A (or adenosine) in the end. In some cases they may be attached to 2′OH group, as mentioned already.

Specific amino acids are attached to specific tRNAs which is carried out by specific enzymes, which we have learnt already are designated as aminoacyl tRNA synthetases. The enzyme must recognise the specific amino acid on the one hand and the specific tRNA on the other hand. As we shall see later this specific recognition is extremely important for recognition of the message (synthesised by the DNA). This is why this recognition is referred to as the second genetic code. If we go slightly deeper into the problem it will become obvious that these enzymes play a vital role in the synthesis of protein putting all the amino acids in the correct sequence. This will be understood better when we come to the translation of the message produced by DNA at the ribosomal site.

When there are more than one tRNA for an amino acid usually the same synthetase loads all of them. There is, however an unusual situation in *E.coli*. There are two synthetases whereas only one tRNA for lysine, and both load to one transfer RNA. As a whole there are two classes of synthetases depending on their structure and mode of action. It is believed that before the modern protein synthesising mechanism evolved there was only one system for 10 amino acids to be dealt with. Another system evolved to tackle another 10 amino acids. Later the two systems converged to one and both the classes of tRNAs were used to handle all the 20 protein-synthesising amino acids known today. This concept, however, is not firmly established.

It may be recalled that Zamecnik required two fractions of rat liver extract for the incorporation of radioactive amino acids into proteins.

The pH 5 supernatant contained tRNAs and aminoacyl tRNA synthetases. The pH 5 precipitate contained mainly fragmented endoplasmic reticular structure. When the cells are disrupted to prepare the extract the reticular structure becomes fragmented and precipitated at pH 5. So it was thought that endoplasmic reticulum is the site of protein synthesis. However, it was realised later that some spherical particles attached to the endoplasmic reticular structure and composed of RNA and protein are actually the site of protein synthesis. These particles were designated as ribonucleoprotein particles of microsomes during a meeting in 1959 at the Carnegie Institute of Washington. R.B. Roberts who organised the meeting was the one who coined the term 'ribosome' by combining the first four letters and the last 4 letters (ribonucleoprotein particles of microsomes) and that became widely accepted. It was mainly through the work of Alfred Tissieres in the laboratory of Jim Watson that preliminary information about the overall structure of *E.coli* ribosomes became available. At that time several stalwarts started working on ribosomes, notable among them was C. G. Kurland of Sweden and David Schlessinger of U.S.A who joined hands with Watson and Tissieres. Sometime later Mayasu Nomura also joined the group. He is one of the topmost authorities on ribosome, specially in the genetic aspects.

Let us first try to understand what type of structure *E.coli* ribosome has. Naturally it is a giant molecule, its molecular weight is 2.7 million daltons, the complexity will be realised with the simple fact that ordinary table sugar (sucrose) has a molecular weight of 342 daltons only. One molecule of simple protein like haemoglobin has a molecular weight of 64,500 daltons. In all the cases we assume that 1 atom of hydrogen has atomic weight of 1 dalton (a unit of mass very nearly equal to that of a hydrogen atom). The *E. coli* ribosomes like all ribosomes dissociate into two subunits designated as 50S and 30S ribosomes (Fig. 2.14) becuase their sedimentation constants are 50S and 30S respectively. The intact *E. coli* ribosome sediments at 70S. In this connection we should remember that the sedimentation rate

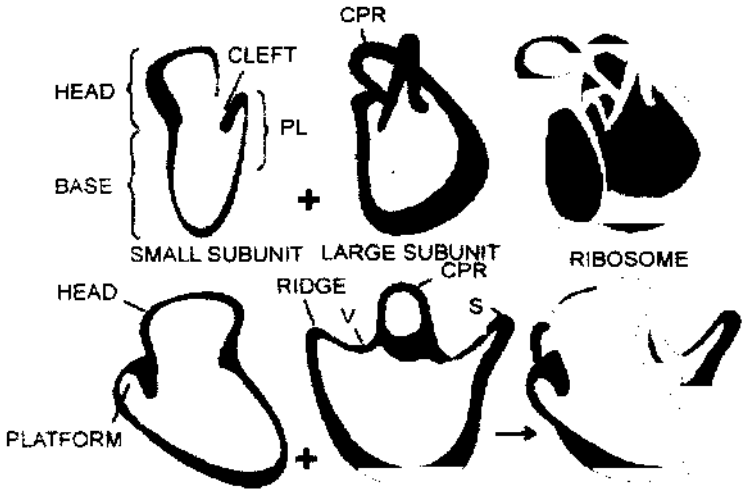



Fig.2.14 *Escherichia coli* ribosomes (70S) and their subunits (30S ar ribosomes). The side and the front views are shown in the figure. These proposed by James Lake (described later) are universally accepted.

will definitely be dependent on the weight of the molecule sedimented but its shape and size are equally important. For exa a globular macromolecule may sediment faster than a heavier mo in case the latter has an elongated shape, which increases the fric force and hinders its sedimentation. The larger ribosome (50S molecular weight, 1.85 x$10^6$ daltons while the smaller one has $10^6$ daltons. So one has approximately half the size of the other tl the sedimentation rate is more than half. Ribosomes from h organisms have equal amounts of RNA and protein but ribo from most microorganisms like *E. coli* have 60% RNA and protein. The ribosomes of higher organisms (eukaryotes) are som larger in size and have sedimentation constant close to 80S ar subunits have sedimentation constants 60 and 40 respect (although those vary from organism to organism). Although R the major component of *E. coli* ribosomes, only three RNA mol

occur in 70S ribosomes (Fig.2.15). The smaller subunit contains a single RNA molecule (or chain to describe it better) which has sedimentation constant 16S and has a molecular weight of 0.55 million daltons. 16S RNA of 30S ribosome contains about 1600 bases. The larger subunit contains two RNA chains, the larger having 3200 bases (molecular weight of the RNA being about 1.1 million daltons) and the smaller RNA molecule (5S RNA) is slightly larger than tRNA molecules. The tRNA molecules having sedimentation constant of

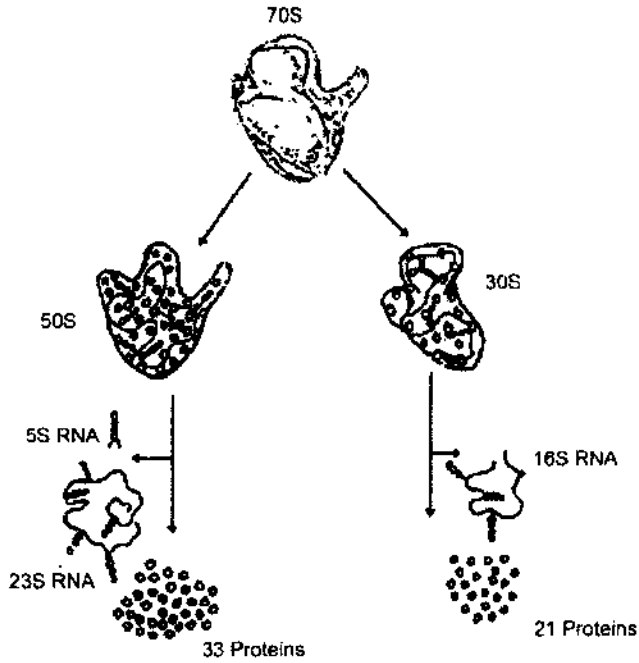



Fig.2.15. Constituents of 30s and 50S subunits of *Escherichia coli*. The smaller subunit is composed of only one RNA (16S) and 21 different proteins, where as the larger subunit is composed of two RNAs (23S and 5S) and 34 different proteins. Except one (L7 / L12 which occurs in four copies) all other proteins are different from each other and occur as one copy each. However L7 is N-acetylated L12.

4S and having an average molecular weight of 26,000 (75-80 bases) are not a part of ribosomal structure. However, these associate with the ribosomes temporarily to bring the amino acids to the correct site so that they could be linked to each other to give rise to protein molecules. The 5S RNA, which is part and parcel of the 50S ribosomes, has about 120 bases and a molecular weight of about 11,000 daltons.

The schematic representations (secondary structures) of 5S, 16S and 23S RNAs are shown in Fig 2.16 a, b and c respectively. The distinctive feature of the three RNAs is the stem-loop structure, which is common to all RNA structures. As the size of 5S RNA is slightly larger than 4S (transfer) RNAs its sequencing was a simpler job. In case of 16S and 23S RNAs the complementary DNAs were synthesized with the help of an enzyme, reverse transcriptase and sequencing was

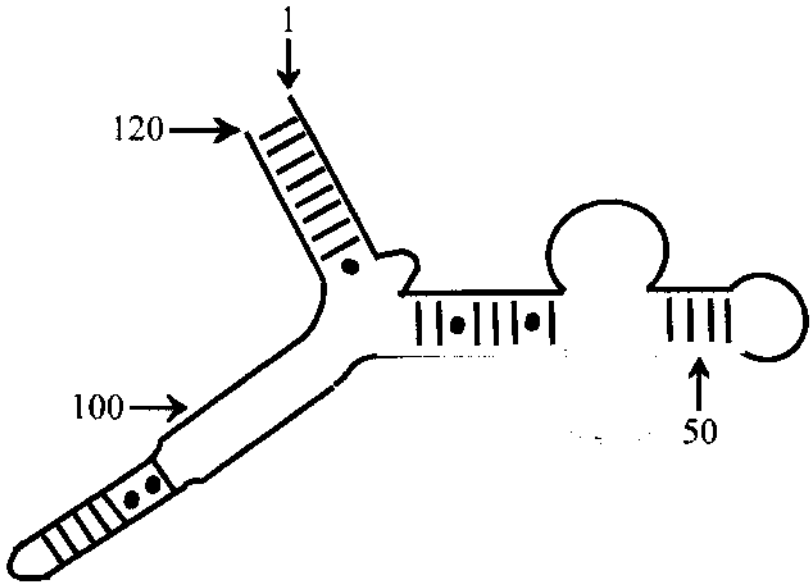
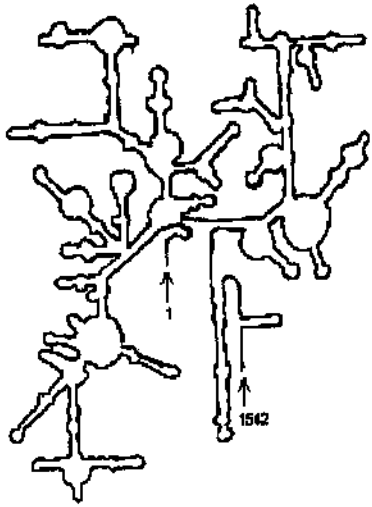



Fig. 2.16 (a) The schematic diagrams of primary structures of a) 5S RNA b) 16S RNA & c) 23S RNA (b & c on the next page).

(b)




(c)

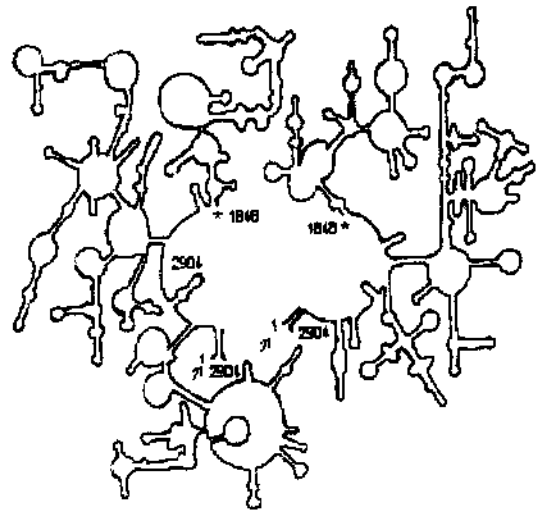



Fig 2.16. continued. (a) on the earlier page.

done at the DNA level by the standard technique. Since RNA is rather labile to alkaline conditions and RNA degrading enzymes occur widely, this approach was adopted by Harry Noller, one of the top ribosomologists of today.

The situation regarding the protein components of the ribosome is rather complicated. The credit goes to Waller who first showed in 1964 that ribosome is made up of a variety of proteins. Actually there are 54 proteins present in *E. coli* ribosome; of these 34 are present in 50S ribosome and 21 in the smaller or 30S ribosome (Fig 2.15). One protein is common to both so the total number is 54 instead of 55. However, one particular protein L8 does not actually exist as it is a combination of 3 proteins (L10, L7 and L12). These proteins are numbered S1-S21 (small subunit) and L1-L34 (large subunit) according to their movement on the polyacrylamide gel during electrophoresis. All these proteins have been isolated, well characterised physically and chemically. The primary structure of most of them is known. This gigantic work was mainly done by late Wittmann and his wife Wittmann-Liebold and their excellent group working tirelessly at the Max Planck Institute of Molecular Genetics at Berlin. There will be several occasions to mention about their work. Wittmann worked earlier on tobacco mosaic virus (TMV) RNA and attempted to break the genetic code from the structure of coat proteins. He presented this work in the International symposium held at Hyderabad which has been mentioned in Prologue. His wife, who was primarily responsible for characterising ribosomal proteins, visited India much later on our invitation and delivered a lecture at the Banaras Hindu University.

An interesting observation made by Nomura[3] while working in Watson's laboratory is worth mentioning here. During ultra-centrifugation of ribosomes on cesium chloride gradient some proteins get detached from the ribosomes and float on top of the gradient while the ribosomes move downwards due to the gravitational field and

take the equilibrium position according to their density. This technique is known as density gradient centrifugation. Ribosomes sediment at a place of the gradient where density is equal to that of ribosomes. The detached proteins, which float on top of the gradient, were called 'split' proteins and the particles devoid of these proteins were referred to as 'core' particles. The detachment was thought to be due to hydrostatic pressure developed during centrifugation but was later shown to be due to salt (CsCl) present at a high concentration. However, the exciting achievement of Nomura was that he could reassemble the intact ribosomes again by incubating the core particles with the split proteins. The preliminary success eventually led him to fully reconstitute the 30S ribosomes from the 16S RNA molecule and 21 different proteins. This process is practically spontaneous with the requirement of a small amount of energy at one step. The incubation at 37°C for a few minutes is enough to provide the energy for assembly. Through his elegant work he also demonstrated the sequential addition of 21 proteins to 16S RNA and the specificity involved in the sequence of attachment of the proteins during the association process. Initially, one or two proteins are attached; subsequently others are added step by step. Unless the core proteins are assembled, other proteins cannot be attached. Nomura thus created an assembly map (Fig. 2.17), which has helped considerably in understanding the structure of the ribosome, and to learn to some extent, the formation of this complex organelle at least under *in vitro* condition. In all likelihood the assembly takes place in the same way under physiological condition also. The assembly of 50S ribosomes had been a somewhat tougher job but was also initially done by Nomura. However, this was easily achieved by Nierhaus in Wittmann's laboratory by a two step procedure. After this crucial work on 'Ribosomal assembly map' Nomura's attention was diverted more and more to the genetic aspects of ribosomal proteins. He mapped the positions of some of the ribosomal protein genes on the *E. coli* genome. It will be quite interesting to know how the synthesis of so many ribosomal proteins is coordinated genetically

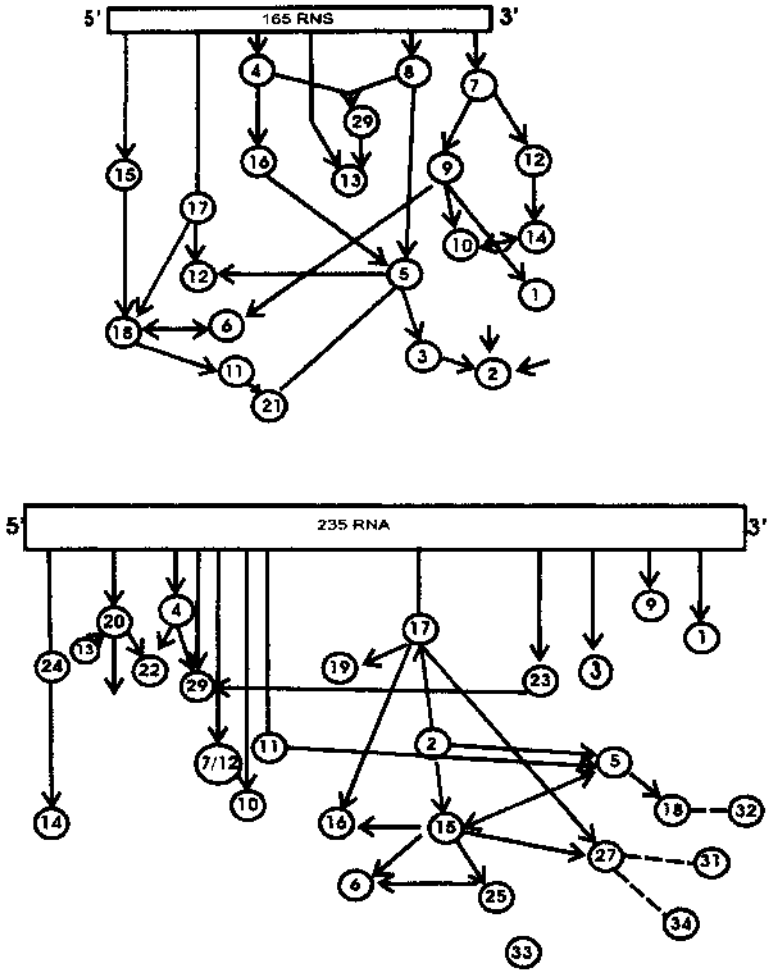



Fig.2.17 Assembly map of ribosomal proteins. The map was built by the painstaking work of Masayasu. Nomura on the complete assembly of the two subunits (a) 30S & (b) 50S from their RNA and protein components. The details have been described in the text.

so that the 70S ribosome is assembled properly. Normally, there is no excess of free ribosomal proteins in the cytoplasm of the growing cell.

Some digression at the stage may not be undesirable. Wittmann's work on tobacco mosaic virus at the Max Planck Institute of Tübingen established his reputation as a biochemist and a molecular biologist. From Tübingen he moved to Berlin. It was a great pleasure for me to meet him for the first time at Tübingen in 1957 when he came to receive me at the railway station. On my way back to India, after spending three years in U.S.A., I traveled through Europe. He was then working at the Institute of virus research headed by Gerhardt Schramm. During my trip around the world in 60 days in 1971, I visited for the first time the Max Plank Institute of Molecular Genetics at Berlin, which had three wings, each headed by a Director. Wittmann was one of them. His tremendous experience in the field of tobacco mosaic virus perhaps motivated him towards 'ribosomology', which needed expertise in protein chemistry to start with. It is unfortunate that ribosome community suffered a lot due to his untimely demise. It was not surprising that Wittmann's first and foremost attempt was to identify all the proteins present in the *E. coli* 70S ribosomes. Reports on the identification of the proteins were pouring in from various laboratories. The method of separation of the proteins was mostly by two-dimensional gel electrophoresis. This reminds me of my early days with two-dimensional paper chromatography and paper electrophoresis. The difference between the two techniques is that polymerised acrylamide instead of paper is used as the bed for the separation and electrophoresis (and not chromatography) is carried out in both the directions. Polyacrylamide has one great advantage, it acts as a molecular sieve and can separate macromolecules according to their size. The double effect of sieving (gel filtration) and electrophoresis resolves molecules of very close properties and sizes. All the 54 ribosomal proteins could be easily separated out. Most of

these proteins were found to be highly basic in nature, only a limited few are acidic or neutral. In the beginning there was a lot of confusion about the nomenclature of these proteins, naturally different numbers were used by different groups to identify them. Finally Kurland's, Nomura's and Wittmann's groups came to an agreement in identifying them by numbers depending on their positions on the two-dimensional gel. As already mentioned, the proteins from the smaller subunit were designated as S1 to S21 and the larger subunit L1 to L34. The molecular weights of the proteins range from 7,000 to 32,000 daltons with an average molecular weight of about 17,000 daltons. The exception was S1 protein of 30S subunit, which had a much higher molecular weight (65,000). For quite sometime this protein was thought not to be a part of the ribosome as it could easily get detached from the subunit but there is no ambiguity about it being a ribosomal protein. Each of these proteins except one occurs in one copy per ribosome. The exception is L7/L12. Actually, L7 and L12 are identical proteins, the only difference being that L7 has an extra acetyl ($CH_3CO$) group attached at its N-terminal end. This protein (L7, L12 being treated as the same protein) occurs in 4 copies per ribosome. It indicates that this protein may have some special role to play, which is not so far known. So, there is lot of complexity in the structure of the ribosome, which is not unexpected as this acts as the factory for the synthesis of all the proteins. This factory may be of varying size depending on organism but the overall architecture and mode of production of the finished material, in this case the proteins, are more or less the same. So, like DNA with the genetic code ribosome is ubiquitous and a 'must' for all existing living systems.

Ribosome is a tiny organelle approximately 200-300Å in diameter and is more or less spherical in shape with special features. Although ribosomes could be easily visualised under the electron microscope, it is still very difficult to recognise its structural features. They were recognized on the basis of their shapes and sizes in different

orientations on the grid of the electron microscope. Attempts were made by various groups of workers including Wittmann's to build models on the basis of them. The Stoffler's group in Wittmann's laboratory modelled 30S subunit as a quasi -symmetrical object having a small head and a large body including a hollow part and having elongated extensions on both the sides. 50S subunit was described by them as an apparently symmetrical object with three conspicuous protuberances, which were almost symmetrically arranged around a depression at the top of the body. One has to stretch their imagination to assume this model to look like an 'arm chair'. Similarly in popular jargon 30S subunit can be looked upon as a 'teddy bear', the teddy bear sits on the arm chair when the 70S ribosome is formed. Another group in California headed by James Lake[4] performed a similar work on model building. They have slightly different versions of the structure of the two subunits. For example, they visualised 30S subunit as a completely symmetrical object with a small head, large body without hollow part and a distinct high shoulder or platform. Unlike Stoffler's group Lake's group suggested that the crests of 50S ribosomes are neither pronounced nor symmetrical. Further, there is a thumb-like protrusion which extends at an angle of 50°C from one side of the central protuberance. Since L7/L12 proteins are found to be present in such structure it is called L7/L12 stalk. A third model built by the Russian group headed by Alexander Spirin and his coworkers at the Institute of Protein Reasearch at Pushchino, more or less agreed with both the models but had somewhat different features. Lake's model is, however, widely accepted and has been depicted in Fig. 2.14. Naturally, a lot more work has to be carried out to refine the models. At present it is being studied by different techniques including cryoelectron microscopy, where samples frozen in liquid nitrogen are subjected to electron microscopic examination. In this method the resolution is less, but the features are not distorted as in ordinary electron microscopy. Joachim Frank's group at Albany, New York, U.S.A. and Martin Van Heel's group at Berlin, Germany are

engaged in this type of exercise and considerable amount of information about the positions of tRNAs within the cavity between 30S and 50S ribosomes have been obtained. It is gratifying to mention that Rajendra Agrawal, my last Ph. D. student has contributed a great deal in this area in Frank's laboratory.

Two sophisticated techniques were extensively employed to understand the structure of ribosomes and have already yielded very useful information. Isolated proteins of ribosomes when injected individually into the rabbit produce very specific antibodies, as expected. Those react only with the specific antigens (in this case ribosomal proteins). The antibody against a specific ribosomal protein is directed against the subunit from which the protein is derived. An antibody molecule is Y-shaped and bivalent i.e., has two sites of recognition with the antigens at the two ends of the branches of Y. Thus, under an electron microscope, pairs of ribosomes linked to Y-shaped gammaglobulin (antibody) can be observed. After examining a large number of these pairs of ribosomes it was possible to visualise the positions of different proteins on the subunit. In this way locations of almost all the proteins on the ribosomal subunits have been determined. This information helped to build models for the ribosome. The most interesting observation made by these studies is that some proteins are localised at more than one site, widely apart on the ribosome, indicating clearly that the proteins concerned are not globular in shape but have elongated structure. Naturally, ribonucleic acid-protein interaction plays the most significant role in producing such a structure, giving rise to the overall organisation of the ribosome. This technique known as 'immunoelectron microscopy' (combination of immunology and electron microscopy) was exploited both by Wittmann's group and Lake's group and others as well.

The other technique, which is somewhat more complicated, is the neutron diffraction analysis. As the X-ray diffraction technique is used to analyse the three-dimensional structure of proteins and nucleic

acids, neutron diffraction is used in a more or less similar way. In case of ribosomes the desired proteins of a subunit are replaced with proteins labeled with deuterium (deuterium is heavy hydrogen having double the mass of normal hydrogen) by Nomura's assembly method. When neutrons are directed against such ribosomes having only two deuterium labeled proteins (prepared with the help of Nomura's assembly method) they act as strong centres of diffraction. Other proteins would act as a background only. From the diffraction pattern the distance between the deuterium labeled proteins could be calculated. In this way the relative position of a protein with respect to another protein (only two of them being deuterated) is thus obtained. By the triangulation technique the three-dimensional positions of three proteins with respect to each other can be worked out. This is a very laborious technique and has been successfully utilised by Peter Moore's group in Yale University and Benno Schoenborn's group at Brookhaven National Laboratory. I. Serdyuk in the Institute of Protein Research at Moscow has also used the technique quite fruitfully. It has already been mentioned that cryoelectron microscopy has also been used for similar studies although the resolution achieved by this technique is somewhat less than ordinary electron microscopy. The ultimate success has been achieved comparatively recently by crystallizing the ribosome and determining their structures by X-ray diffraction which would be discussed in Chapter 4.

I spent 15 years of my active life in 'ribosomology' as it is called these days. How did I get involved in ribosomology? That was an example of serendipity. I love 'serendipity' because the word is of Indian origin ('Sarandip' means 'golden island') which was accidentally discovered. For some compelling reasons we were purifying the enzymes in *E. coli* extract which degrade RNA. One of the enzymes, RNaseI, behaved rather awkwardly. To start with there was very little enzyme present in the extract. On further purification the activity increased. It was quite unnatural. Normally, the activity

of an enzyme is lost during purification due to poor recovery or loss of an enzyme activity during the handling procedure. Even if there is no loss, the activity is not expected to increase. The increase in enzyme activity could be accounted for by assuming that there was some inhibitor of the enzyme present in the extract, which was removed during purification. So we tried to purify the suspected inhibitor. It was falling apart during purification and the only information we could get about the inhibitor was that it was complex of ribonucleic acid and protein. There was no other alternative but to go to the library and find out that the RNA degrading enzyme (RNaseI) which we were purifying gets associated with ribosomes (specially the smaller subunit) and becomes latent. David Elson and his wife at the Weissmann Institute of Science in Israel made this observation for the first time. We got the answer. Magnesium ion (Mg++) has to be around to keep the ribosomes intact for carrying out protein synthesis otherwise it will be chewed up by the enzyme. It happened so. When we had the answer we thought of finding out why the enzyme becomes latent in contact with ribosome. That led a few interesting findings and kept us busy for 15 years or so. It should be mentioned here that in order to avoid this now a days a mutant of *E.coli* lacking RNaseI is invariably used to prepare intact ribosomes. It is interesting to note that this enzyme in the cell is located in the periplasmic space (space between the two membranes of the bacteria) and therefore cannot do any harm to the ribosomes in the cytoplasm. This was discovered by Leon Heppel of the National Institutes of Health with whom I was in close contact.

Both David Elson and his wife Nina were highly active in ribosomology. The meeting on 'ribosomes', which was held in Israel, was organised by them. That was a unique opportunity for me to meet them although I had visited Israel earlier. The two were not only excellent scientists but also very good hosts. Unfortunately, David Elson died a few years ago. The International Conference on Ribosomes in 1989 at Montana, USA was held in memory of both H. G. Wittmann and David Elson[5].

## References

1. *Introduction to protein structure.*, C. Branden and J. Tooze, Garland Publishing Inc., New York (1991).
2. *Biophysical Chemistry.* Part III. The behavior of biological macromolecules. C. R. Cantor and P. R. Schimmel, W. H. Freeman and Co., SanFrancisco (1980).
3. *History of ribosome research: a personal account.* M. Nomura, in reference 5 (below)
4. *The ribosome.* J. Lake. Scientific American, Volume 245 (August), pp 84-97 (1981).
5. *The ribosome structure, function and evolution edited* by W. E, Hill, K.A.Garrett, P. B. Moore, D. Schlessinger and J, R. Wernar. American Society for Microbiology, Washington, D.C. (1990).

*****

# Chapter - 3
# Notation of the Music
## (Genetic Code and Messenger RNA)

*It will soon be clear that just as man-made music is written in musical notation, the music of life is written in coded form in DNA. Before playing the music the coded message has to be transcribed in a tape made of RNA, which is rightly called the messenger RNA. Although there was a hunch about the existence of such a messenger for quite some time it took two to three decades to prove it. The earlier chapter duly created the backdrop and this chapter will provide the finishing touches. The transcription process following replication is the forerunner of the process of translation, the story of which has been partly told in the earlier chapter and will be completed in the next. The present chapter smells Indian to some extent, which could not be avoided by any means.*

The direct relationship between genes and enzymes was postulated as early as 1902 by Archibald Garrod, a clinician who had a famous geneticist friend William Bateson. It was an uncommon case when a mother would be startled to find her child passing out urine, which would soon turn black after voiding. There seemed something wrong with the child; some fire inside the body. She would rush to the clinician who apparently had no answer. Whenever Garrod came across such a case he would start enquiring among his relatives as to whether any other child or an adult in the family had the same problem. He would invariably find one or two. His inquisitive mind worked out that there must have been some hereditary (genetic) changes in

metabolic processes of the child. The disease was well-known and called alkaptonuria. Later several similar hereditary diseases like sickle cell anemia, glycosuria etc., were identified. All of these supported Garrod's hypothesis that the defect was inherited by the child from the parents. The words 'inborn errors of metabolism' crept into his mind and the terminology was apparently coined by him. The only information he did not have was where the error was. Fortunately this particular defect, which is due to defective metabolism, is not destructive. The disease was due to the defect of metabolism of an amino acid tyrosine as shown in Fig. 3.1. Homogentisic acid accumulates as a result and is excreted in the urine. When in contact with air it gets oxidized and turns black in colour. No wonder, the mother gets startled to see her child's urine turning black. In normal individuals tyrosine is degraded by a pathway in which homogentisic acid is an intermediate. In alkaptonuric individuals further metabolism of homogentisic acid is blocked, so it accumulates and is excreted through the urine. After voiding, homogentisic acid is oxidized when it comes in contact with air and turns black. Hence the disease is called Alkaptonuria (black urine). Fortunately this defect does not cause any harm to the physiological status of the individual who in turn lives happily throughout his life.

'When I was working at the National Institutes of Health with Bernie Horecker, an expert in pentose metabolism, I required a lot of xylose, a sugar, which was terribly expensive. In the same building was located the beautiful research hospital which admitted cases of people suffering from rare diseases. Here walks in a patient who excretes a lot of xylose in his urine, the disease being known as xylosuria. Xylose could be easily isolated from his urine. The individual was beyond doubt a factory for the production of xylose, which is one of the pentoses (five carbon sugars). All the patients who excrete any of the pentoses are generally referred to as cases of pentosuria.

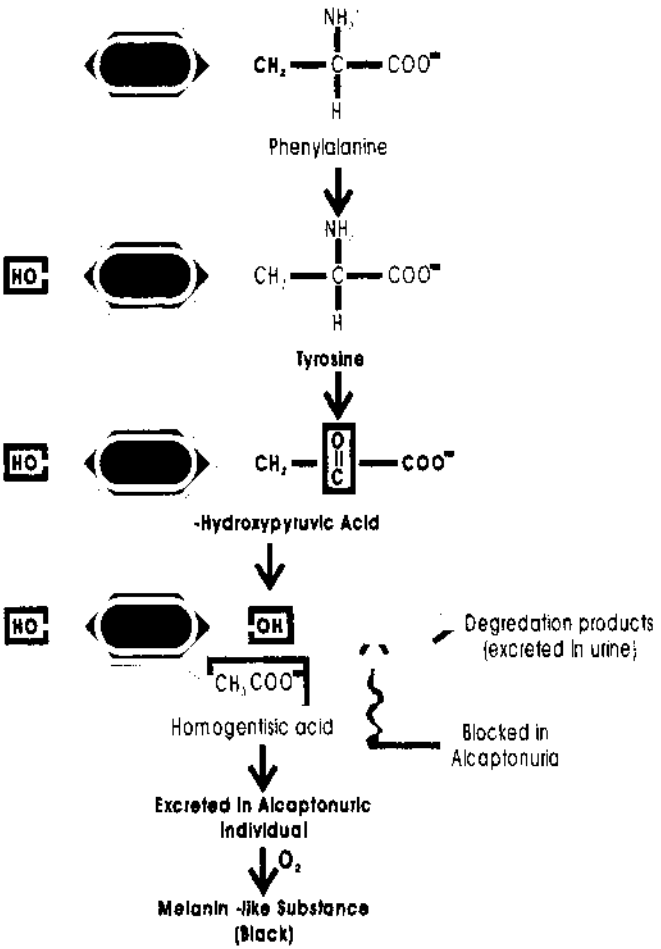



Fig.3.1. Hereditary (genetic) defect in the metabolism of tyrosine leading to alkaptonuria. The pathway of metabolism of tyrosine and the block in the alkaptonuric individuals is shown. Details have been discussed in the text.

Preliminary studies indicated that the hereditary diseases are due to defects in the genetic material or so called 'genes'. But the basic question that arose in the minds of the biologists was the kind of relationship which existed between genes and the enzymes. It took the efforts of a large number of scientists for many years to establish the relationship between genes and enzymes. This was possible when the mutant bacteria requiring certain metabolites for growth (auxotrophs) became available. Metabolic block at any step leads to the termination of the pathway at that particular stage resulting in the accumulation of the intermediate. One of the best examples is the biosynthesis of the amino acid arginine from glutamic acid by seven consecutive chemical reactions. Auxotrophic mutants of the bacteria were isolated with block at each step and it could be directly shown

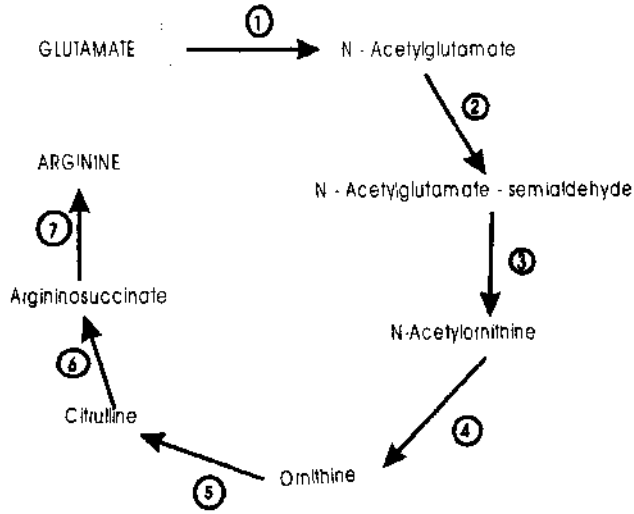



Fig 3.2 Metabolic defects in the biochemical pathway of synthesis of arginine from glutamate in bacteria as model system. The above depicts the metabolic pathway of biosynthesis of arginine, an amino acid formed from glutamic acid, another amino acid. Mutants defective in individual steps have been isolated by simple genetic manipulation.

which specific enzyme was missing in each of them (Fig.3.2). All the mutants were naturally defective in arginine biosynthesis and required arginine for growth. This way it was also demonstrated that one gene produces one specific enzyme. However, when it became known that one enzyme may have more than one polypeptide chains it was also realised that one gene could not synthesise the whole enzyme but one polypeptide chain only. Polypeptides associate to form the enzyme. Thus, the hypothesis of one gene-one enzyme was modified to one gene-one polypeptide hypothesis. With more and more progress in biological research it was realised that genes may code for proteins but did not directly synthesise them. It became known later that genes directly produce a kind of RNA known as messenger RNA and those in turn synthesise the polypeptides as per code contained in DNA. This led to the Central dogma enunciated by Crick DNA→RNA→Protein which was eventually found not to be absolutely correct and had to be somewhat revised.

Jacob and Monod's operon concept was the first of its kind to explain the induction of enzymes in prokaryotes (bacteria including *archaea*). One of the examples is β-galactosidase, an enzyme which catalyses the hydrolysis of a galactoside for example, lactose to its two components glucose and galactose. Lactose is a galactoside having a combination of galactose and glucose. Although the gene for β-galactosidase (β- is the type of linkage between two sugars) is present in *E. coli* the enzyme is not normally produced but upon addition of lactose (a β- galactoside) the enzyme gets induced. So it is an inducible enzyme and different from the constitutive enzymes which are normally produced and do not need any inducer. There are two more enzymes (permease and transacetylase) which are also simultaneously induced. The genes of these two enzymes along with that of β-galactosidase constitute the galactose operon. (Fig.3.3a). Jacob and Monod who shared the Nobel Prize in 1968 for this outstanding work rightly visualized that there is a site on DNA from which the message

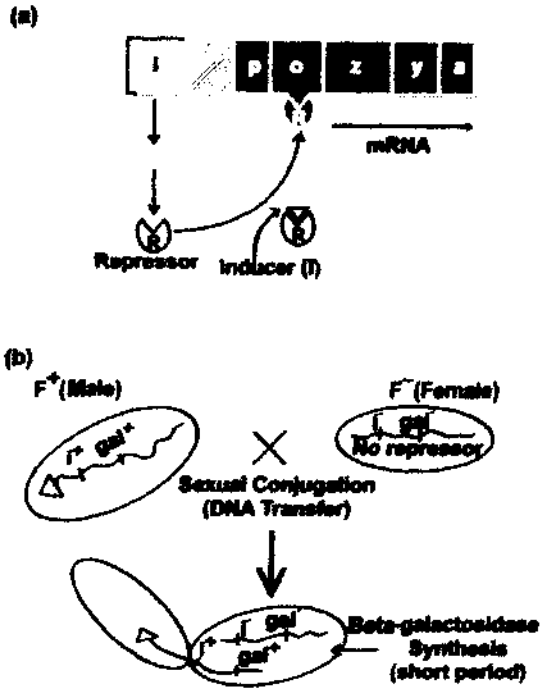



Fig 3.3 Operon concept of Jacob and Monod and the breakthrough Pardee-Jacob-Monod (PaJaMo) experiment to establish it. (a) The lactose (lac) operon as originally conceived by Jacob and Monod consists of the following sites and genes in sequence, repressor (i) gene, promoter (p) region, β-galactosidase gene (z), permease (y) gene, and transacetylase (a) gene. The operator (o) has also been depicted. For details see the text. (b) PaJaMo experiment was based on the transfer of genetic material in bacteria by sexual conjugation. The male bacteria ( $F^+$) has the intact operon, so cannot synthesize β-galactosidase due to the presence of repressor and female ( $F^-$ ) is a mutant which has both repressor and β-galactosidase genes defective, so it cannot synthesize β-galactosidase either in presence or absence of inducer. However, on receiving the genetic material from the male by sexual conjugation it can induce synthesis for a short period in receptor free environment.

for production of the enzyme(s) is regulated and they named the site as the 'operator'(o). The *i* gene produces a repressor which binds to the operator and blocks the production of the message as shown in Fig. 3.3a. However, when an inducer like lactose is added it binds to the repressor and alters its shape. As a result it fails to bind to the operator, so the expression of the message starts. As long as the inducer is present the operator remains open but as soon as the inducer is exhausted (due to its breakdown or any other reason) or withdrawn there is no more production of the enzyme(s) as the repressor blocks the operator. There are two more genes apart from the β-galactosidase gene, permease(*y*) and transacetylase(*a*) which are simultaneously regulated by the same operator. This constitutes the galactose operon. The operator practically acts as a 'switch' of the whole operon which is switched 'on' or 'off' with the help of the inducer or the repressor respectively. The crucial experiment to establish this fact was done by Arthur Pardee in U.S.A. and Jacob and Monod in France independently and eventually by Pardee in Jacob and Monod's laboratory. The famous experiment is jokingly referred to as Pa Ja Mo (Pardee, Jacob and Monod) experiment (Fig.3.3b). This was based on the gene transfer from a male bacterium to a female one (known as sexual conjugation). The male bacterium had a normal β-galactosidase and repressor genes, so it could not produce the enzyme without the addition of the inducer. The female had both β-galactosidase and repressor genes defective and thus failed to synthesize the enzyme. But when the two were mated (DNA is tansferred from the male to the female) the female (recipient of the male gene) immediately started synthesizing the enzyme. The male gene which was in a repressed state in the male became induced without an inducer as there was no repressor in the female environment. It took a while to build the requisite concentration of repressor expressed from the DNA of the donor (male cell). Hence, after some time the β-galactosidase synthesis was blocked. This is illustrated in Fig.3.3b. The induction of an enzyme or enzymes with an inducer is similar to induction of a

latent virus (in lysogenic state of the host) by ultraviolet light or some chemicals to lytic state of the host. Some bacterial viruses (bacteriophages) can stay in latent state in the host. Its genetic material is integrated with the host genome and stays latent (repressed state) due to the synthesis of the repressor molecules. On exposure to U.V. (or some chemicals) the induction takes place and the viral genome becomes detached from the host and starts multiplying. These are two entirely different phenomena but the mode of repression is somewhat similar. This development in molecular biology reminds us of a similar situation in the early history of biochemistry. Chemists who were investigating the basic features of fermentation of sugars to alcohol and physiologists who were studying the mechanism of digestion ended up with the same answer, enzymes being responsible for both the phenomena. The concept of regulation of metabolic processes was rather new in the 1960s and first led to the realisation of the existence of a message for the production of an enzyme. It was realised at that stage that enzymes could not be synthesised in the absence of not only amino acids but also uracil, an ingredient of RNA. Therefore, it was the forerunner of the concept of messenger RNA. This idea was proved to be absolutely right. According to the operon concept of Jacob and Monod, the production of more than one enzyme is regulated by one operator through a messenger which turned out to be RNA. As mentioned already, the genes of the three enzymes (β-galactosidase, permease and acetylase) are linked and present in the same operon regulated by the same operator. The order of location of the genes were also correctly postulated. Jacob and Monod were ahead of time in their visualisation of the regulatory system present in the living cell. However, there are some controversies in the matter, which have been recorded by Judson in his highly interesting book[1]. It should, however, be remembered that the regulation is quite different in the eukaryotic (higher) organisms where most of the enzymes are expressed in an unlinked fashion and the message is interrupted. There is hardly any such 'operon'. As already discussed in Chapter 2

Pauling's concept of hydrogen bond and his suggestion of α-helical structure of protein were responsible for the initiation of the understanding of the overall structure of proteins. The subsequent development was intimately connected to the development of molecular biology. However, he neither coined the term nor became a protagonist of that. Similar is the situation in another entirely different area, which was not only outstanding but a breakthrough in biology, the concept of molecular diseases. The term is not extensively used today, as we understand the cause of such diseases thoroughly. Sickle cell anemia among the Africans was known for a long time. When Pauling isolated hemoglobin from the sickle cell anemic patients and subjected them to electrophoresis to check its mobility in an electric field he must have had the intuition that something was wrong with the molecule. He was absolutely right. The sickle cell haemoglobin (HbS) moved faster than the normal adult haemoglobin (HbA) as shown in (Fig. 3.4). It was later shown that the former is less soluble (under physiological condition) and also has less oxygen binding capacity than the latter, which leads to the suffering and sickle cell crisis and eventually leads to death. Under less oxygen tension red blood cells (RBC) having a beautiful biconcave shape assume the sickle-like appearance in the sickle cell patients and due to loss of their flexibility the sickle cells create the crisis when they pass through the capillaries. Less solubility of HbS leads to rigidity. The most outstanding observation which Pauling made following the isolation of haemoglobins from both the parents of the child suffering from sickle cell anaemia to check their movement in the electric field and was not surprised to find two bands of haemoglobin at the position of the normal and the other at that of sickle cell haemoglobin. He became sure that they must have had the sickle cell trait, as the disease was hereditary. Both the parents had one good gene and one defective gene. The poor child by chance inherited the defective genes from both the parents and produced sickle cell haemoglobin only. This supported Archibald Garrod's concept of inborn errors of metabolism.

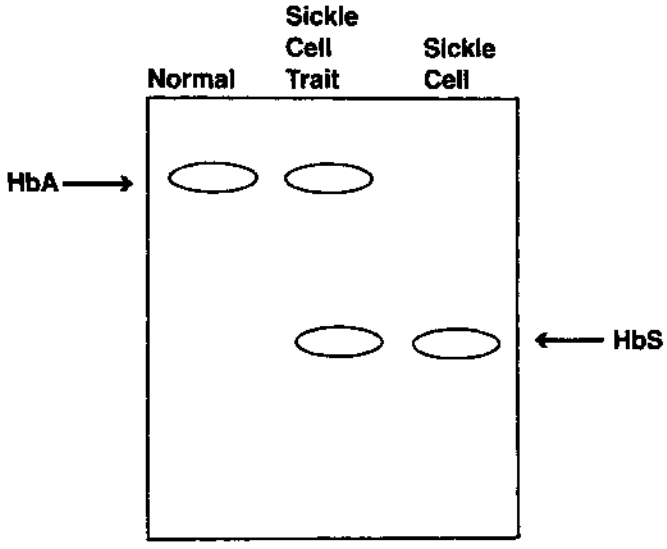



Fig 3.4. Separation of sickle-cell hemoglobin (HbS) from adult hemoglobin (HbA) due to charge difference during starch gel electrophoresis. Due to charge differences adult hemoglobin (HbA) and sickle-cell hemoglobin move at different speeds in an electric field. HbS moves faster than HbA.

Later the defect was clearly understood at the molecular level as V. Ingram developed a simple fingerprint type of analysis of the closely related proteins, specially haemoglobins. Proteins were partially digested by a proteolytic enzyme (trypsin in this case) and the digest was analysed by two-dimensional electrophoresis on filter paper. The spots (peptides) from normal and altered haemoglobins, which did not match, were analysed and the amino acids in the peptides were identified. It clearly showed that there was difference in only one peptide, which had one of its amino acids, glutamine, replaced with valine. This may be looked upon as equivalent to the differences in finger prints of individuals, It was further shown that the amino acid was in position 6 of the protein chain, which had 146 amino acids. Out of the two chains ($\alpha$&$\beta$) of haemoglobin, the $\beta$-chain was

defective in only one amino acid which created this havoc. This simple change resulted in less solubility, less oxygen binding capacity and polymerization leading to rigidity of red blood cells. The answer became clearer at the molecular level when the genetic code became known. The problem was understood better at the DNA level. There was a single base change (A to T), the code (GAA/GAG) for glutamic acid was changed to code (GTA/GTG) for valine, therefore base T was inserted in place of A. This will be absolutely clear from the genetic code to be discussed later.

As early as in 1942 it became clear from the ultraviolet spectrophotometry of the cell (*in situ*) developed by T. Caspersson and the cytochemical work of J. Brachet that RNA had something to do with protein synthesis. It was also demonstrated that RNA is synthesised in the nucleus of the eukaryotic cells and transported into the cytoplasm. At that time the classical paper of Jacob and Monod in the *Journal of Molecular Biology* (1961) had not yet been published but everyone was aware of their concept of messenger RNA and also the contents of the paper. The flow of information must be from DNA to RNA to protein. In this connection E. Volkin and L. Astrachan's highly important observation indicating the base sequence of RNA produced in phage-infected cell, mimicked the viral DNA, was now being widely accepted. These two authors had clear indication that viral messenger RNA produced after infection took control of the host cell machinery following infection. Unfortunately, they did not put up the case strongly without any ambiguity. Later I could share their dismay during my visit to their laboratory. So the ground was prepared to discover the enzyme responsible for the synthesis of RNA (so-called messenger RNA) on DNA template, which was just the experimental proof of the concept that was developing rapidly. Before telling the story of how I myself got involved in the race for the discovery of the enzyme, DNA-dependent RNA polymerase, let me tell you how the existence of messenger RNA in the cell was directly demonstrated in two different laboratories almost simultaneously. It

should, however, be remembered that the enzyme was discovered earlier.

The discovery of the enzyme synthesising RNA on DNA template was quite different from the discovery of messenger RNA. The term messenger was coined by Monod which was documented in the famous 1961 paper of Jacob and Monod in *The Journal of Molecular Biology*. The term was, however, being tossed around for quite some time. On the basis of Volkin and Astrachan's observation that infection of *Eschericia coli* with bacteriophage T4, the phage DNA-specific RNA (similar in base composition) is synthesized, Sydney Brenner and Francois Jacob planned an experiment based on Meselson and Stahl style for DNA synthesis (see Chapter 1) with the idea to trace the RNA (with heavy isotopes) if it got associated with the ribosomes (which were already recognised as the sites for protein synthesis). They did the successful gradient experiment during their visit to Caltech, California. Simultaneously, a group at Harvard headed by Watson tried to perform a similar experiment in normal *E.coli*, which was not infected with phage. Francois Gros, Walter Gilbert and others showed the formation of an unstable RNA (like Volkin-Astrachan RNA), which later associated with the ribosomes. This broke the myth of ribosomal RNA acting as messenger, which was floating around for quite sometime. Both the work from Caltech and Harvard were published back to back in Nature in the year 1961. Thus a third type of RNA (other than transfer and ribosomal RNAs) was recognized and later established as the *handiwork* of DNA-dependent RNA polymerase. All types of RNA in bacteria are synthesized by this enzyme.

To get back to the discovery of DNA-dependent RNA polymerase I should start with Arthur Kornberg who could be called the 'Wizard of DNA' and is the living example of an enzyme chemist who kept his eyes on the synthesis of the gene as Watson on its structure. Kornberg's path was unusually long and torturous and he did not start with a big bang like Watson and Crick but in a more painstaking way starting from the very building blocks of the genetic material. The

life-long association between Kornberg and DNA is one of the examples of the spirit behind science. After returning to the Bose Institute in 1957 from the first trip abroad I got a letter from Kornberg informing that he wanted to visit Calcutta after attending a meeting in Japan. Since I worked with Bernie Horecker who was an intimate friend of Arthur Kornberg the latter knew me well. I was excited and requested the director D. M. Bose to take care of his stay. He immediately agreed to arrange for his accommodation in the old residence of Sir J. C. Bose, the Founder-Director. That was not a wise decision as it was not quite suitable for modern-day living but I could not protest. Kornberg arrived with his wife. The Director was kind enough to lend his own car to show them around. That was almost a vintage car and the Kornbergs were amused to have a ride in that. I was showing them the usual places of interest to the tourists but Kornberg was not happy. He wanted to see the Kalighat temple. The source of knowledge of the existence of the temple must have been the tourist's guidebook. I took them reluctantly to the temple, filthy and infested with beggars of all kinds. That was one of the most horrible scenes to show to a Westerner. Mrs. Kornberg was almost fainting at the sight, so I immediately diverted them to the nearby Ballyganj lake area, one of the best residential places in Calcutta those days. The lake and the adjoining area(s) were beautiful. She almost cried with joy at the sight. 'Why do not the people come here instead of going to the temple?' After breathing fresh air she felt good and this was the right opportunity to bring them back to the Institute.

The Kornbergs were followed by Herman Kalckar. During his sojourn to Europe in search of gene Watson first started his exercise in Kalckar's laboratory in Denmark. Kalckar's mental attitude was quite different from others. He liked everything in Calcutta, the horrible city and enjoyed his stay thoroughly in the same place which the Kornbergs hated. The musical soiree on the streets in the early morning hours would wake him up and he felt the bliss. The Hindu philosophy was very much in his mind when he saw the funeral procession of an old man accompanied with music.

**Plate 1 (a)**
**Fig. 1.3 : Skeletal and space-filled models of DNA.**

(a) The skeletal model clearly shows the major and minor grooves in both B(left) and A(right). DNA occurs naturally in B form and assumes A form in less hydrated state. For discussion see the text.

**Plate 1 (b)**

(b) The space filled model is built taking into account the actual sizes of the atoms involved. It is clear from the model that DNA is not a tubular structure. There is hardly any space between the two strands. Hydrogen-bonded base pairs fill up the gaps.

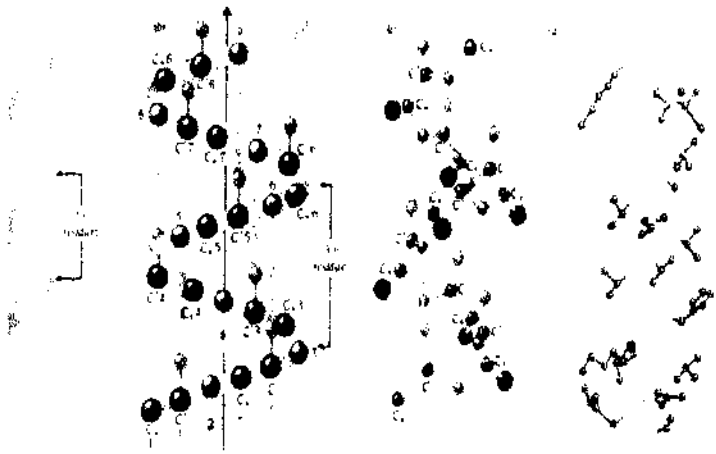


**Plate 2.**

Fig.2.7. α-helical structure present in protein. Oxygen atom (part of the peptide linkage) of the first amino acid residue is hydrogen bonded with the N atom (part of the peptide linkage on the 4$^{th}$ amino acid residue). Numerous hydrogen bonds, although weak make the α-helical structure stable.

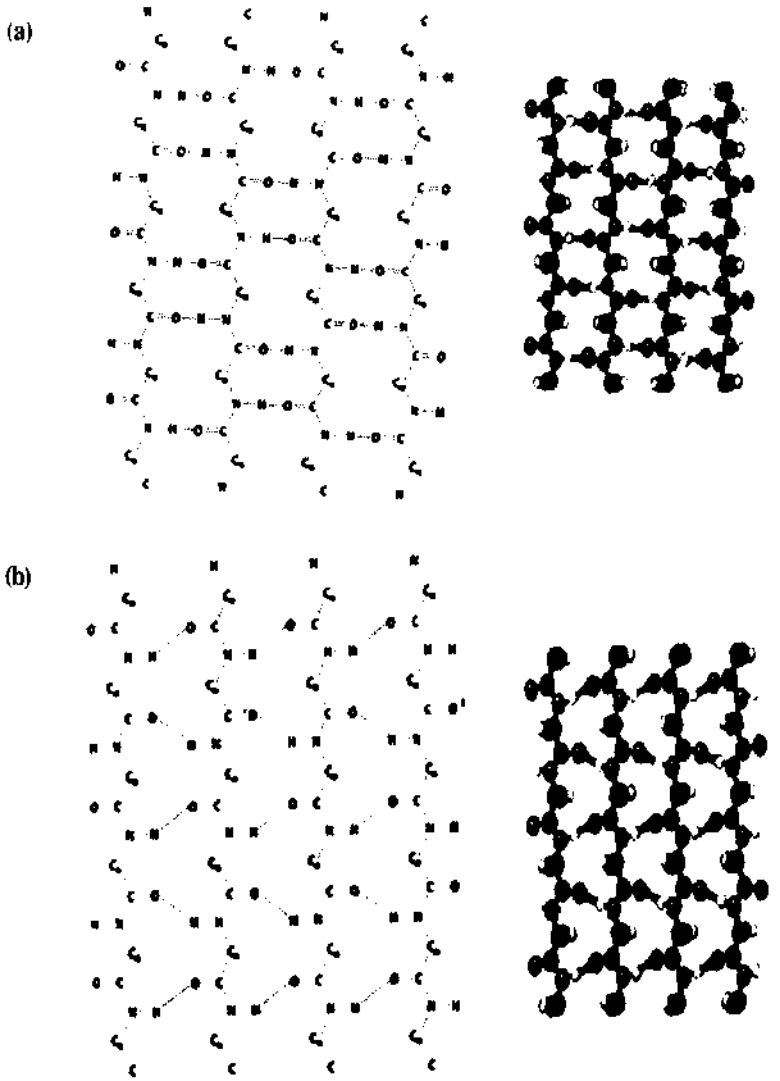



**Plate 3.**

(Fig.2.8) β-sheet structures in protein, schematic (left), and space-filled models (right). (a) anti-parallel β-sheets (b) parallel β-sheets. Details are described in the text. (c) & (d) are positioned in the text.

(a)

β chain of haemoglobin

Myoglobin

(b) Haemoglobin has four chains (two pairs of identical chains of α and β). The four chains are held together by various types of forces as mentioned in Figs.2.5

(c) The polypeptide chain of the enzyme pyruvate kinase has both α and β structures (red: α / β barrel, blue: α, green open twisted structure). Cylindrical structures α represent helices, flat arrows represent β strands. There are three domains in the structure. N and C indicate terminals of the chain are also indicated.

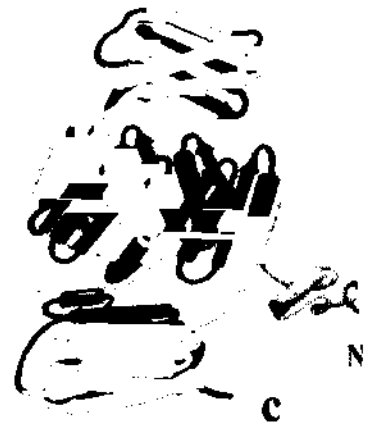



**Plate 4.**

(Fig.2.9) Quarternary structures of (a) myoglobin, β chain of haemoglobin, (b)haemoglobin and (c) pyruvate kinase

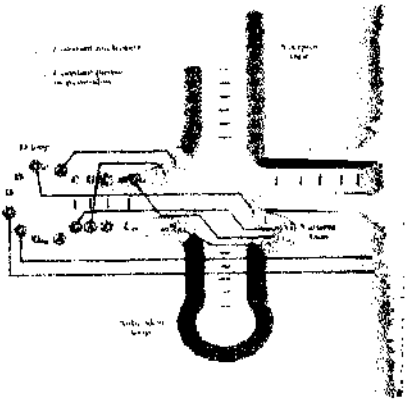


**Plate 5**

(Fig.2.12). Clover leaf model of transfer RNA depicting its secondary structure. This is based on stem-loop structure, generally present in RNAs. The stems having A type of structure are formed according to Watson-Crick base pairing rule. The CCA terminal and the different loops, TψC, dihydrouridine, anticodon etc. are marked and explained in the text.

**Plate 6**

(Fig.2.13). Three-dimensional structure of transfer RNA. The structure was elucidated from the X-ray crystallographic studies of Alexander Rich and Aaron Klug. The structure is derived mainly by folding TψC and pseudouridine loops. The CCA terminal on the one hand and anticodon loop on the other (both functional) remain exposed. As a whole the structure looks like the inverted letter L with the two functional groups at two ends.

He would murmur, 'That's the way one should adorn the reality of life and death'. The cellar of his house at Bethesda was filled up with drinks and books, mostly philosophical, of which many were on Indian philosophy. He would talk to me for hours about Indian philosophy of which I had very little knowledge. It did not matter, he would sometimes murmur to himself. He was misinterpreted by Watson in 'Double Helix' without realizing that like Leloir's contribution his would be largely in the area of glucose to galactose conversion. Lactose, which is the major sugar present in milk, is a combination of glucose and galactose and for this reason is a poison for galactosemic children. His greatest contribution was the discovery of the cause of this hereditary disease known as galactosemia. Milk, which is the source of life for children, is poison for those suffering from the disease. He did show that one of the enzymes which catalyses the step 2 in the conversion of galactose to glucose

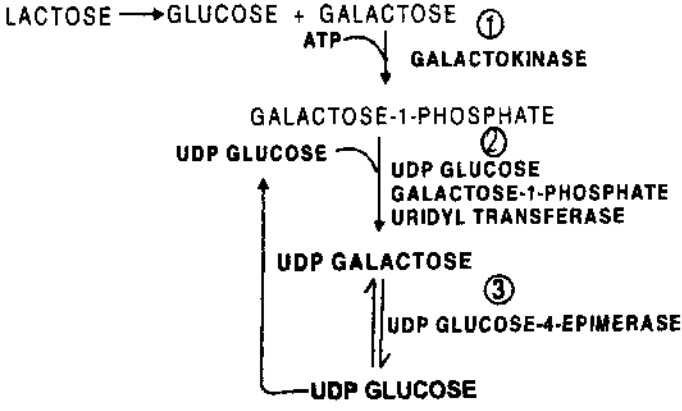



Fig. 3.5. Metabolism of galactose to glucose and the defect in galactose metabolism leading to galactosemia. Galactose (specially present in milk, primary food for children) is converted to glucose through uridine diphosphate glucose (UDPG) pathway. The step 2 of the pathway is usually blocked in galactosemic individuals. As a result galactose-1-phosphate accumulates and interferes with the metabolism of glucose and also induces toxicity.

(Fig.3.5) was missing in such children. He subsequently developed the method for detection of the enzyme UDPglucose galactose-1-phosphate uridyl transferase in the liver biopsies of the children and eventually in blood samples which was of great help in early diagnosis. Avoiding milk or any galactose-containing food was the only treatment, in a true sense. The defect lay in the parents of the child. As we already know, we inherit two sets of chromosomes one from the father and the other from the mother. If one is defective normally it does not matter much but if both are defective then there would be a problem because of the absence of the enzyme. If both the parents have the defective gene in one of their chromosomes and by chance the child inherits the defective genes from both the parents he would suffer from the disease. If galactose accumulates without being converted to glucose it will act as poison. The discovery of the cause of galactosemia is next in importance to the discovery of the cause of sickle cell anaemia by Linus Pauling and later Ingram. Those were preliminary studies on hereditary defects at the genetic level on the basis of one gene-one peptide theory. That was also the forerunner of the knowledge of the defects of DNA in hereditary diseases. The scientists and some times the historians also hardly put all the facts in proper perspectives.

Those were the golden days for Bose Institute as far as I was concerned. Joshua Lederberg followed Herman Kalckar. He was returning to U.S.A from Australia through India after doing the famous experiment with Gustav Nossal in Mac Farlane Burnet's laboratory to prove Burnet's clonal selection theory. It should be mentioned here that Niels Jerne, the Danish immunoligist is mostly credited for this theory. Lederberg knew how to isolate and examine a single cell under a microscope. That was of great help to establish the clonal selection theory for the production of antibodies, which Burnet strongly propounded in the face of a number of alternate suggestions. Lederberg is one of the few scientists who made fantastic contributions for which he won the Nobel Prize at an early age. He discovered two important

phenomena, transformation and transduction, both involved in gene transfer. In his usual authoritative style Lederberg gave a brilliant lecture. He had already done the Nobel Prize winning work on transduction but did not yet get the award at that time. There was a large crowd in the lecture Hall to hear him. Suddenly, J. B. S. Haldane arrived in his loose Indian saffron-coloured *Pyjama and Kurta.* Lederberg's facial expression changed immediately when he saw the eccentric giant with the unfathomed talent enter the lecture hall. Haldane had done all sorts of experiments, even some on himself. He was one of the early propounders of one gene-one enzyme hypothesis. His contributions to biology and statistics were profound. His words were final in most of the matters. He left England in disgust and chose the Indian soil as his home for living. But his love for the Indians did not last too long. At that time he was in Calcutta working at the Indian Statistical Institute. He was in trouble himself or might have created trouble for the Director Prasanta Chandra Mahalanobis who was the father of statistics in India. Haldane sat in the front row and looked like he chewing pan (betel leaf). Within a short time following the start of the lecture he closed his eyes and his head drooped. It appeared that he was in a slumber. Lederberg's jubilant feeling vanished. The important 'audience' who adorned the lecture hall seemed not to be interested in his lecture. There could not have been a greater disappointment for Lederberg. When the lecture was opened for discussion, Haldane, the indifferent 'audience' started spearheading the discussion with piercing questions. Lederberg handled those joyfully and brilliantly. His jovial expression reappeared. The giant must not have slept as he thought. Was it only an outward behaviour? Was his inner mind fully awake? Nobody could say. After all he believed in Indian Philosophy! It was fortunate for our country that he chose it as his last resort but the country could not recognize the 'true Indian' in him. That was the tragedy!

In spite of all these activities I was getting frustrated with my

beloved Institute, more so with the once loving Director-Cum-Dictator, who induced me to get involved in biology. There were a number of reasons behind it. I was seriously thinking of writing to several well-known scientists in U.S.A. to find a temporary berth somewhere. One of them was Severo Ochoa, then at the Department of Biochemistry of the New York University School of Medicine. That was a big disappointment; there was no space in his laboratory. But the second letter offering the post doctorate fellowship quickly followed. I was surprised and overjoyed. Later Bernie Horecker (who was at that time the Chairman of the Department of Microbiology, a floor below) told me that Ochoa met him on the stairway after writing the refusal letter, mentioned above. He changed his mind after Horecker overenthusiastically recommended me. Very soon Nobel Prize awards of 1959 were announced. Ochoa and Kornberg (master and disciple) shared the prize in medicine that year, Ochoa for RNA synthesis and Kornberg for DNA synthesis. Before leaving for Sweden to attend the award ceremony Ochoa wrote to me 'Although the prize is cited for RNA synthesis the enzyme which I discovered does not synthesize RNA under physiological condition. We have to look for another enzyme.' He was fully aware that the citation of the award was wrong.

The enzyme which he initially discovered and was most talked about those days was polynucleotide phosphorylase. It catalysed the synthesis of RNA-like material but not RNA, which occurs in nature. The enzyme was originally isolated from the cell extract of *Azotobacter vinelandii*, a nitrogen fixing organism. Later this enzyme was found to occur widely. A typical example of the reaction catalyzed by the enzyme is the following. It should be noted that $Mg^{++}$ is essential for this reaction.

$$nUDP \rightleftharpoons \text{Poly U} + nP_i \text{ (inorganic phosphate)}$$

(Uridine diphosphate) (Polymer of UMP)

In other words, an RNA-like material consisting of uracil only was produced from UDP. The reaction can be carried out in reverse

direction as well (hence the double arrow). For example, if synthetic PolyU is treated with the enzyme in the presence of excess of phosphate PolyU will be phosphorolysed (phosphodiester linkage is broken with the help of phosphate) and UDP (with the addition of $P_i$ to UMP) will be produced. Hence the enzyme was named polynucleotide phosphorylase from the reverse reaction it catalyses. If a mixture of ADP and UDP are used as substrates for the enzyme a mixed polymer (copolymer) of poly AU is produced. It is a random copolymer (positions of A and U are not in any particular sequence). Similarly, if all the diphosphates (ADP,UDP, GDP, and CDP) are used in the mixture, the polynucleotide synthesised will have all the four nucleotides but not in any particular order.

$$nADP + nUDP + nGDP + nCDP + Mg^{++} \rightleftharpoons (AUGC)_n + 4nPi$$

Random polymer

This was the reason why there was doubt from the very beginning if such enzyme would synthesize RNA under physiological conditions in the cell. There was another strong objection; the reaction is freely reversible. Phosphate concentration under physiological condition is rather high and expected to force the reaction in the opposite direction or the enzyme will catalyze the breakdown of RNA instead of synthesizing it. It has been discussed earlier (Chapter1) that the synthesis of biological polymers (polysacharides, fatty acids, proteins etc.,) never take place by the reversal of degradation processes. The same mistake was repeated in the case of RNA.

Although the enzyme was ruled out as the candidate for RNA synthesis in the cell its discovery led to the understanding of many properties of RNA and eventually helped in breaking the genetic code. It was rather easy to synthesise a large amount of homopolymers, copolymers, etc. and study their properties. For example, poly A and poly U formed a double stranded polymer (Poly A: Poly U), as expected. One could study the dissociation of two strands (technically called the melting of the copolymer) in a spectrophotometer and

determine its melting temperature ($T_m$), which is defined as the temperature at which 50% dissociation takes place. The very basic work done with these polymers as well as DNA by Paul Doty and Julius Murmur was a milestone in molecular biology (Fig.3.6). GC-rich DNA had a higher melting temperature (due to three hydrogen bonds between the two bases G and C) than AT-rich DNAs (having two hydrogen bonds between A and T). Actually many of the subsequent discoveries were dependent on the knowledge and use of the synthetic polymers.

To Ochoa research was like 'horse racing'. He was one of the foster fathers of modern biochemistry and had not only watched biochemistry to grow but himself contributed richly to the growth of the discipline. He was practically a 'grand-father' to the younger biochemists. His ventures in a number of fields were invariably crowned with grand success. Those were primarily based on hard competition with others. With his great interest, enthusiasm and far sight he had invariably been at the forefront. But I could not somehow

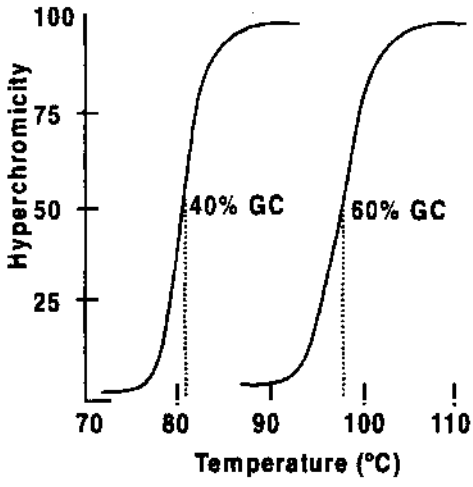



Fig 3.6. Thermal melting of double-stranded DNA or RNA. On gradually increasing the temperature the absorbance of the solution at 340 nm starts increasing and eventually attains the maximum value due to separation of the strands. The temperature at which the mid-point of the S-shape curve lies indicates 50% dissociation and is known as melting temperature. The process is called thermal melting. Higher is the GC-content higher is the melting temperature.

or the other reconcile with his 'horse racing' attitude towards science. To me he was a symbol of biochemistry, which had grown, in his own hands. As mentioned already, he got the Nobel Prize just before I joined his laboratory. It was long overdue but I was absolutely sure, the Prize was not his final aim. Even after receiving the award he had made significant contributions in the area of protein synthesis and remained quite active as long as he could physically do that. By the time I joined his laboratory he had already realised that the polynucleotide phosphorylase he had discovered might be a very important enzyme and yet it was not responsible for RNA synthesis under physiological condition. The story was the same for glycogen synthesis as well as fatty acid synthesis, which has already been told. Several groups involved in the game were fully aware that a different enzyme would be responsible for RNA synthesis and that would most probably use DNA as template. In the mean time a short communication appeared in the *Journal of American Chemical Society* from Samuel Weiss's laboratory at Chicago claiming the existence of such an enzyme in the nucleus of rat liver cells. When I joined Ochoa's laboratory in early 1960, a German post doctoral fellow (Hans Kruger) was struggling with the rat liver nuclei preparations but was unable to repeat Weiss's experiment. After my arrival and further attempts with no success in repeating Weiss's results Ochoa decided to switch over to his favourite organism *Azotobacter vinelandii*. We had some preliminary indications regarding the existence of such an enzyme in the cell-free extract of this microorganism but were always under tension as we knew that two other laboratories, one, the laboratory of Jerard Hurwitz in the sister Department of Microbiology and the other of Audrey Stevens then at the National Institutes of Health were also frantically searching for the enzyme. There were other laboratories as well engaged in a similar pursuit. Two more Indians besides myself (B.B. Biswas of Bose Institute in Richard Abraham's laboratory in the University of Pittsburgh at Pittsburgh and Nirmala Maheswari in James Bonner's laboratory in Caltech, California) were also involved

in the race. The former was working with the thymus nuclei and the latter with pea embryo, or in other words, all the three systems, bacteria, animals and plants were explored with the conviction that all organisms should have the enzyme to translate the message of DNA. So in the language of Ochoa it was really "horse racing". Following Kornberg's work on DNA synthesis all of us were using ribonucleoside triphosphates as precursors of RNA and one of them was radioactive. We could demonstrate that the radioactivity from $^{14}$C-ATP was incorporated into trichloroacetic acid-insoluble fraction (RNA being precipitated with this acid and thus separated form free radioactive triphosphate) in the presence of other three triphosphates (GTP, UTP and CTP) but the dependency on DNA of such synthesis was not highly reproducible. By that time the idea had emerged that RNA synthesis under *in vitro* conditions must take place using DNA as template.

To be precise, each competing group knew that they were near the goal and each group was aware of the shortest route but the credit depended on the forceful projection of the sketchy data. That actually happened in the Federation Meeting held at the Atlantic city during April, 1961. Ochoa presided over the session and presented our paper. I felt somewhat disappointed as I was hoping that he would ask me to do the same. Perhaps he did not have faith in me regarding the thrust I should put during the presentation. Junior authors from different groups like B. B. Biswas from Abrams' laboratory and Tom August from Hurwitz's Laboratory presented their results. Audrey Steven's work was highly applauded. She presented the work, as she was the sole worker in her laboratory. Some amount of confusion prevailed in that meeting but it was quite obvious that the enzyme RNA polymerase was discovered in several laboratories simultaneously. Textbooks usually give the credit to Audrey Stevens for this. However, it took some time to purify the enzyme thoroughly and understand the reaction catalyzed by the enzyme. My next attempt was to

characterize the RNA synthesized on the DNA template and to establish the relationship between the two structures. But as the enzyme was partially purified the results were not unequivocal. So those could never be published although I was fairly convinced that one of the strands of the double-stranded DNA was copied, the other one was not. These observations were eventually made in many laboratories. It was left to Bob Chamberlin, a young graduate student of Paul Berg at Stanford Medical School, to purify the *E. coli* enzyme and his method is accepted as a standard protocol even today. Our unfinished job in Ochoa'a laboratory was later completed by Joe Krackow who purified and characterised the *A. vinelandii* enzyme thoroughly and studied its properties. One should remember that the discovery of the enzyme was indeed a milestone in the route of understanding the function of the gene at the molecular level. Another milestone (breaking of the genetic code) was laid almost at the same time with the identification of messenger RNA and the demonstration of its association with ribosomes, on one hand and the discovery of DNA-dependent RNA polymerase, on the other hand which finally paved the way of breaking the genetic code.

The mechanism of RNA synthesis is very much similar to DNA synthesis. Ribonucleoside triphosphates (ATP, UTP, GTP and CTP) are the substrates and DNA (double stranded) is the template. $Mg^{++}$ is also required in this synthesis as in DNA synthesis. The reaction is written in the following fashion.

$$n(\text{ATP, GTP, CTP, UTP}) + Mg^{++} \xrightarrow{\text{DNA-dependent RNA polymerase}} (A,G,C,U)_n + 4nPPi$$
$$+ (\text{DNA as template})$$

The difference between the two polymerases is that DNA polymerase copies both the strands while only one of the two strands (so-called sense strand) is copied by RNA polymerase. Initially it was believed that only one strand carried the message while the other strand

did not. But this notion was found to be incorrect as it was observed later that the message could be on either of the two strands. However, if the message is on a stretch of one strand the complementary strand normally would not have any message as shown below.

RNA 5′→ 3′ 5′ → 3′ 5′→3′
DNA 5′----------3′
DNA 3′----------5′
RNA 3′ ← 5′ 3′← 5′

As in the case of DNA synthesis, RNA synthesis also takes place from 5′ to 3′direction although it may not necessarily start at the 5′ end of DNA. It may be recollected that copying of DNA starts from the origin of replication, which could even be in the middle.

Not only mRNAs but also all types of RNAs like transfer RNAs, ribosomal RNAs etc., are synthesized by the RNA polymerase on the DNA template or, in other words, DNA contains the information (in form of complementary base sequences) for all types of RNAs. In eukaryotes, however, three different enzymes do the different jobs. RNA polymerase I located in the nucleus is mostly concerned with the synthesis of ribosomal RNAs, whereas RNA polymerases II and III are involved in mRNA and tRNA (as well as 5S RNA) syntheses respectively. Mitochondria contain only one enzyme (different from other polymerases) similar to bacteria.

It is interesting to recall at this stage that the giant enzyme chemist Kornberg missed the enzyme involved in RNA synthesis. [He published a paper in the Proceedings of National Academy of Sciences on the isolation and purification of polynucleotide phosphorylase in *E. coli*.] Actually he was misled by the earlier discovery of the enzyme in *Azotobacter vinelandii* by his *guru* Severo Ochoa. Kornberg and Littauer (an Israeli scientist working as post-doctoral fellow in his laboratory) observed the incorporation of radioactivity from $^{14}$C-ATP

to RNA in the extracts of *E. coli*, which might have been due to DNA-dependent RNA polymerase. But they thought that ATP was breaking down to ADP and thus being incorporated into RNA catalyzed by polynucleotide phosphorylase as Ochoa's enzyme did. Littauer looked for the enzyme and got it and therefore lost interest in RNA synthesis. During my visit to Kornberg's laboratory I asked him, 'Do you think that you would have discovered the DNA-dependent RNA polymerase'? The wizard of DNA said, 'Yes I was misled, I was not smart enough with RNA.'

I was busy in the lab to wind up because my plan to return to India and to try my luck again there was finalized. I was supposed to leave USA after a month's tour in the country under the Rockefeller Foundation travel grant. Peter Lengyel stepped into my lab then. Peter was a Hungarian refugee, doing his Ph.D. in Ochoa's laboratory at a late age. He was quite intelligent and it was natural that Severo Ochoa liked him very much. Peter asked me, 'Debi, how about trying poly U as messenger RNA?' On the basis of today's knowledge that would have been the most desirable experiment. Jacob and Monod's concept of messenger RNA was practically established although not published till then. The DNA-dependent RNA polymerase was already discovered and the message that DNA is transcribed as RNA was on everyone's lip. Naturally, the stage was set for the discovery of the genetic code. All the elements were ready to crown the discoverer with the Nobel Prize. It was practically a case of good luck for someone. The same situation developed earlier but did not quite happen for the discovery of DNA-dependent RNA polymerase.

I have wandered away from Peter's pointed question. I don't blame Peter but I was over-cautious and believed too much of Crick's piece of statement or utterance. There was a lot of theoretical work going on for the genetic code, a good deal of assumption was made from time to time, calculations were based on doublet and triplet codes (two or three bases standing for an amino acid). George Gamow

initiated a 'Tie club'[2] in California to break the code by wearing ties marked with different amino acids. Twenty codes for 20 amino acids, quite natural. But nature is sometimes unnatural and not always sympathetic to Crick. He had published a fascinating (at least for me) paper in *Nature* on triplet code. He had a strong argument for non-overlapping triplet code. Everyone was aware that 'doublet' code would not serve the purpose as it would make only 16 ($4^2$) codes (or codons as used these days), which will fall short of 4 for the 20 amino acids present in the proteins. Triplet codes generated from four bases would lead to 64 ($4^3$) possibilities, too many for 20 amino acids. Crick had a clever idea. Triplets UUU, AAA etc. should be ruled out as those codes might be misleading. So 60 (64-4) codes were expected to serve the purpose. If one assumes that the code is non-overlapping then one-third would be the safest number not to confuse the reading. One third of 60 leads to the magic number of 20. Without any imaginative power I reciprocated to Peter, UUU is not likely to be a code as 'argued by Crick'. What a fool I was! Peter had no faith in me. He asserted, 'Still I shall try but I have no experience in protein synthesis' (he was working at that time on methylmalonylCoA isomerase, an enzyme which catalyses the inter conversion of methymalonylCoA and succinylCoA). He continued, 'I have to call Joe from Cold Spring Harbor for doing that experiment.' Joe Speyer was working with Ochoa on protein synthesis without much progress. That summer he went to Cold Spring Harbor to attend the famous phage genetics course. Peter also applied for admission but for some reason or the other he was not offered a chance. He immediately called up Joe 'I have a clever idea, we have plenty of poly U around. How about trying it as messenger RNA in your protein synthesizing system?' He gladly agreed to come to New York during the weekend and try the experiment. Post lunch Joe called back and gave the most exciting and equally disappointing news (for Peter). The experiment had already been done in Nirenberg's laboratory[3] at the National Institutes Health, Bethesda. John Mathaei who was involved in the experiment

had also gone to Cold Spring Harbor to undergo the training. My wife Maharani was also attending the course. During lunch Joe narrated the clever idea of trying poly U as messenger RNA. Mathaei jumped and said that, the experiment already been done; poly U acts as the message for polyphenylalanine synthesis, or in other words, UUU is the code for phenylalanine'. As long as he kept it a closely guarded secret as Nirenberg had planned to make the dramatic announcement at the International Congress of Biochemistry to be held at Moscow shortly. Mathaei found it risky to keep it a secret any more. Still Joe came to New York, performed the experiment together with Peter. Thus the race began. This part of the story has not been placed in proper perspective by Judson.[1] True to his habit, Ochoa jumped in the area and took help of various types of synthetic polymers made with the help of his polynucleotide phosphorylase. This was a fantastic use of the products for the enzyme wrongly cited for RNA synthesis.

The royal battle started. Both Nirenberg's and Ochoa's groups used homopolymers and copolymers of nucleotides (A, U, G, C) and went on breaking code after code. Actually, that was a somewhat uncertain game as the exact sequence of the bases in the polymers was not known and thus did not yield absolutely certain results. Slowly codes of different amino acids were becoming clearer. At the crucial juncture Har Govind Khorana stepped in. When Nirenberg and Ochoa were struggling to determine the codes for the twenty different amino acids then Khorana at the University of Wisconsin had perfected the coupled chemical and biochemical methods to synthesize stretches of DNA and RNA of desired length and sequence of bases, as required. So it was easy for him to understand the genetic code without any ambiguity. His first synthetic messenger was UCUCUC..... which has the precise repeating sequence of U and C. Such polynucleotide is expected to have only two types of codes, UCU and CUC provided the codon was a three lettered one. He tested it in the protein synthesizing system and the polypeptide that was produced was identified as serine-leucine-

serine-leucine.... As shown below, if the reading starts from U the most likely product is serine-leucine-serine- leucine..... But if the translation starts from the second nucleotide there may be some ambiguity about which codon is for which amino acid. However, if the translation starts from the third nucleotide the situation will be the same as starting from the first nucleotide. It will be clear subsequently that it is possible to assign the correct codon (UCU stands for serine and CUC for leucine).

```
 1  2  3  4  5  6  7  8  9 10 11 12 13 14
 U  C  U  C  U  C  U  C  U  C  U  C  U  C
*<1 ser 3><4 leu 6><7 ser 9><10 leu 12>
   *<2 leu 4><5 ser 7><8 leu 10><11 ser 13>
      *<3 ser 5><6 leu 8><9 ser 11><12 leu 14>
```

Similarly he used another copolymer of AC with the sequence ACACAC.... This led to the synthesis of a polypeptide containing threonine and histidine in sequence. It was deduced that ACA is the code for threonine and CAC for histidine. That had to be further confirmed.

```
 1  2  3  4  5  6  7  8  9 10 11 12 13 14
 A  C  A  C  A  C  A  C  A  C  A  C  A  C
*<1 thr 3><4 his 6><7 thr 9><10 his 12>
   *<2 his 4><5 thr 7><8 his 10><11 thr 13>
      *<3 thr 5><6 his 8><9 thr 11><12 his 14>
```

Similarly RNA with three bases in a repeating pattern was expected to yield three different types of polypeptides (homopolymers) depending on whether the reading frame starts from first, second or third letter as shown below.

```
 1  2  3  4  5  6  7  8  9 10 11 12 13 14
 A  A  G  A  A  G  A  A  G  A  A  G  A  A
*<1 lys 3><4 lys 6><7 lys 9><10 lys 12>
   *<2 arg 4><5 arg 7><8 arg 10><11 arg 13>
      *<3 glu 5><6 glu 8><9 glu 11><12 glu 14>
```

Actually individual homopolymers of lysine, arginine and glutamic acid were obtained and from further experiments it was established

that AAG, AGA and GAA stood for lysine, arginine and glutamic acid respectively. Similarly, copolymers of tetra nucleotide UAUC resulted in a single polypeptide with the repeating pattern of four amino acids Tyr-Leu-Ser-ile.

```
 1  2  3  4  5  6  7  8  9  10 11 12 13 14
 U  A  U  C  U  A  U  C  A  A  U  C  U  A
*<1 tyr 3> <4 leu 6> <7 ser 9> <10 ile 12>
   *<2 ile 4> <5 tyr 7> <8 leu 10> <11 ser 13>
      *<3 ser 5> <6 ile 8> <9 asn 11> <12 leu 14>
```

In all the three reading frames similar polypeptides of repeating four amino acids in a particular sequence were as expected and experimentally proved to be so.

However, (AUAG)n and (GUAA)n led to the synthesis of only di and tri peptides and thus aroused the suspicion that there must be codons for chain termination. This resulted in the identification of two stop codons (chain terminators) UAG and UAA From similar experiment the third stop codon UGA was also ascertained.

```
 1  2  3  4  5  6  7  8  9  10 11 12 13 14
 A  U  A  G  A  U  A  G  A  U  A  G  A  U
*<1 ile 3> <4 asp 6> <7 arg 9><10 stop 12>
   <2 stop 4> <5 ile 7> <8 asp 10><11 arg 13>
      *<3 arg 5> <6 stop 8> <9 ile 11> <12 asp 14>
```

```
1  2  3  4  5  6  7  8  9  10 11 12 13 14
G  U  A  A  G  U  A  A  G  U  A  A  G  U
*<1 val 3> <4 ser 6> <7 lys 9> <10 stop 12>
   *<2 stop 4> <5 val 7> <8 ser 10><11 lys 13>
      *<3 lys 5> <6 stop 8> <9 val 11> <12 ser 14>
```

Within a short time all the codes (codons) became known (Table 3.1a). Crick[4], the grand master of DNA easily put all the codes in a tabular form (Table 3.1b) which immediately became the text book material. The first vertical row indicates the first letter of the codon,

the horizontal row indicates the second letter of the codon and the last vertical row indicates the third letter. All the amino acids (except methionine) have more than one code. Leucine has maximum number of six. The degeneracy is mostly due to the third position. This is known as "Wobbling". There are three stop signals (chain terminators) as discussed in the text. AUG (for formyl methionine) acts as the start codon in prokaryotes and methionine in eukaryotes. However, AUG is the codon for internal methionine as well. GUG which codes for valine may also act as a start codon. These codes can be presented in circular fashion as well (Fig.3.7) which appears to many as a better option in comparison to the tabular form. A wristwatch gifted to me by one of my students has the circular genetic code in place of regular dial.

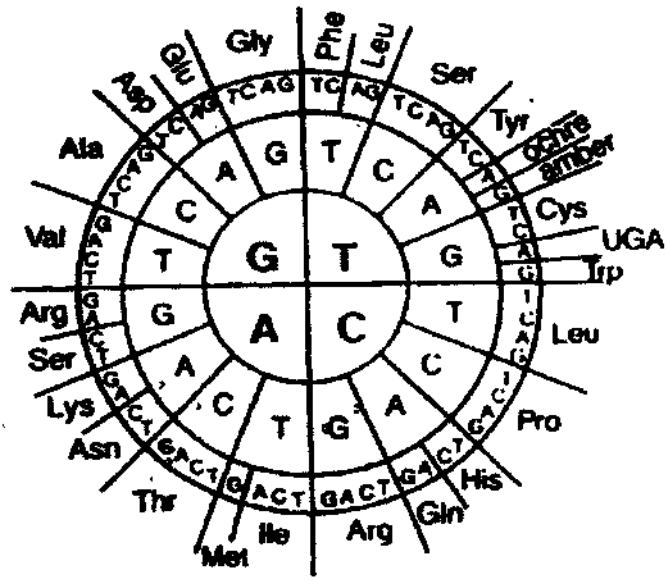



Fig 3.7. Genetic code presented in the circular form. Amino acids for which codons stand are written in the periphery of the code. The first base of the codon is written in the innermost ring, the middle base in the middle circle, the third base in the outermost circle. "Wobble" bases can be readily identified although one has to look around to identify all the codons for a particular amino acid.

**Table 3.1 a**

**The Genetic Code**

| Amino Acid | RNA Codon |
|---|---|
| Alanine | GCA GCC GCG GCU |
| Arginine | AGA AGG CGA CGC CGG CGU |
| Asparagine | AAC AAU |
| Aspartic acid | GAC GAU |
| Cysteine | UGC UGU |
| Glutamic acid | GAA GAG |
| Glutamine | CAA CAG |
| Glycine | GGA GGC GGG GGU |
| Histidine | CAC CAU |
| Isoleucine | AUA AUC AUU |
| Leucine | UUA UUG CUA CUC CUG CUU |
| Lysine | AAA AAG |
| Methionine | AUG |
| Phenylalanine | UUC UUU |
| Proline | CCA CCC CCG CCU |
| Serine | AGC AGU UCA UCC UCG UCU |
| Threonine | ACA ACC ACG ACU |
| Tryptophan | UGG |
| Tyrosine | UAC UAU |
| Valine | GUA GUC GUG GUU |
| Stop Codon | UAA UAG UGA |

**Table 3.1 b**

**Genetic Code**

| First Position (52 end) | Second Position U | C | A | G | Third Position (32 end) |
|---|---|---|---|---|---|
| U | Phe | Acr | Tyr | Cys | U |
| | Phe | Acr | Tyr | Cys | C |
| | Leu | Acr | Stop | Stop | A |
| | Leu | Acr | Stop | Trp | G |
| C | Leu | Pro | His | Arg | U |
| | Leu | Pro | His | Arg | C |
| | Leu | Pro | Gln | Arg | A |
| | Leu | Pro | Gln | Arg | G |
| A | Ile | Thr | Asn | Ser | U |
| | Ile | Thr | Asn | Ser | C |
| | Ile | Thr | Lys | Arg | A |
| | Met | Thr | Lys | Arg | G |
| G | Val | Ala | Asp | Gly | U |
| | Val | Ala | Asp | Gly | C |
| | Val | Ala | Glu | Gly | A |
| | Val | Ala | Glu | Gly | G |

It was no wonder that Nirenberg and Khorana shared the Nobel Prize in Medicine in 1968. Holley who determined the structure of tRNA for the first time (see Chapter 2) also shared the prize with them. The discovery of the genetic code was the second breakthrough after the seminal work of Crick and Watson. Molecular biology advanced at a rapid stride following the second one. There is a big difference between the first and the second. The structure of DNA as proposed by Watson and Crick in 1953 was quite unexpected but the second one was waiting in the corner to be solved. Someone only needed luck to strike the gold mine. The first one in a sense was a conceptual breakthrough, the other one was based on proper experimental design. It was a hearsay that Nirenberg and Mathaei used poly U in the incubation mixture for protein synthesis to keep RNase (in the extract) busy and leave messenger RNAs intact. Further, they were using $^{14}$C-phenylalanine as the marker amino acid to be incorporated into protein. It may or may not be true that they did not design the experiment to break the genetic code. Once it was broken Khorana's synthetic approach came to the rescue. Subsequent events developed very fast thereafter. The genetic code was found to be degenerate (one amino acid having more than one code) and universal (with limited exceptions and that also in case of some mitochondrial genes). The latter has great implications in the history of living systems, a strong indication that all the existing organisms were practically derived from one ancestor. It may not be exaggerated if we say on this ground that the bacteria and humans have the same origin.

Several new information later obtained about messenger RNAs was, however, unexpected. For example, it was observed mostly in animals and plants that the messages are present in split form (called exons) as shown in the Fig.3.8a. The introns divide the exons and normally do not carry any message. Exons are spliced (cut out and joined together) to make the message continuous and do its job as

shown in the figure. There are conflicting theories about the origin of exons and introns and the problem has remained unsolved. Some groups feel that originally there were separate pools of exons and introns and those were joined together in a way those are presented today. Others feel that introns fragmented the message and created the exons by inserting themselves within the messenger RNA. The so-called eukaryotic messages are further modified. They are sometimes capped with some specific groups like methyl guanosine (Fig.3.8b) and mostly they have polyA tails (Fig.3.8c). The function of the tail is not firmly established but it is widely believed that it decides the life of the messenger RNA. Unlike other RNAs messenger RNAs are unstable and having variable turnover rates. In prokaryotes their turnover is faster as compared to the eukaryotes. Another way of modification of the message following transcription is editing (Fig.3.8d) which will be discussed in Chapter 5. It has recently been found that messages could be overlapping. From one message two or more proteins can be synthesised by shifting the reading frame (Fig. 3.8e). This is not a modification in the true sense as the original message is read in different frame to generate a different protein. Elaboration of these complexities is beyond the scope of this book.

At this stage it will be of interest to many the narration of an incident regarding Khorana, which throws some light on his personal character. During my short visit to Khorana's laboratory one of his Post-doctoral fellows showed me a book[5] by Khorana, which depicts his basic work on phosphate esters, specially the chemical synthesis of polynucleotides with the use of carbodiimide as the condensing agent. This led to the phenomenal success in his research career leading to the Nobel Prize. I went through the book and decided to buy it. Khorana's postdoc insisted, 'Why don't you get a complimentary copy from him?' Naturally I asked him, 'Did you get this one in that way?' The reply was 'No, I tried but failed; why don't you try?' I did try but took a different approach. I asked him, 'Whether your book will

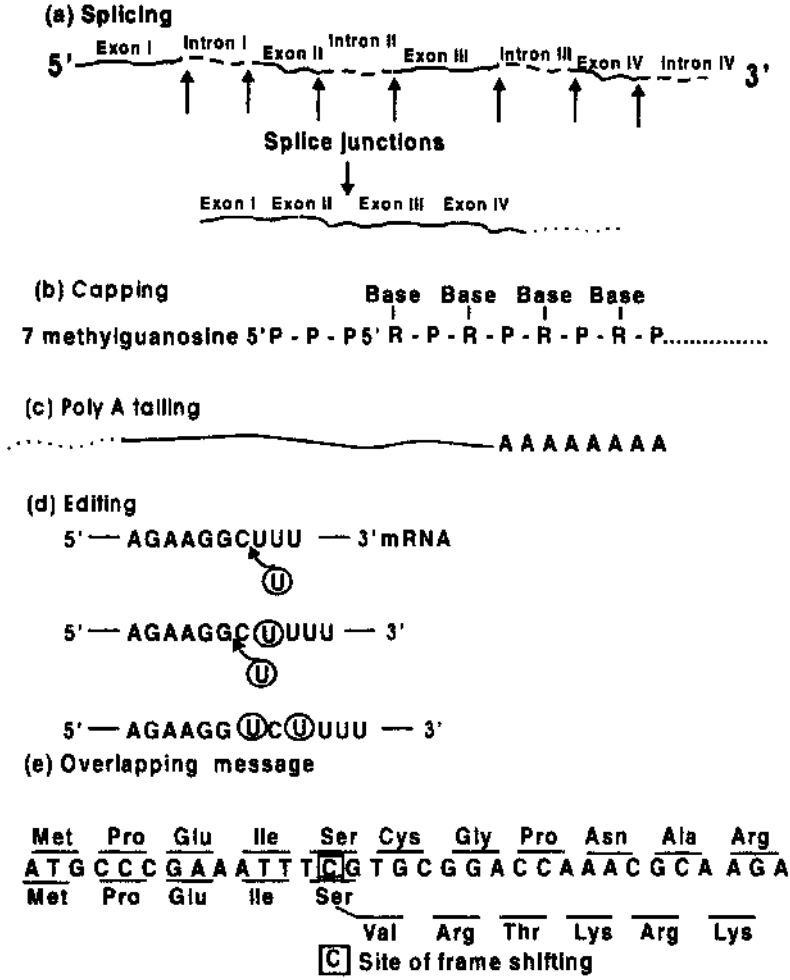



Fig. 3.8. Modification of messenger RNAs. (a) Splicing – the exons which are consecutive parts of the message separated by introns are joined together. (b) Capping - 7 methyl guanosine is added to the 5′ end of mRNA . (c) Poly A tailing - the 3′ end of the mRNA is modified by a chain of poly A. (d) Editing – the correct message is generated by substituting bases to make it meaningful (to be discussed in Chapter V). (e) Overlapping message – one reading frame may have more than one message which can be read by shifting the reading frame.

be readily available in the nearby bookshop so that I can buy it'. 'Don't worry I would give you a complimentary copy' was the reply. When I mentioned that to the postdoc he was very upset and tried to console himself saying that he gave it to you because you are an Indian. I objected, as it may not have been completely true. There is no doubt that he has a soft corner for India and his main concern was poverty in his country. He was keen to help India and Indians as much as possible.' I don't know whether he was convinced but became pacified. I still have the prized possession with Khorana's own handwriting; 'To Debi Burma' and his signature. Khorana loved India and he was a true 'Indian'. As I mentioned in the Prologue, he did not get a suitable job in the country although he was ready to sacrifice. As he hardly visited India; once I asked him, 'Why don't you visit India once in a while'? His reply was, 'I would rather help my relatives monetarily than to spend money on travel to visit them.' The main reason may be that he shuns publicity. The rousing reception he would have received from the Indian public would have overwhelmed him no doubt but that would have deterred him from working. In this connection a story befitting Khorana, published in a local newspaper way back, is interesting to recall. Unfortunately, I have forgotten the details and the name of the writer. In short, it is as follows. Khorana was offered the directorship of one of the National Laboratories of India and was taken straight from the airport to the Director's residence, which was located on the top floor of the Institute building. He was shown his itinerary for the next day which ran more or less like the following : welcoming some very important politicians visiting the Institute, chairing the building committee meeting, reception to some very important persons holding high office in the Government etc.etc... He innocently asked,'When do I do my experiments?' The prompt reply was, 'You will find plenty of opportunity for that. That is the reason why your residence is located on the top floor of the Institute building'. Next day airport passengers saw a gentleman with a brief case marked 'HGK' sneaking away from the International Airport

and boarding a flight for U.S.A. Though it is a story it is a correct depiction of the character of Har Govind Khorana. I want to repeat what I stated earlier. Whenever I had the opportunity to meet him he used to ask,'Is it possible to work in India?' He himself knew the answer better.

## References

1. *Eighth day of creation (the makers of the revolution in biology)*. H F. Judson, Simon and Schuster, New York 1979.
2. *What mad pursuit (a personal view of scientific discovery)*. F.H.C.Crick, Basic Books, U.S.A., 1988.
3. The Genetic Code II. M.W.Nirenberg, *Scientific American*, Volume 208 (March), pp 80-94 (1963).
4. The Genetic Code III. F.H.C.Crick, *Scientific American*, Volume 215 (October), pp 55-62 (1966)
5. *Some recent developments in the chemistry of phosphate esters of biological interest*. H.G. Khorana, John Wiley & Sons, Inc. (1961)

# Chapter - 4
# The musical instrument: The way it is played
## (*Ribosomes and Protein Synthesis*)

*The music of life is played by DNA (deoxyribonucleic acid). The beautiful features of this amazing player have been described in Chapter 1. The tune of the music is the messenger RNA (ribonucleic acid). The story of its discovery and function has been recorded in Chapter 3. The link between the two is ribosome, the musical instrument. The history of its discovery and function (known before the messenger RNA was discovered) has been told in Chapter 2. The details of the operation of the instrument in which the author himself was involved to some extent, are elaborated in this chapter.*

It is well-known and already mentioned that Paul Zamecnik could have received the Nobel Prize if he did not start with an eukaryotic system (rat liver) to explore the mechanism of protein synthesis[1]. As we have learnt, the prokaryotic systems are simpler than eukaryotic ones. In 1960, the meeting of the Federation of Biological Societies of U.S.A. was to be held at Atlantic City, New Jersey. Junior research workers (usually post doctoral fellows) presented their work done under the monitoring eyes of their mentors. I had just joined Severo Ochoa's laboratory and had sent the abstract of the work done in Bose Institute before leaving Calcutta. The paper was on 'Protein synthesis in the extract of *Azotobactor vinelandii*', perhaps the first example of a work done on protein synthesis in a prokaryote. The work was

initiated in the laboratory of R. H. Burris in 1956 and continued at the Bose Institute after my return to India. The work done in collaboration with Maharani, my would-be-wife, was scheduled to be presented in a session to be chaired by Paul Zamecnik. But I had no fund for travel and approached Ochoa to ask if he would support my travel although the work was not done in his laboratory. He reluctantly agreed. Zamecnik congratulated me for the presentation and specially mentioned in his concluding remarks that the work was done in a developing country. This minor incident has already been mentioned in Chapter 2 but is again mentioned here only for the reasons that work with bacterial system eventually solved many problems of protein synthesis. That was the start of my work in the area of protein synthesis to which I devoted the major part of my research career. The particulate preparations we used for those studies were nothing but ribsosomes (the term was already coined by that time). Shortly after joining the Banaras Hindu University at Varanasi in 1964 I got deeply involved in understanding the structure and function of ribosomes, an important area of molecular biology, which will be dealt with in this chapter.

Let me recapitulate whatever has been discussed about ribosomes so far. That will help us to understand the complications of the whole system better. There is no harm in repeating an exciting story. What is life? We are struggling hard to unravel that but as yet we are in darkness. What is the music of life? Do we know? Perhaps we do. The molecules of life, DNA, RNA, protein etc., play the music in concert. How? It took us almost 50 years, the entire period of the second half of the past century to understand that. We understand a lot, so we enjoy the music, sometimes even without understanding it. In the hustle-bustle of the present day life a cassette player is a good companion; a tape of choice provides the mental peace which all of us long for. If we have a glimpse of Nature we are overwhelmed with her exquisite beauty but if we look deeper into her we are amazed

with the mechanistic features, the engineering feats, of the musical instrument. Nature evolved its tape recorder or cassette player billions of years ago shortly after the creation of life, not just to play the music but also to sustain it. It took us a long time to discover the instrument among the numerous organelles present in the 'living cell'. Obviously, I am referring to ribosomes, which play the music of life.

The message of life which biologists refer to as genetic information is stored in a coded form in the master molecule called deoxyribonucleic acid or DNA, in short. The stored information is mainly in the nucleus of the cell. The message has to be transcribed and translated. The translational apparatus "Ribosome" is in the cytoplasm of the cell outside the nucleus. This is, however, not true for the prokaryotes, which have no partitioned nucleus. Any way, the message has to reach this apparatus to play the music. Messenger RNA does the job of the tape, copies the message from DNA, crosses the nuclear membrane and presents itself to the ribosomal site to produce proteins which are the architects of the cell. Further, DNA, the basic molecule of life can duplicate itself because it is composed of two complementary strands. The necessity of the double-stranded structure of DNA is reflected in the process of replication which is a 'must' for continuation of the species and even when the question of evolution of one species from another becomes a necessity.

As discussed in the earlier chapters, ribonucleic acid (abbreviated as RNA) has multiple attributes. It replicates itself like DNA, it is produced on DNA template and can reconvert itself into DNA, carries out protein synthesis and many other functions. Actually every step of protein synthesis involves one type of RNA or the other. Amino acids, the building blocks of proteins literally float around in the cytoplasm of the cell and are activated with ATP (the molecule which is the source of energy) before they can join hands to form the protein. The job has to be done in perfect sequence. A class of RNAs initially called as soluble RNAs and now transfer RNAs are referred to as

'adaptor molecules' which act as handles to pick up the activated amino acids from the cytoplasm and transport those to the site of protein synthesis (the cassette player or ribosome) which in turn, is itself composed of RNA and protein.

The message to arrange the amino acids in precise order comes from DNA. That message is in coded form and when transcribed produces the messenger RNA which is like a tape to be fed to the ribosome to translate the coded message. Thus, the translation of the message takes place and the protein is synthesized. Therefore, the central dogma was enunciated as follows:

$$\text{DNA} \xrightarrow{\text{Transcription}} \text{RNA} \xrightarrow{\text{Translation}} \text{Protein}$$

This sets the tune of the music of life.

We have already discussed that ribosome is a ubiquitous organelle present in each and every living system (except viruses which can not 'thrive' on their own), be it an animal, plant or microorganism. Without ribosomes cells cannot synthesise proteins, their functional entities. Without the production of proteins the cell would die. So ribosomes are one of the vital organelles (like organs of the human body) of the cell. Let us discuss here how it is done. Every sentence that is written in English starts with a capital letter. The words that constitute the sentence are placed in proper order to make sense and those are written invariably in one direction from left to right (unlike right to left in some languages). No language, I believe, writes in both directions. As the sentence is completed a full stop is put (sometimes a question mark is placed if the sentence raises a question). Without the stop signal there would be confusion due to the merger of two sentences. The next sentence again starts with a capital letter. So the capital letter and full stop help to convey the correct sense of a sentence and avoid confusion between the consecutive sentences. A genome of

either a lower or a higher organism synthesises a number of proteins and their messages are written in coded language in DNA. As it has been repeatedly mentioned, the message is transcribed first in the form of RNA and then translated in the form of a protein. Like individual sentences individual messages (for proteins) are translated individually although several messages may be present in one messenger RNA especially in the case of prokaryotes. Nature has devised a clever way of avoiding the confusion between the consecutive messages. Each message starts with a start signal and ends in a stop signal. Accordingly, the steps of translation can be divided into three distinct stages, start of the reading of the message, continuation of the reading till the stop signal is reached and the protein synthesis is terminated. All the three steps are shown in a summarized form in (Fig. 4.1).

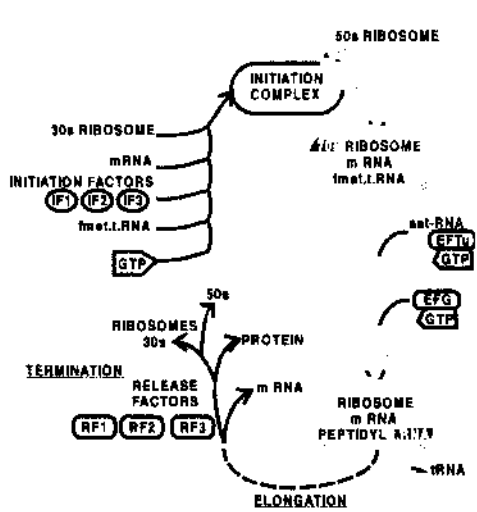



Fig. 4.1. Three steps of protein synthesis (initiation, elongation and termination). Extraneous factors (initiation, elongation and termination) help to carry out the functions at different stages. These factors are temporarily associated with ribosomes and then recycle. GTP provides the energy by its hydrolysis to GDP and $P_i$.

The process of assembly of the amino acids on the ribosomes to form a protein chain is a multistep process and as mentioned above, can be divided into three distinct stages (i) initiation, (ii) elongation and (iii) termination. At every stage a number of factors which themselves are proteins are involved. These are accordingly called initiation factors (IF), elongation factors (EF) and termination factors (TF) or release factors (RF). Finished protein chains are released from the ribosomes with the help of termination factors. The steps will be elaborated here with the help of the bacterial protein synthesising system, specially that of *Escherichia coli*, as the protein synthesis in eukaryotes is somewhat more complicated. It should be remembered that these factors are not ribosomal proteins but occur in the cytoplasm. Those associate temporarily with the ribosomes and get dissociated when the job is finished. Another important aspect to be emphasised at this stage is that at least the first two steps are energy-dependent and the supplier of energy is GTP instead of ATP. This was first detected by Fritz Lipmann, 'Master of biochemical energy' It should, however, be remembered that the very first step of protein synthesis i.e., activation of amino acids requires ATP and not GTP. At the start or the initiation step 30S ribosomes of *E.coli* form the initiation complex. Messenger RNA and three initiation factors (IF1, IF2 associated with GTP and IF3) are required for the formation of the complex (Fig. 4.1). Protein synthesis is initiated with the smaller subunit, the 30S ribosomes. The 50S subunit associates with it when the messenger RNA and formyl methionyl tRNA remain associated with 30S subunit and only when all the three initiation factors, specially IF3 get dissociated. This way the initiation complex formed is ready to start the elongation cycle during which activated amino acids are brought in with the help of the elongation factor Tu with bound GTP as per codons on the mRNA. Another elongation factor EF-G bound with GTP is involved prior to peptide bond formation. Both the elongation factors transiently associate with 70S ribosomes at each and every step of elongation. Finally when the synthesis of

the protein chain is completed it is released from the ribosomes with the help of one of the three release factors RF1, RF2, and RF3. The free 70S ribosomes then get dissociated to its two subunits to start the whole process all over again.

The initiation step is rather critical and needs some elaboration. The discussion will again center around bacterial system due to its comparative simplicity (Fig.4.2a). In the bacterial system all proteins are supposed to have the amino acid methionine at the start i.e., N-terminal end. The starter amino acid is, however, formyl methionine ($NH_2$ group of methionine is formylated as shown in Fig.4.2b). There is a special reason for starting with formyl methionine. The first peptide bond is to be formed by the addition of the amino group of the next amino acid (according to the genetic code) to the carboxyl group of methionine (attached to tRNA). However, there is equal probability of extending the chain in other direction i.e., forming the peptide bond between free $NH_2$ group of methionine and COOH group of the incoming amino acid. Since methionine is formylated (formyl group attached to NH of methionine) this probability is ruled out. In higher organisms (animals and plants) the system has been developed in such a way that no blocking of $NH_2$ group is necessary and the protein chain extends in one direction only. In this way the protein chain in any organism extends invariably in one direction, $NH_2$ terminal to COOH terminal. fMet is brought to the ribosomal site by the initiator RNA (which has initiator anticodon CAU) with the help of initiation factor 2 (IF2). Formylation of methionine takes place only after it is loaded on its own transfer RNA which is the initiator tRNA recognising the initiation codon AUG (Table 3.1a). The problem with methionine code should be discussed at this stage. Methionine has a single codon AUG. Although it is the starting amino acid in bacteria it may also be present within a protein chain, any position (according to the code). How can a single codon (AUG) decide for starting position (which is crucial) and also any other internal position?

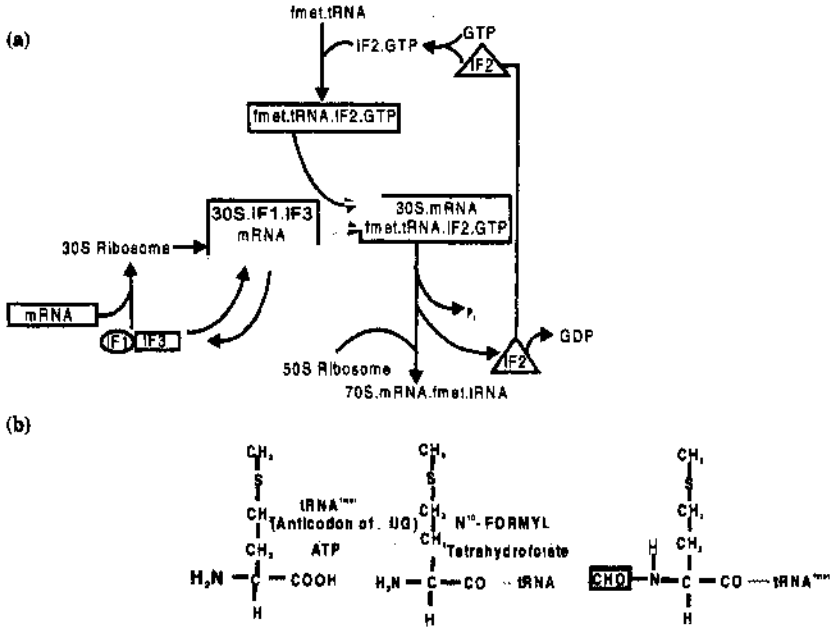



Fig 4.2. Initiation of protein synthesis (a) 30S ribosome forms the initiation complex with initiation factors IF1 & IF3 and mRNA. IF3 dissociates eventually. Formyl methionyl tRNA.IF2.GTP associates with 30S.mRNA to form the initiation complex. GTP is hydrolyzed and IF2.GDP dissociates from the ribosome. All the factors recycle as shown above. Finally 50S ribosome joins with the complex to form 70S.mRNA.fmet.tRNA complex. (b) Following its attachment to initiator-tRNA methionine is formylated with the help of tetrahydrofolate. (c) Difference in tRNA structure determines whether it is to be formylated or not.

Actually, there are two different transfer RNAs for the purpose. Transfer $RNA_f$ reads the starting anticodon, $tRNA_m$ reads the internal methionine codon. Methionine attached to the former (and not the latter) is formylated with the the help of tetrahydrofolate (Fig.4.2b) and this formylated fmet.$tRNA_f$ recognises the initiator AUG and hence does not allow met.$tRNA_m$ to this position. The latter (and not the former) is recognized by the internal AUG. One should, however, remember that the same methionine activating enzyme is responsible for loading both tRNAs having the same anticodon CAU (Fig.4.2c). However, the formylation of methione attached to $tRNA_f$ ( but not $tRNA_m$) takes place by the methylating enzyme with the help of tetrahydrofolate. Nature has thus devised a clever way to start the protein chain. Many proteins do not have methionine at the start as in most cases this is removed later. It is also known that GUG (code for valine) may also act as a starting codon. There is no problem with that as this amino acid has more than one codon.

At this stage the mode of functioning of IF2 needs some elaboration (Fig.4.2a). 30S ribosomes are associated with the initiation factors IF1 and IF3 as well as the mRNA. After these two factors dissociate leaving mRNA behind fmet.tRNA associates with 30S ribosomes with the help of IF2 and GTP. IF2 forms a complex with GTP. The binary complex IF2.GTP subsequently forms a ternary complex with formyl methionyl transfer RNA (fmet.tRNA). On binding of IF2.GTP.fmet.tRNA to 30S ribosomes GTP is hyrdolysed to GDP and Pi and IF2.GDP dissociates from the 30S ribosomes leaving the messenger RNA associated with fmet.tRNA behind. In this connection it should be remembered that as long as IF3 remains attached to 30S ribosomes, 50S ribosomes can not associate with it. Once the initiation complex of 30S ribosomes with messenger RNA and fmet.tRNA is formed, 50S ribosomes associate with 30S ribosomes (initiation complex) to form 70S ribosomes. Following the association of 50S ribosomes with 30S ribosomes the elongation of protein chain starts.

It has been shown by *in vitro* experiments that IF3 controls the association of 30S and 50S ribosomes in the way as shown in Fig 4.3. It was originally known as dissociation factor but later its name was changed to anti-association factor because it does not allow association of 50S ribosome as long as it is attached to 30S ribosome. However, there is no doubt that IF3 controls the dissociation and association of ribosomes. Much later Bernard Davis, the famous microbiologist who

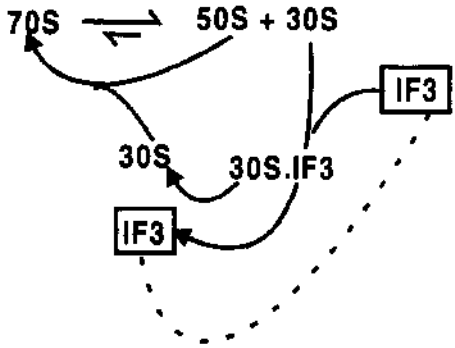



Fig 4.3. Initiation factor IF3-dependent association and dissociation of 30S and 50S ribosomes. As long as IF3 is associated with 30S ribosomes, 50S ribosomes can not associate with them to form 70S ribsomes. IF3 was originally known as dissociation factor but later designated as antiassociation factor (see text).

visited our laboratory at Varanasi, showed anguish for replacement of the term 'dissociation' with 'antiassociation'. This dig was at me due to my association with Severo Ochoa, who was responsible for it. Davis originally demonstrated the dissociation of ribosomes by IF3.

Before discussing the elongation step in detail it is better to throw light on the sites of binding of tRNAs to the ribosomes. As per the classical model, ribosomes are supposed to have two sites for transfer RNAs, A and P sites (Fig.4.4). Initiator tRNA is positioned at the P

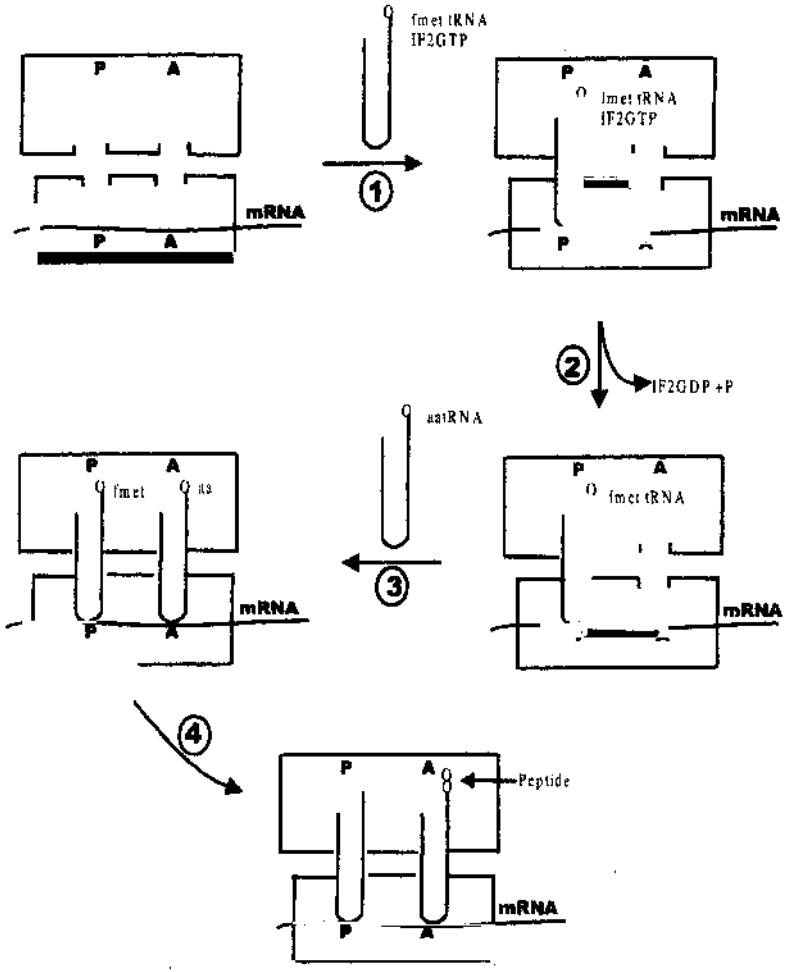



Fig. 4.4. Two-site-model of 70S ribosomes. Initiator tRNA associated with formylmethionine directly binds at the P (peptidyl) site but all other amino acids are positioned at the A (aminoacyl) site and are eventually transferred to the P site following the peptide bond formation.

site (peptidyl site, so called for reasons which will be discussed later). The other site is A site (or aminoacyl site). Later, three site model (with A, P, and E sites, the latter indicating exit site) was developed and is now universally accepted. This will be discussed a little later. Once the initiator tRNA along with formylmethionine is positioned at the P site the next transfer RNA according to the codon in the messenger RNA is brought in position with the help of another factor called elongation factor T or EF-T in short (Fig.4.5). There are two elongation factors EF-T and EF-G. The first one is actually composed of two factors Tu and Ts. A molecule of GTP can displace Ts from the complex and bind with Tu to form EF-Tu.GTP. This binary complex, in turn, binds to tRNA carrying the activated amino acid and brings it to the ribosomal site known as the aminoacyl site (A site). The site is so-called because the aminoacyl group is attached to incoming tRNA. GTP is then hydrolysed and EF-Tu.GDP is released from the ribosome leaving behind the tRNA along with the amino acid (rather aminoacyl group) at the aminoacyl (A) site. Thus, the first amino acid i.e., formylmethionine and the next amino acid (according to the code) are positioned next to each other and are subsequently connected by a peptide bond. The formylmethionine is transferred from its own tRNA to the amino acid attached to the aminoacyl tRNA (to be discussed shortly). The tRNA after transfer of formyl methionine to aminoacyl tRNA leaves the ribosome making the peptidyl site (P site) vacant. Simultanesouly, the aminoacyl tRNA moves to the peptidyl site (P) from amino acyl (A) site along with the peptide and the aminoacyl site (A) again becomes vacant and ready to accept the next tRNA with the third amino acid as per codon. Now it is clear that the peptidyl (P) site is so called because the peptide carrying tRNA is stationed at that site at every step of elongation. The movement of tRNA (along with mRNA) from A site to P site is called translocation and this is effected by the other elongation factor EF-G. Both EF-T and EF-G are, therefore, called translocation factors.

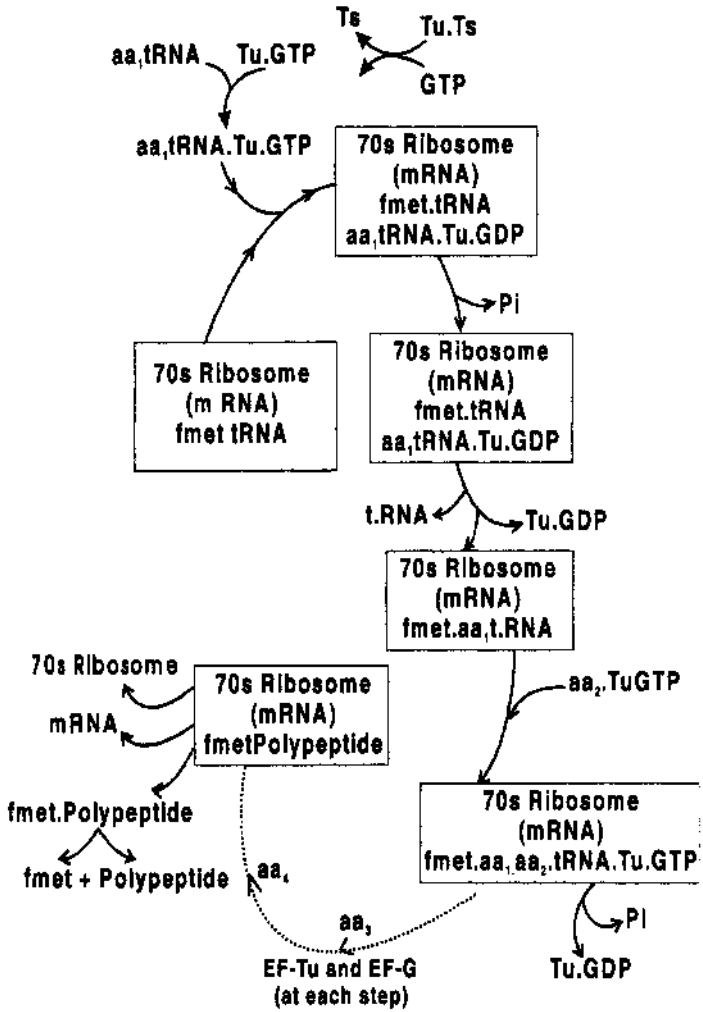



Fig. 4.5. Elongation of protein chain on 70S ribosomes. The elongation step effected by two elongation factors EF.Tu and EF.G which act in succession. EFT helps to bring aminoacyl.tRNAs to the A site on the 30S ribosomes and EFG responsible for the translocation. GTP is essential for the functioning of both th factors. The details have been described in the text.

Later it was shown that there could be a third site called the Exit site or E site. It has been demonstrated that at least three codons at a stretch or 9 bases (3 bases form one codon) are associated with the ribosomes at a time as shown in the Fig. 4.6. Once the tRNA at the peptidyl site delivers its amino acid for the peptide linkage to the amino acyl moiety attached to the tRNA at the amino acyl site the unloaded tRNA (now without the peptide) does not immediately leave the ribosomal site but migrates to the Exit site, naturally still in contact with the mRNA. Once it comes to the Exit site and is ready to leave it gets ejected from the ribosome during the subsequent movement of mRNA. This is known as the three-site model and now holds off

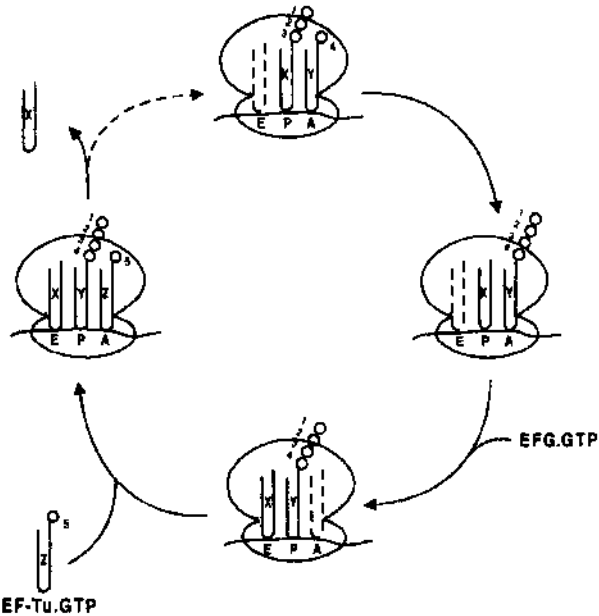



Fig. 4.6. Three-site-model for protein synthesis. Aminoacyl tRNA after receiving the peptide moves from A site to P site. Free tRNA transiently associates at the third site (E) on ribosome before its final departure. The process continues in a cyclic fashion as new tRNA enters at the A site, dictated by the successive codons on the mRNA.

general consensus. Recently X-ray crystallographic studies have confirmed the existence of the three sites, which would be discussed at the end.

As mentioned before, elongation factor T is composed of two components, Tu and Ts (Fig.4.5). When the factor interacts with GTP it dissociates due to the binding of Tu to GTP.

$$Tu.Ts + GTP \longrightarrow Tu.GTP + Ts$$

EF Tu.GTP then associates with an aminoacyl tRNA and the complex. EFTu.GTP is then ready to be transported to the aminoacyl site of the ribosome for the first peptide bond to be formed. It should be remembered that the particular amino acid is selected according to the next codon on the mRNA. Before the peptide bond is formed GTP associated with aminoacyl.tRNA is hydrolysed and EF.Tu GDP so formed dissociates from ribosome leaving aminoacyl tRNA behind. EF.Tu subsequently released associates with Ts and Tu.Ts so formed becomes ready to initiate the next cyclic process. Once the peptide bond is formed between formyl methionyl (bound to tRNA) and incoming aminoacyl (also bound to its own tRNA) moieties, tRNA bound to formyl methionine becomes free to be transported from the P site to the Exit site. The process is coupled with translocation of the peptidyl tRNA from A site to P site effected by the other elongation factor G (EFG). The latter associates with GTP and is then capable of binding to 50S ribosomes to carry out the operation. In the process GTP is hydrolysed to GDP and Pi and EFG.GDP leaves the scene as it is no more capable of remaining bound to ribosome. It becomes ready again when it is recharged with GTP. Thus both the factors (EFG and EFT) act catalytically. The elongation process is also repeated in the same way for each and every incoming amino acid as directed by the codons on the mRNA (Fig.4.5)

EFTu and EFG cannot bind simultaneously to the ribosome as they share overlapping sites (Fig.4.7). Therefore, unless one factor

leaves, the other one cannot bind. This is important as the two have to work one after another in a cyclic fashion. Another interesting fact is that EFTu (MW. ~40,000 Da) has similar shape and size as that of a part of EFG, which is of higher molecular weight (MW. ~70,000Da). However, when EFTu associates with tRNA it takes the shape and size of EFG as the additional part of the latter looks similar to tRNA. This is a kind of molecular mimicry between protein and RNA. Following the binding of aminoacyl tRNA to ribosome, EFTu leaves. Apparently, the part of EFG similar to EFTu in shape and size takes its position when it moves from the A site to the P site. It is quite likely that EFG in this way is responsible for the translocation of tRNA. Subsequently, the process repeats itself and the elongation of protein chain continues. So the incoming transfer RNA at the A site receives the peptide chain already synthesised and along with it moves

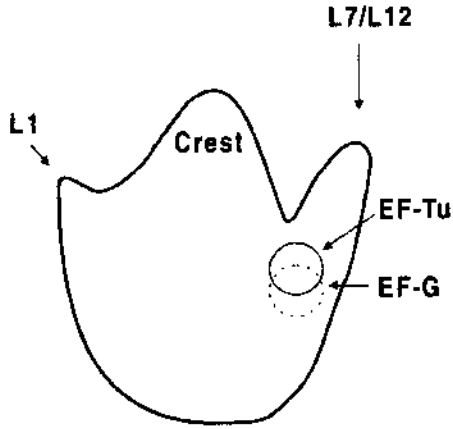


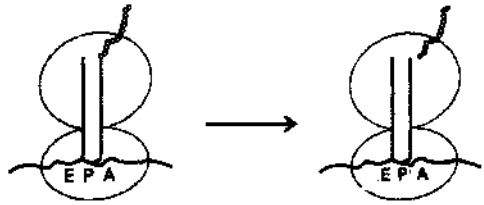

Fig. 4.7. Overlapping sites of binding of elongation factors EF-TU and EF-G, complexed with GTP. Both the factors bind at the bottom of the L7/L12. But cannot bind simultaneously. As soon as peptidyl tRNA moves from A site to P site EFG binds at or near A site and is perhaps directly responsible for the translocation of peptidyl tRNA. Details are described in the text.

from A site to elongations of P site till the elongation of the protein chain is completed.

The site of binding of both EFTu and EFG is at the base of the L7/L12 stalk region of 50S ribosome as shown in Fig. 4.7. It is believed that the stalk could be involved in the translocation process. For a long time it was assumed that the protein factors for initiation, elongation etc., interact directly with the ribosomal proteins. However, in case of EFG it was shown for the first time that it binds directly to 23S RNA, the major RNA component of 50S ribosome. As soon as GTP is hydrolysed, EFG.GDP cannot remain bound to 23S RNA due to change of conformation and leaves the ribosomal site. EFG.GTP has the proper conformation to bind to 23S RNA, which EFG.GDP does not have. After leaving the ribosome EFG dissociates from GDP and associates with another molecule of GTP, ready to go back to the ribosomal site of binding and perform another step of translocation. This cyclic operation goes on continuously. EFG acts catalytically, one molecule catalyzing several steps of translocation. As discussed already, EFTu does the same, it brings an aminocyal tRNA to the site of protein synthesis, leaves the ribosome and goes back to the cytoplasm to pick up the next one. EFTu and EFG bind alternately to ribosomes, and can never bind together. Two should function one after another and not simultaneously. Further, large amounts of these factors are not necessary as they act catalytically. The two factors have evolved in such a way that they can compete with each other for the same site on the ribosomes. The molecular mimicry between an RNA and part of a protein molecule may throw light on the mechanism of translocation.

The last stage in ribosome-mediated protein synthesis is the termination step (Fig. 4.8). If there is no stop signal at the proper site, the elongation of protein chain will continue beyond limit. Actually, three stop codons (UUA, UAG and UGA) are used by nature (see Table.3.1a). Out of a total number of 64 codons these three codons do

not code for any amino acid. The stop signal is not enough and one of the two termination factors generally called release factors (RF1 and RF2) have to interact with the ribosomes to make the termination effective. As shown in the Fig. 4.8, RFI is necessary for UUA and UAG to be functional whereas RF2 is necessary for UUA and UGA to act as termination codons. So UUA has the option to use either RF1 or RF2. Nature is clever enough to sometimes put two stop codons in tandem to prevent the unlikely incident of accidental elongation of the chain beyond the desired limit.




| Termination Codons | Release Factors |
|---|---|
| UAA | RF1 / RF2 |
| UAG | RF1 - |
| UGA | - RF2 |

Fig. 4.8. Termination of the protein chain. There are three termination codons (UAA, UGA, UAG) which act as stop signals. Sometimes two stop signals are consecutively used to ensure termination. The termination step needs help of the termination factors, known as release factors. Each termination codon uses specific release factor(s) as shown above.

One other complication in protein synthesis should be clarified at this stage. It was mentioned earlier that protein synthesis starts with the formation of an initiation complex with 30S ribosomes alone. However, in practice the same ribosome after completing one round of protein synthesis can initiate another round on the same messenger

RNA. For this an additional release factor known as ribosome recycling factor (RRF) is necessary. It acts in collaboration with EFG. The crystal structure of RRF has been determined. Although it is a protein, its three dimensional configuration is similar to that of tRNA. This is another example of 'molecular mimicry'

To recapitulate whatever has been discussed already, the reading frame of mRNA starts with the initiation codon AUG which is complementary to the anticodon at the base of the initiator tRNA or formylmethionyl tRNA ($tRNA_f$). In prokaryotic systems all proteins start with the amino acid formylmethionine. Formyl methionyl group acts more or less like the capital letter at the start of a sentence. The transfer RNA charged with formylmethionmine enters the ribosomes at the peptidyl (P) site at the initiation stage. As discussed already, the message is recorded in triplet, three bases make a code which codes for an amino acid, rather tRNA carrying the particular amino acid. As already discussed in Chapter 2, tRNA is L-shaped. The L is put in a reverse fashion at the ribosomal site; the bottom contains the anticodon to recognize the codon while the top part holds the amino acid in position. The amino acid is not an ordinary one; it is pre-activated with the energy-rich compound ATP (amino acyl adenylate). Two such tRNAs can somehow be fitted side by side at the site of protein synthesis. The calculation shows that the space is just enough to place them at an angle to each other. However, there should be a kink in the message to fit in the two. The two activated amino acids have to be brought sufficiently close to be linked by a peptide bond. It was assumed that an enzyme called peptidyl transferase catalyses the peptide bond formation. No such enzyme has so far been discovered. Earlier it was thought that a number of ribosomal proteins make the peptidyl transferase centre. Lately it is conjectured that ribosomal RNA creates the peptidyl transferase centre and is responsible for peptide bond formation. The peptide bond is formed rather automatically, if not, spontaneously, although it is still debatable.

What is necessary is the proper aligning of the two tRNAs, atleast transiently. During this transient state the peptidyl moiety is transferred to the amino group of other aminoacyl moiety attached through its carboxy group to the tRNA in the aminoacyl (A) site. Therefore, it may be conjectured that ribosome as a whole acts as the peptidyl transferase, which is responsible for the peptide bond formation. This would be discussed a little later on the basis of a very recent study of the crystal structures of the ribosomes.

It is worth recording here that the naked 16S and 23S RNAs (isolated from 30S and 50S ribosomes respectively) free from any protein can form the binary complex (16S.23S RNA) as demonstrated for the first time in our laboratory and later in other laboratories. Further, the complex is capable of carrying out the steps of protein synthesis with the help of two or three ribosomal proteins and added factors although to a very small extent or, in other words, it just mimics the ribosome-like activity. Therefore, ribosomal proteins may not be directly involved in protein synthesis. A few years later an Internationally known American ribosomologist (Harry Noller) showed that ribosomes from a heat-loving microorganism (*Tetrahymena thermophile)* could retain the full peptidyl transferase activity even when 90-95% proteins are removed. This made every one believe that ribosomal RNAs having the peptidyl transferase activity can act as ribosomes. It should be mentioned here that the periodic meetings on ribosomes that are held regularly are the records of the advancement of ribosomology[2,3,4].

Recently a breakthrough in 'ribosomology' has created a big hope that the intricacies of protein synthesis could be fully revealed through the X-ray crystal structures of ribosomes determined by three different laboratories. It should be pointed out here that the mechanism of peptide bond formation which remained unsolved for the last three decades may also be understood fully in the near future. All along it was suspected that 23S RNA (specially its domain V) could act as a

ribozyme. Noller already visualized a ribozyme type of structure at the interface of the two ribsomal subunits. Very recently, Peter Moore and Thomas Steitz have proposed on the basis of their structural studies that the mechanism may be just the reverse of proteolysis (or in other words, hydrolysis of a peptide bond by a protease). There is not much scope of discussing the proposed mechanism here but it suffices to point out that a tetrahedral intermediate (carbon attached to four different groups) is formed by the nucleophilic attack of amino ($NH_2$) group of one amino acid on the carboxyl (COOH) group of another. A single adenine base of 23S RNA acts as acid base catalyst for this purpose. This is, however, not universally accepted. It may not be out of place to mention here that in a collaborative model building work we have suggested that the 2′ hydroxyl group of peptidyl transfer RNA is involved in the peptide bond formation.

Whatever might be the mechanism of the peptide bond formation there still remains the question of movement of mRNA along with the tRNAs through the ribosome or *vice versa*, one codon at a time. It can be visualised that all the three codons associated with the three tRNAs according to the three site model are located at or near the so-called peptidyl transferase (PT) centre on the ribosomes. The cavity that is created between the two ribosomal subunits houses the two (or three) tRNAs at the so called A, P and E sites. The mRNA has to wriggle through this site as a thread does through the eye of a needle; but what device threads it through? As already mentioned, at every step of addition of an amino acid, tRNA at the aminoacyl site has to move to the peptidyl site as soon as a new peptide bond is formed, to make space for the incoming tRNA with the next amino acid. During translocation, tRNA (stripped of the amino acid) located at the Exit (E) site has also to leave. The picture is clear but how does the translocation take place?

Before we consider the details it should be clarified that any tRNA (with the corresponding amino acid) cannot land at the aminoacyl

to a specific site and there is only one copy each of the proteins in the 30S ribosome. It is also interesting to note that all the ribosomal proteins are available to the external reagents even molecules as large as antibodies. Therefore, ribosomes have a somewhat open structure. As discussed in Chapter 2, the models of ribosomes (RNA and protein together) have been built with the combined efforts of several laboratories which used two powerful tools, immunoelectron microscopy and neutron diffraction. Recently a new technique called 'cryoelectron microscopy' which provides a better picture of the ribosomes has been introduced although its resolving power is somewhat lower than that of an ordinary electron microscope. However, under frozen conditions (in liquid nitrogen) the features of ribosomes are not distorted, and accurate pictures with details can be obtained. Eventually their crystal structures have also been obtained.

The outlines of the three-dimensional structure of *E. coli* 50S ribosome have also been shown in Fig. 2.14. This is a consensus model, so to say, and may not be absolutely correct in minute details. 50S ribosomes are almost double to the size of 30S ribosomes and have other interesting features. It has a head (sometimes called crest) similar to but not exactly like that of 30S ribosomes, a protuberance called L7/L12 stalk on one side which extend into the medium (such stalk does not exist in 30S ribosomes) and a small protuberance on the other side. There is no platform region like that of 30S ribosomes. It is believed that when the two subunits join together the interfacial space between the two, near the platform (of 30S ribosomes) acts as the site for protein synthesis. The head region of 30S ribosomes appears to have a rocking motion during protein synthesis and this may be due to changes in the conformation of 16S RNA. A better understanding of L7/L12 stalk, which is one of the most interesting features is also necessary to understand the mechanism of protein synthesis completely.

Although translocation factors (EFTu and EFG) are known for

quite sometime their mechanism of action is still not absolutely clear. Several laboratories are engaged in unraveling the mechanism of translocation and a number of models have been proposed from time to time from different laboratories. These models are, as expected, basically different from one another but all of them advocate the movement of the messenger RNA (along with tRNAs) coupled to the peptide bond formation either simultaneously or immediately thereafter. From historical perspectives it is interesting to recall some of those. For example, Lipmann suggested a pulsation (alternate contraction and expansion) model. Spirin argued for locking-unlocking model. According to him ribosome is like a box with a lid and the two subunits are linked by a hinge just like a box. The two remain associated at one site and open up and close just like the box in a pulsating type of movement. He assumes that this movement is mechanistically responsible for the translocation. Later Spirin proposed a more sophisticated model where two tRNAs are aligned at an angle to each other and become transiently parallel resulting in peptide bond formation and simultaneous movement of the mRNA, one codon (three bases) at a time. It is interesting to note that Hardesty originally proposed an inchworm theory of the movement of mRNA. It has already been mentioned that there should be a kink in the mRNA to accommodate the two tRNAs side by side. The latest model proposed by Hardesty (displacement reaction model) is very much similar to Spirin's revised model. Another interesting proposition made by Woese is based on the ratchet mechanism. A directional movement like that of a ratchet (clockwise or anti-clockwise) results in the movement of the mRNA. Does L7/L12 act as a ratchet? All the models are based mostly on theoretical considerations and hardly any experimental evidence has been provided for any one of those.

Our model of translocation based on the conformational changes would be described here. Under *in vitro* condition a simple cation like $Mg^{++}$ (magnesium ion) can also control association and

dissociation of 30S and 50S ribosomes as IF-3 does under *in vivo* condition. At low $Mg^{++}$ concentration 70S ribosomes dissociate and they associate at high $Mg^{++}$ concentrations. Under physiological condition this cannot happen as in a cell. $Mg^{++}$ concentration remains fairly constant. However, many laboratories used this $Mg^{++}$-dependent association-dissociation for routine studies of translocation. Our approach was also similar, although we used IF-3 in some of our studies to verify whether Mg++ is mimicking the role of IF-3 in regulating the association and dissociation of the ribosomes. We argued that there are two populations of ribosomes in *E. coli*. As the cells grow at a fast rate the number of ribosomes increase. It may increase from a few hundreds at the stationary phase when the bacteria stop growing to a few thousands at the exponential phase of growth. Most of the ribosomes in the rapid phase of growth are in a form called 'tight couple'. Tight couple ribosomes can be dissociated at a very low $Mg^{++}$ concentration. They do associate at 4mM $Mg^{++}$. Ribosomes present in the stationary phase of growth are, mostly in 'loose couple' form. These dissociate rather readily at 4mM $Mg^{++}$ and require a much higher $Mg^{++}$ concentration (10 mM or so) to associate under *in vitro* conditions. These loose couple ribosomes have less biological activity than tight couple ones and are therefore thought to be damaged. However, we do not agree to this for reasons which are mentioned below.

Our model of translocation (Fig. 4.9) is based on the assumption that tight and loose couple ribosomes are the intermediate products of the translocation cycle[3]. Tight couples are in the pretranslocation stage. On addition of EFG.GTP to tight couple 50S ribosomes (step1) it binds to 23S RNA at the base of the L7/L12 stalk and two changes take place. The conformation of RNA at the base changes and L7/L12, which was folded towards the body, becomes extended into the medium (this is reverse of the belief that tight couples have normally the stalk extended into the medium). Thus tight couple 50S ribosomes

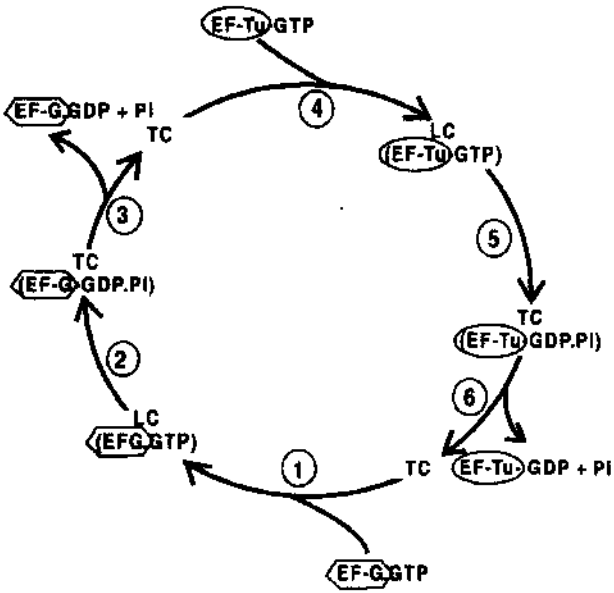



Fig. 4.9. Translocation cycle involving tight and loose couple ribosomes. There are two populations of 70S ribosomes in which 30S and 50S ribosomes are tightly and loosely bound due to the conformational change of 50S ribosomes during translocation (the model was proposed from this laboratory and confirmed by cryoelectron microscopy in Frank's Laboratory).

become converted to loose couples. Subsequently, GTP hydrolysis takes place on the ribosome itself (step2) and loose couple ribosomes are converted back to tight couple form and EFG.GDP dissociates from the ribosome (step3) leaving it in the GDP state (of tight couple). It is converted to loose couple (GTP conformation) again if EFTu.GTP binds to it (step 4). As soon as GTP is hydrolysed on the ribosome to GDP and Pi, loose couple is converted to tight couple (step5). Following the release of EFTu.GDP and Pi tight couple becomes free (step6) and is ready to be associated with EFG.GTP. Thus the cyclic process of interconversion (as shown in Fig. 4.9) is repeated at every step. EFG is known to undergo conformational changes when bound

with GTP. EFG.GTP has different conformation than EFG.GDP. Due to this difference EFG.GTP strongly binds to tight couple 50S ribosomes. That is why the loose couple ribosomes have much less biological activity than tight couple ones. There is one more change during the interconversion of tight and loose couple ribosomes. It has been shown in our laboratory that there are two sites of association between 30S and 50S ribosomes. One site is common while the other is not. Therefore, there is a spatial separation (movement with respect to each other) between the two subunits during translocation, which has been supported by experimental evidences from other laboratories, specially by cryoelectron microscopy.

It is known that on the addition of phenylalanyl tRNA.Tu.GTP (ternary complex) to the vacant ribosomes the binding of phenylalanyl tRNA originally takes place at the A site and then switches over to the vacant P site. However, we further observed that the tight couple ribosomes are converted to loose couple ones as well during this process. If we block the P site with phenylalanyl tRNA then the entry of another tRNA at the A site does not add upon any further conformational change to 50S ribosomes. Binding of EFG.GTP or EFTu.GTP leads to similar conformational change. It should be remembered that EFTu and EFG occupy overlapping sites on ribosomes and therefore, must have a similar function as far as conformational change of 50S ribosomes are concerned. However, the two cannot bind simultaneously. A German group working on pressure-induced dissociation of ribosomes supported our model strongly and has explained their results with the help of our model. So our model based on conformational change of 50S ribosomes may stand the test of time.

It might not be out of place to mention here that my last Ph.D. student Rajendra Agarwal conducted an extensive work using the technique of cryoelectron microscopy in the laboratory of Joachim Frank. They showed direct evidences for the first time in favour of

conformational changes of ribsosomes during translocation. On the basis of their detailed observations they suggested a ratchet type of movement (as mentioned earlier) of the two subunits with respect to each other. According to their model the two subunits move with respect to each other during translocation leading to an intermediate state known as hybrid state, as suggested by Noller earlier(Fig 4,10). Peptidyl tRNA and aminoacyl tRNA occupy P and A sites respectively on both 30S and 50S ribosomes. Following peptide bond formation and during translocation the hybrid state A/P and P/E is produced that is aminoacyl tRNA occupies A site on 30S ribosome and P site on 50S ribosome whereas peptidyl tRNA occupies P site on 30S ribosome and E site on 50S ribosome. In the next step the normal state is created, as expected. It is becoming more and more obvious that the conformational change of 50S ribosomes induced by the translocation factor EFG is the key behind translocation.

A phenomenal development took place in ribosomology when three different laboratories determined the crystal structures of ribosomes. X-ray crystallography is usually the final step in the determination of a structure of small and large molecules. It was unthinkable in those days to determine the X-ray structure of the giant molecules like

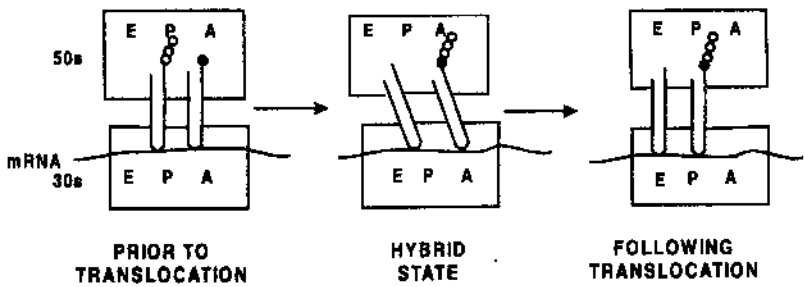



Fig.4.10. Hybrid structure of 70S ribosomes during translocation (proposed by Noller). The details have been described in the text.

ribosomes. Yet, work in this area was initiated in the laboratory of Ada Yonath in Israel quite sometime ago (in 1980s) but the progress was rather slow. Suddenly three more laboratories jumped into the arena and beautiful X-ray diffraction pictures of 30S, 50S and even 70S ribosomes were obtained. Those yielded a lot of information about the architecture of the protein-synthesizing machinery. Prior to that two other laboratories were working with another technique known as cryoelectron microscopy, mentioned early, which also became quite helpful in this respect. One of the major difficulties in electron microscopy is the distortion of the structures to be visualized during the preparation of sample on the grid by drying in vacuum. To avoid this problem the technique of cryoelectron microscopy was introduced. In this case the grid with the sample in state of solution is dropped into liquid nitrogen. The sample instantaneously freezes and the water becomes glassy ice. The features of the preparation in which the samples are not distorted can be visualized through the electron microscope. Joachim Frank of Albany perfected this technique for studying the structure of ribosomes. Numerous pictures of the ribosomes were taken by tilting the grids at various angles and the overall three-dimensional structure was built. Frank, a physicist, developed powerful computer programmes to analyse those. The disadvantage with such techniques is the low resolution, which was initially in the order of 10Å. This has been gradually improved and Frank has not only visualized the three-dimensional structures of both the subunits individually but also in a state of association and has determined the positions of the three transfer RNAs at three distinct sites (A, P and E) in the cavity between the two subunits. This has been supported by X-ray diffraction studies as well which will be discussed shortly. Frank's group in U.S.A. and the other group headed by Martin Van Heel at Berlin, Germany are in fierce competition with each other. Actually the two groups have similar pictures of ribosomes but are not in full agreement regarding the positioning of tRNAs on the ribosome. However, the beauty of this technique is that the

ribosomes in any state during protein synthesis can be instantaneously frozen; therefore, the changes involved during each and every step can be analysed in detail. We shall discuss a little later about how this technique has helped some X-ray crystallographers to analyze the details of the crystal structures. As the resolution power is being increased cryoelectron microscopy is turning to be a more powerful technique.

X-ray crystallography of ribosomes reminds me of a meeting on 'Ribosomes' held in 1984 at Marine Science Institute, University of Texas, at Port Aransas. At the end of the sessions Peter Moore was asked to review the proceedings and he mentioned in a rather oblique way that Ada Yonath is engaged in the studies of ribosome structure by X-ray crystallography for the past few years but she has not been able to achieve much. Yonath became furious and tried to defend herself by saying that X-ray diffraction pattern of such a giant molecule composed of large number of proteins and two long and one short chain RNAs is so much complicated to be analysed quickly. In spite of that she had some breakthroughs already, one of them being the detection of a channel in *Bacillus* 50S ribosomes, starting from the protein synthesising cavity and opening into its backside which was confirmed by others much later. Yonath started her crystallography work in Weizmann Institute of Science at Rehovot (Israel) but since the facilities of her laboratory were not that good she started collaboration with H.G. Wittmann at the Max Planck Institute of Molecular Genetics, Berlin. After Wittmann's death she worked in the Max Planck Institute at Hamburg, Germany to continue her crystal studies tenaciously. In subsequent meetings she presented more and more information due to various improvements in the techniques but the progress was rather slow. In the meantime three independent groups had crystals of 30S, 50S and 70S ribosomes and presented their work in Ribosome meeting held at Denmark (Helsingor Conference). Finally, three groups almost simultaneously published three papers,

two in *Nature* and one in *Science*. It should be pointed out that three different groups crystallized ribosomes from two different thermophilic microorganisms, none of them used *E. coli*. However, the structures are not much different from those of *E. coli*. The sequences of RNAs and proteins and the three-dimensional organization of the ribosomes are highly conserved with only minor differences, which was of great advantage.

One of the three groups is headed by Venkit Ramakrishnan who performed the neutron diffraction work in Peter Moore's laboratory. I remember meeting him for the first time during the International Biophysics congress held at Mexico, where he presented a poster on his neutron diffraction work done in Moore's laboratory. It was a pleasant surprise to learn that his biochemist father in India was known to me. The X-ray diffraction work was carried out with crystals of 30S ribosomes of *Thermus thermophilus* at 5Å resolution at the University of Utah, Salt Lake City. He followed the footsteps of Ada Yonath and strongly emphasised that his work was possible due to the lead she already had with 50S ribosomes but his pictures were far more important than her. The picture of 30S ribosome that emerged from his studies clearly showed the phosphate backbone of 16S RNA (Plate 7, Fig. 4.11). The α–helices of proteins are also distinctly visible. He could also fit in the structures of several proteins, which were individually crystallized earlier. Double helical regions of RNA (mostly in A form, some in non-A form as well) are unequivocally identified. The fold of the central domain of 16S RNA is also distinctly visible. As a whole the path of 16S RNA chain in the subunit could be traced for the first time. The platform part of 30S ribosomes parallel to the inter subunit space appeared to consist of two helices, one RNA-rich with practically no protein. This supported the concept of the mobility of the platform. All the cross linking and biochemical data involving RNA and protein carried out earlier by Noller's group and others were in good agreement with the crystallogrphic data. The painstaking work carried out over decades was thus fully vindicated

by X-ray crystallography.

Peter Moore who criticised Yonath for the slow progress decided to join hands with the famous X-ray crystallographer Thomas Steitz at the Yale University where both worked. Steitz had already crystallized a number of important biomolecules including DNA polymerase and threw considerable light on their functions. The collaboration resulted in the beautiful crystallographic pictures of 50S ribosomes of *Haloarcula marismortui*, a salt loving bacterium at 2.4Å resolution (Plate 8, Fig. 4.12). One of the difficult jobs in X-ray crystallogrphic work is to determine the phases of the diffraction pattern but they derived that from 4 heavy atom derivatives, inter crystal density averaging and density modification. More than 300 base pairs in A form were fitted into the map. In this case as well phosphate backbone of RNA (both 5S and 23S RNAs) was clearly visible. In the case of 30S ribosomes RNA is in non-A form duplex. Single-stranded regions and the tetra loops which are functionally important could be identified without any ambiguity. It was further observed that long rods of RNA crisscrossing the subunit arose from the stacking of short separate double helices. Again not all of them are in A form. In many places proteins crosslink two or more of these rods. Polypeptide channel observed originally by Yonath was marked by four tungsten atoms. The exit channel is so called because it is the exit route of the polypeptide chain synthesised on the ribosome. Translation factor binding sites were identified by fitting in the crystal structures of L6, L11 and L14 and α-sarcin-ricin loop of RNA (involved in protein synthesis) that binds L11 protein. Elongation factors G and Tu (complexed with aatRNA and GTP) have also been localized in the same region. The observations agree with the earlier electron microscopic data as well as elaborate biochemical studies carried out by various groups. It is interesting to recall that I met the couple Joan Steitz (well known for her work on small nuclear ribonucleoprotein particle i.e. SnRP) and Thomas Steitz (an internationally known X-ray crystallographer) at Delhi in a symposium organized by the National Institute of Immunology. Both are

highly respected in their individual fields of research.

In order to get the information about the structural domains of ribosomes Noller, the 'Master of Ribosomology' was the one who first carried out extensive experiments based on chemical crosslinking, foot printing etc. Subsequently, he joined the race on evaluation of ribosome architecture by X-ray crystallography. The story is slightly different from other competitors. A Russian couple Yusupova, who got married while working in Alexander Spirin's laboratory in Pouschino near Moscow got good crystals of ribosomes but had little facilities to work out their structures. So they migrated to France to do the job. There too they could not make much progress, so it was no wonder that they readily joined Noller's laboratory on his invitation. Noller also recruited one young American crystallographer Jarmie Cate and utilized the X-ray diffraction facilities of Thomas Earnest at Berkeley Radiation Laboratory. Their joint work using crystals of 70S ribosomes of *Thermus thermophilus* at 7.8Å resolution was published in Science almost simultaneously with those of the other two groups in *Nature*.

The information obtained from cryoelectron microscopic studies of Frank, was used extensively by Noller at every stage. Actually, he combined the two tools and strongly believes that both are necessary to resolve the structure of functional complexes. As usual phosphate backbone could be traced in both the subunits in association (Plate 9, Fig. 4.13). A RNA-rich region was found at the interface of the two subunits. Penultimate helix of 16S RNA was also located at the interface. The functional triangular space between the two subunits has been visualised and the locations of the three tRNAs in that space have been determined without any ambiguity. Details of tRNA interactions with ribosomes have been worked out and those agree, in general, with the biochemical studies specially carried out in his laboratory. Numerous contacts are found between the 30S subunit

and peptidyl tRNA. In contrast the anticodon region of aminoacyl tRNA is much more exposed. Besides those, quite a few RNA-RNA and some RNA-protein bridges have also been localised. Finally, X-ray diffraction studies have directly demonstrated the binding of three tRNAs in the cavity between the two subunits (Plate 10, Fig.4.14). Noller suggests that the relay between the different segments of 23S RNA would eventually help to pin down the mechanism of peptide bond formation which has still remained elusive. He has also identified a ribozyme type of structure at the interface of both the subunits. It may be recalled that the author demonstrated way back in 1985 that 16S: 23S RNA complex with a limited number of proteins is capable of peptide bond formation. Seven years later, Noller demonstrated that ribosomes treated with trypsin retain full biological activity even when more than 90% protein is removed. As already mentioned, the author in a collaborative work has suggested that the 2' OH group of peptidyl tRNA may be involved in the formation of the peptide bond through a six member cyclic intermediate. Peter Moore is, however, not in agreement with this proposition but he himself has found that the OH group is located in the functional region. Noller strongly feels the need of a type of movie camera based on cryoelectron microscopy and X-ray diffraction to take pictures of functional complexes at various stages to elucidate the mechanism of the peptide bond formation. It appears that the two techniques applied in conjunction will finally unravel the mysteries of structure-function relationship of ribosomes, which has so far remained obscure. The latest work on X-ray crystallography has been discussed by the author in a popular article.[5]

At the start of this chapter ribosome was compared to a cassette player but the analogy ends there. Unlike a cassette player ribosome is assembled from two parts, a large and small subunit. No one knows the evolution of this apparatus. It is sometimes tempting to speculate that primitive ribosomes might have been composed of RNA only,

devoid of any protein and a single RNA chain performing the function of both the subunits. Ribosomal RNA, tRNA, messenger RNA, rather all kinds of RNA are derived from DNA. So, the interactions between these RNAs could have led to the evolution of primitive translational apparatus. This issue is being raised at the end just to emphasize that the primitive cassette player and its tape might have been a much simpler device composed of RNA only. Evolution through millions of years makes them look different today but they still are functioning in old fashion. Due to their common origin, the mechanism of translation of the genetic message appears to be the same universally and the translational apparatus has naturally become ubiquitous.

At early stages when various models of the two subunits of *E. coli* ribosomes were being proposed, one of these (proposed by the German group) was popularly called "Teddy bear seats on the arm chair". Obviously 'teddy bear' refers to 30S ribosomes and 50S ribosomes represent the arm chair. At that time I was just getting acquinted with ribosomes. Bernie Horecker (one of my earlier mentors) was visiting Calcutta to attend a symposium. He was chairing the session where I was presenting my work. This was at Bose Institute, my old haunt. To amuse him I made a special slide representing 30S ribosome as a female and 50S ribosome as a male. Naturally, the female was sitting on the lap of the male in the teddy bear style. Bernie was no doubt amused but declared in a serious tone, `I did not know that Debi is now doing pornography only'. I retorted, `I was induced to it by my mentor'.

In this connection the following extract taken from one of my popular science articles may not be completely out of place.

"The word 'two' is a magic word, rather, the number, two is a magic number and a keystone or a touchstone whatever you may call it. The number is not only intimately connected with the living world, but also lies at the bottom of the evolution. A man and a woman, a male and a female, a he and a she. When the great deluge came Noah

took care to have a pair of each animal in his ark to save the variety created by God who also created Eve out of the rib of Adam to provide a pleasant company to him. Later she became a source of trouble ....."

As a man of science I should not be carried away by the magic number. In spite of my poor knowledge in biology and even before meeting my biologist wife I learnt that bacteria grew in number by asexual reproduction, which was news to a physical chemist! A single bacterial cell grows larger when it feels (or rather needs) and divides into two; the microbiologists call the newly created ones as 'daughter' cells. Again I do not know why they prefer 'daughter' to 'sons' but still the number two comes into the picture. May be, I am somehow biased towards the number. But I was told that bacteria enjoy (who can tell!) or rather can have sexual conjugation. Joshua Lederberg along with his *Guru* Edward L. Tatum who shared the Nobel Prize in medicine in 1945 demonstrated that for the first time by a simple trick and coined the term 'sexual conjugation'. One bacterium is capable of donating its genetic material DNA to another bacterium. In retrospect one may remember the famous PaJaMo experiment (described in Chapter 3) was based on the transfer of genetic material from a male bacterium to a female by sexual conjugation. The donor is branded 'male' and the acceptor 'female'. Lederberg and Tatum definitely liked the involvement of the number two but an amazing thing happens in this case. The female is converted to a male by 'sexual conjugation'. I do not know whether 'women's lib' group will be pleased at this performance but this is reality.

## References

1. P.C. Zamecnik in Cold Spring Harbor Symposium on Quantitative Biology vol.34,page 1 (1969)
2. *A history of ribosome research : A personal account.* M. Nomura in '*The*

*ribosome structure* and evolution' edited by W.E. Hill, R.A. Garrett., P.B. Moore, D. Schlessinger and J. Werner, American society for Microbiology, Washington D.C. (1990).

3. *Conformational changes of 50S ribosomes during protein synthesis* by D.P. Burma et al., in 'Structure, function and genetics of ribosomes' edited by B. Hardesty and G. Kramer, Springer Verlag, New York, (1986).
4. *Frontiers on translation*, edited by A.T. Matheson, P.P. Dennis, J.E. Davis and W.E. Hill, Biochemistry and Cell Physiology, vol. 73, No.11 & 12, (1995).
5. *Architechture of the protein synthesizing machinery of the cell.* by D.P. Burma, *Science and Culture*, vol. 67, page 3 (2001).

*****

# Chapter - 5
## Who is the original player?
### (DNA World versus RNA World)

*The story of the music of life was started (Chapter 1) depicting DNA as the master musician creating the music of life. Comparatively in recent time it has been realised that RNA sometimes acts as the master at least in some viruses, which have RNA as the genetic material. It might have originally been the genetic material before DNA evolved. A majority of the molecular biologists today believe that the RNA world preceded the DNA world, as originally suggested by the Nobel Prize-winning physicist Walter Gilbert. The evidences in favour of it would be elaborated here, one of the strongest evidences being that DNA in the DNA world cannot be synthesised without the help of RNA and its ingredients. It is tempting to speculate that RNA, which was once the master of life had chosen to be the servant to the present day master whom it itself created in its own image for quite valid reasons.*

Before I begin let me recollect one of the pleasant memories of my life. Indian Science Congress Association holds its annual session at different places in India by rotation. Quite a few foreign scientists from different countries are invited on this occasion. In 1968, the fifty fifth session was scheduled at Banaras Hindu university at Varanasi where I worked. I was amazed to learn that the illustrious Professor Alexander I. Oparin of the then USSR, who was invited to the Congress, had expressed his desire to visit our laboratory (Molecular Biology Unit) at the Institute of Medical Sciences. Oparin was a world-renowned expert in the area of 'Origin of life'[1] and speculated that the atmosphere of the primordial earth was rich in

methane, ammonia, water and essentially devoid of oxygen. According to Oparin, electrical energy of lightning discharge and heat energy of volcanoes led to the formation of such organic compounds, which dissolved in primordial soup, the ocean. Over millions of years these simple molecules led to the formation of complex molecules, which assembled spontaneously and formed membranes, catalysts (enzymes) and various other constituents of the cell. These in turn led to the formation of primitive cells. In course of time the cells developed the capacity to replicate themselves. I could not trust my eyes when I read the notification from the organizers about his visit. At that time I had initiated my work with ribosomes, which was far from the exciting area of the origin of life. He arrived in due course and spent quite a bit of time visiting the laboratories and discussing a lot about ribosomes. I realised then that he knew A.I. Spirin of the Institute of Protein Research at Pushchino in the then U.S.S.R. and was quite familiar with his work on ribosomes, which might have led to his visit. Spirin himself visited my laboratory later. However, the final words that Oparin said about ribosomes[2] were rewarding. 'Understanding about the evolution of ribosomes will help a lot to unravel the mystery of the origin of life'. This would be a part of the discussion in this chapter.

We are living in a DNA world. All living beings we see around us have DNA as the genetic material. As we have learnt by now, the genetic material contains all the information about the living system, which is written through the four letters of DNA alphabet. All living systems starting from a tiny microorganism that could be seen only under a microscope and ending in large size animals like whale, elephant etc., all contain DNA as the information-carrying material. As far as we know, all the extinct organisms (including the dinosaurs, trinosaurs etc.,) contained DNA as the genetic material. It makes us believe that life started on the earth with DNA as the genetic material. This was strengthened further as the genetic code was found to be

universal. Hardly any exception is found in the list of 64 three-letter codons although there may be some minor variations, for example, in mitochondrial DNA. The usage of these codes does vary from organism to organism but no new code beyond the three letter codes has been discovered till date. Therefore, reasons are strong enough to make us believe that the DNA is globally the genetic material and the triplet codes contained in them are also global. So there should not be any hesitation to admit that we ourselves and the living world surrounding us of which we are a part, constitute the DNA world. The issue, however, may justifiably be raised that some viruses (bacterial, plant and animal) that have no DNA but RNA instead are exceptions. We know for sure that in their cases RNA acts as the genetic material and the information is written in the form of RNA although some of them may form DNA as the intermediate in their life cycle. This may be true but it is doubtful whether viruses should be identified with the proper living entities. Viruses multiply only in living organisms. They do not have the full machinery to reproduce and function independently. All of them are invariably dependent (though not fully) on the genetic information of the host, may be a microbe, plant or animal. From their life style it immediately strikes one that they are familiar with the machinery of their hosts, the pivot of which is DNA and utilize that information fully. Truly speaking, they are also part of the DNA world and not outside of it. We may, however, justifiably concede that RNA may act as genetic material in these exceptional cases. Then why do we talk of the RNA world? Walter Gilbert who got the Nobel Prize in 1980 for developing one of the methods of sequencing DNA coined the term' RNA World' in 1986. The question is 'Why?'

As we have already learnt, unlike DNA, RNA is multifunctional. The source of music of life is no doubt DNA but RNA plays the actual music, as it has been now realised. At various stages of expression of the living system RNA is involved. It is sometimes said that the DNA

is the master molecule, RNA is its most obedient servant carrying out the orders of the master. At the early stages the informational content of DNA, if any, would have remained useless if its message was not transcribed in the form of RNA, as discussed in Chapter 3. We talk of junk DNA because the message of some part of DNA is not transcribed. Once the message is transcribed its next important duty is to synthesise proteins for which other types of RNAs are also required, for example, transfer RNAs which act as handles for amino acids to bring those out of the cytoplasmic pool and carry them to the ribosomal site for protein synthesis (Chapter 2). Not only that the RNAs are capable of translating the messages in form of messenger RNA it also acts as part of the machine for protein synthesis (Chapter 3). The translation is one of the most important functions in the living system. The next important role is that of ribosomal RNAs which are also transcribed from DNA. Ribosomal RNAs (along with the ribosomal proteins) form the integral parts of ribosomes, which are the sites for protein synthesis. To recapitulate, ribosomal RNAs are of various types, 5S,16S and 23S (in prokaryotes), 5.6S, 7.5S, 18S and 28S (in eukaryotes). These not only interact with proteins but also give rise to the three dimensional structure of ribosomes which is extremely important for their functions. There are also other types of RNAs doing some specialised jobs.

Another important issue, which should be discussed at the very first step is that in the DNA world no DNA can be replicated without RNA[3]. The first step, which we have already mentioned in Chapter 2 is the need of an RNA primer which is the starter for the synthesis of DNA chain. Okazaki fragments, the intermediates in DNA synthesis are synthesized with RNA primers (Fig 5.1 a). The primer is formed by copying the DNA by an enzyme called primase. This RNA primer chain is extended as DNA chain. Later, the RNA primer is removed and replaced by DNA. Therefore, without RNA synthesis DNA synthesis cannot start. In other respect also DNA synthesis cannot

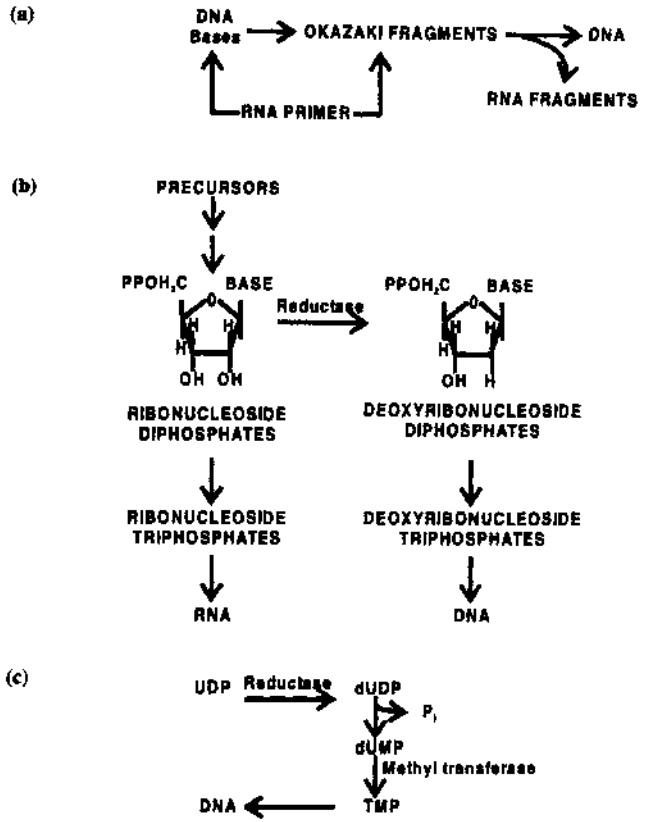



Fig. 5.1. Dependence of DNA duplication on RNA and RNA components. (a)RNA acts as a primer for DNA synthesis. A short ologoribonucleotide chain synthesized by an enzyme called primase from the 5´- end, acts as a primer and is extended by DNA polymerase. (b) Synthesis of deoxynucleoside diphosphates from ribonucleoside diphosphates. Ribonucleoside monophosphates are synthesized *de novo* and are converted to diphosphates from which deoxyribonucleoside diphosphates are derived and then converted to triphosphates which act as the precursors of DNA synthesis. (c) Synthesis of thymine from uracil. Thymidine monophosphate, the constituent of DNA cannot be synthesized *de novo* but is derived by methylation of deoxyuridylate (dUMP).

take place without RNA ingredients as deoxyribonucleoside triphosphates are polymerised to DNA by the enzyme DNA polymerase and deoxyribonucleoside triphosphates are not synthesised de *novo* in the cells. Actually, ribonucleoside diphosphates are synthesised *de novo* and those are reduced to deoxyribonucleoside diphosphates (Fig. 5.1b) for which enzymes called reductases are required. If these enzymes and ribonucleeoside diphosphates are missing, DNA cannot be synthesised. Deoxyribonucleoside diphosphates are eventually converted to triphosphates, which are required for DNA synthesis. Further, thymine, one of the four bases present in DNA cannot be synthesised *de novo* unlike uracil. Deoxyuridylate (dUMP) is methylated to thymidylate (TMP) with the help of an enzyme called thymidylate synthetase responsible for the transfer of the methyl group (Fig.5.1c). Without the base thymine derived from uracil DNA synthesis is impossible. What do all these mean? Today's method of DNA replication evolved at the time when RNA was already available. It is quite likely that RNA appeared on the earth before DNA. We can also think one step further that RNA was primarily the genetic material and later the responsibility of storing genetic information was transferred to DNA. It is known that RNA is much more unstable than DNA. The former gets easily degraded under alkaline conditions. Further, enzymes responsible for RNA degradation are all around. Single-stranded regions of RNA are very much vulnerable to enzymatic attack. So it is quite likely that the genetic function of RNA was eventually handed over to DNA and RNA took up all other responsibilities for the cells to survive, i.e., replication, transcription and translation. The slave built the master in his own image and started carrying out all the functions itself. Let us discuss, how far we can justify this assumption.

Since I started teaching Biochemistry I used to emphasise in the class that so far a non-proteinous enzyme has not been discoverd or in other words, enzymes are invariably protein molecules and there is

no exception to that rule. It was just like the central dogma DNA→RNA → Protein enunciated by Crick but had to be modified (DNA ⇆ RNA →protein) as soon as it was discovered that DNA could be synthesised on RNA template. Similarly, the dogma of enzymes being invariably proteins had to be modified when the outstanding discovery was made independently by Sydney Altman and Thomas Cech and their coworkers who demonstrated that RNAs can also act like enzymes and the term ribozyme was eventually coined[4]. Altman was working with Rnase P, an enzyme which helps in the maturation of precursor tRNAs to mature tRNAs. It was well known that ribosomal RNAs (5S, 16S and 23S) as well as transfer RNAs are copied from DNA as large precursor RNAs (larger in size than mature RNAs), which are subsequently processed by specific RNases to somewhat oversized individual RNAs which are eventually converted to mature RNAs. Thus, the overall process in some cases involves several steps. The reaction catalysed by the enzyme involves the removal of a block of nucleotides from the 5′ end of a precursor tRNA to produce the mature 5′ end of tRNA. Although the enzyme was found to contain both RNA and protein Altman had a suspicion that the protein part of the enzyme was not acting as an enzyme. It was somewhat unusual thinking for those days. Even when all the proteins were removed and only RNA was left behind it still retained the catalytic activity indicating clearly that RNA at least in this case was acting as an enzyme. Although he provided considerable evidences many were still not convinced. The situation was similar when it was first demonstrated that DNA and not protein is the genetic material in *Pneumococci*. Final proof was produced by Altman in a different way. In those days cloning of the genes came into practice. So RNase P gene was cloned. The product RNA (free from any protein) was found to catalyze the maturation of tRNA from the precursor tRNA. Under physiological conditions, however, the enzyme existed as a RNA-protein complex. The RNA component of RNase P of *E. coli* is 377 residues (nucleotides) long and is known as MI RNA. Its structure is

shown schematically in Fig.5.2. The RNase P reaction is dependent on the structure of RNA and requires a divalent cation ($Mg^{++}$ or $Mn^{++}$) for the reaction but does not require any energy source. It is believed that a hydroxyl ligand of $Mg^{++}$ acts as a general base responsible for the catalytic activity. Although the protein component of RNase P is dispensable in *vitro* its presence is essential under *in vivo* conditions. It is believed that the main function of the protein is to act as an electrostatic sink for the catalytic RNA.

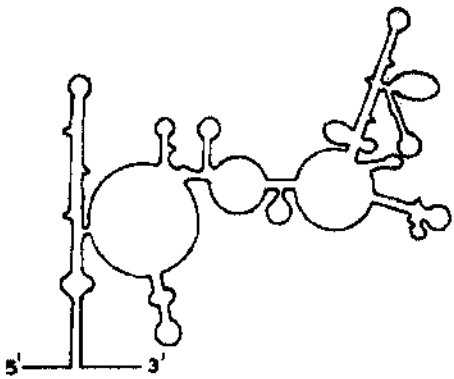



Fig. 5.2 Secondary structure (schematic) of catalytic component (MI RNA) of RNase P of *Salmonella typhimurium*. The above is the sketch of the primary structure of the catalytic center of RNase P which is 377 nucleotides long. As mentioned in the text, RNase P removes RNA from the 3′ end of the precursor tRNAs. RNase P is a typical example of a ribozyme

The other startling observation made almost at the same time by Thomas Cech and his coworkers became a turning point in the history of RNA research. While studying RNA from the protozoa, *Tetrahymena thermophilus,* they found that the precursor ribosomal RNA could splice itself. It has already been discussed in Chapter 3 that messenger RNAs in the eukaryotic cells are not usually synthesised in mature form but as a longer molecule, which contains an interrupted message. Or, in other words, the message is transcribed in segmented fashion. Portions of the message (as per the genetic code) known as exons are separated from each other by the noncoding regions (introns). The introns are removed and exons are joined together in an orderly way so that the continuous message in form of

a mature messenger RNA is produced. The process is known as splicing. The same is found to be true for tRNAs and ribosomal RNAs in the eukaryotic systems. We have already mentioned about the cleavage of precursor tRNA to mature tRNA. This is one way of processing. Some tRNAs have to undergo splicing where a non-functional portion of the RNA is removed from the structure and the two ends are joined together to give rise to mature tRNA. This also happens in the case of ribosomal RNA of *Tetrahymena thermophilus*. (Fig.5.3). As shown in the figure, the portion indicated by the dotted line with arrows, a part of the RNA sequence has to be spliced to give rise to mature ribosomal RNA. This splicing takes place spontaneously although guanosine or a gunosine nucleotide (GMP, GDP or GTP) is required to carry out the reaction. Guanosine carries out specific attack at the 3´end of exon-intron junction attaching itself to the 5´ end of

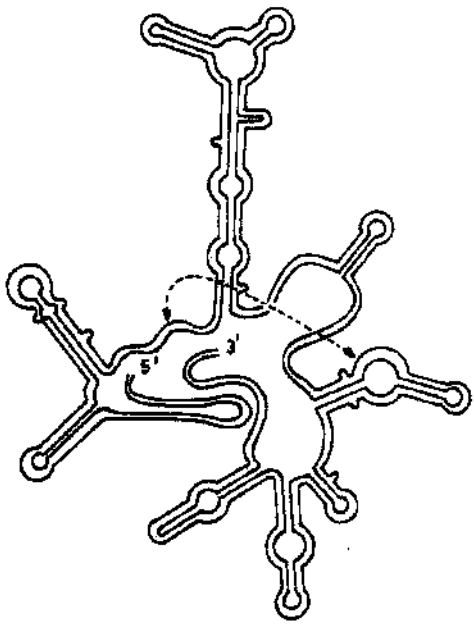



Fig. 5.3 Secondary structure (schematic) of the catalytic centre of *Tetrahymena* self-splicing intron. The portion (intron) indicated by the dotted line with arrow heads needs to be spliced out to make the mature ribosomal RNA (rRNA).The arrows indicate the bases in close proximity in the three dimensional structure and the sites of attack by self-splicing without the help of an external enzyme.

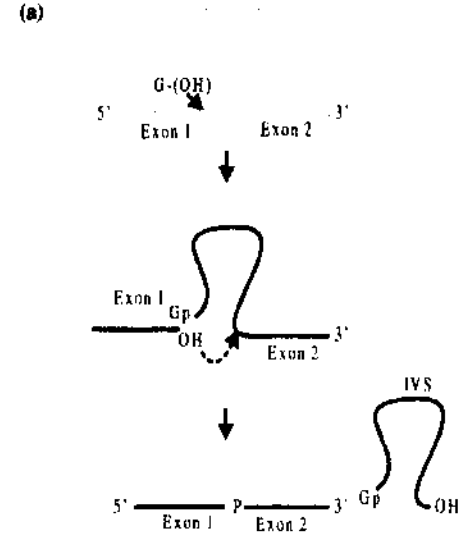



Fig 5.4.. Mechanism of self–splicing *Tetrahymena* rRNA.

(a) The splicing takes place spontaneously no doubt but guanosine or guanine nucleotide (GMP,GDP,GTP) is required to carry out the specific attack as shown. The free OH group at the end of exon 1 attacks the junction point of intron and exon 2 and carries out a transesterification reaction due to which exon 1 and exon 2 are joined together.

(b) Following the excision of the intron (414 nucleotides long) it undergoes a series of intramolecular cyclization (cleavage) reactions which take place leading to the loss of 19 nucleotides from the 5' end. The remaining linker RNA (395 nucleotides long) called L-19 IVS (intervening sequence lacking 19 nucleotides) can act as a catalyst (as shown in Fig. 5.5)

the intron. The free 3'-OH group created at the end of exon makes the subsequent attack at the junction of intron and exon carrying out a transesterification reaction, as a result two exons (exon I and exon II) are joined together (Fig.5.4a). The linear intron, 414 nucleotides long (IVS), thus generated undergoes a series of reactions (Fig.5.4b). In the first step 19 nucleotides are auto catalytically removed from the 5′ end of the spliced intron and the remaining portion is known as L-19IVS (intervening sequence-19 nucleotides). The system is very precise in its excision reaction due to the internal guide sequence, which correctly organizes the bonds to be cleaved and rejoined.

Although ribosomal RNA self-cleaves its intron thus acting as an enzyme, the question arises, it cannot act in a catalytic way. Once the cleavage is done the resultant mature rRNA is not able to carry out the reaction any more. A catalyst has to act repeatedly in a reaction (similar to the protein enzymes). However, it was shown later that the 395 base containing linear nucleotide RNA (the spliced intron) produced after self -splicing can act as a catalyst like a true enzyme after loss of 19 nucleotides from the 3′ end. The resultant linear RNA (known as L-19 IVS or intervening sequence missing 19 nucleotides) has the catalytic property of adding bases to RNA oligonucleotides. A cycle of transesterification reactions takes place as shown in (Fig.5.5). The best substrates are those oligonucleotides which can base pair with L-19 IVS, for example, a synthetic C5 oligomer. Enzyme activity in extending the oligomer is similar to self-splicing reactions in which OH group of guanosine is involved. In this case G-OH group procures a C residue from an oligomer ($C_5$) transfers it to another substrate ($C_5$) and elongates it to C6 by adding one C residue. In this process L-19 IVS is not used up and can carry out the same reaction repeatedly extending to 100 substrate molecules per hour. Although the reaction rate is rather slow in comparison to protein catalysts, yet RNA molecule is quite effective as an enzyme. This type of synthesis is not carried out in the present day cell but might have taken place

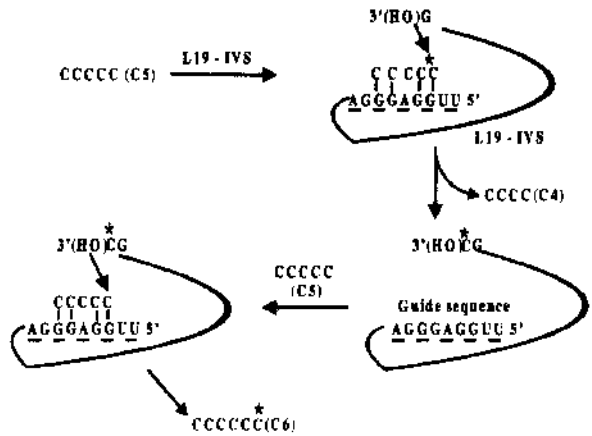



Fig.5.5. Catalytic property of L-19IVS which acts as a ribozyme. The 3′ OH of the guanine residue present at the 3′ end of L-19 IVS plays the key role in the catalytic reaction and the guide sequence marked with broken lines at the 5′ end binds a polymer (say $C_5$) by hydrogen bonding. C* residue is excised from the polymer and transferred to another $C_5$ chain and a $C_6$ polymer is thus formed. This can repeat a number of times and the chain can be lengthened up to a few residues.

extensively in the RNA world and thus may be a relic of that world. Those RNAs which have similar catalytic functions are, in general, called ribozymes. These discoveries shattered another age-old dogma 'Enzymes are invariably protein molecules'. Proteins act as enzymes because they have some special groups (the constituent amino acids carry those). Although ribonucleic acids have only a very limited number of such groups (specially OH and $NH_2$ groups of bases) yet they have the potentiality to act as catalysts. This was overlooked earlier as the vast majorities of enzymes discovered so far turned out to be proteins. Although there are limited number of ribozymes and their enzymatic activity is comparatively poor yet this has great implications in the evolutionary process. RNAs not only act as the genetic material and are deeply involved in transcription/translation

system but also act as enzyme molecule. Thus only RNA molecules seem to carry out a variety of vital functions, which no other molecules can. This led to the wide spread belief that RNA could carry out all the functions in the absence of DNA and even protein. Credit goes to Walter Gilbert[5] to conceive of the RNA world preceding the present day DNA world on the basis of the findings discussed above.

By the time Cech and Altman were making the Nobel Prize winning discoveries, I got myself deeply involved with *E.coli.* ribosomes while working in the Department of Biochemistry of the Institute of Medical Sciences at the Banaras Hindu University. The story has been partly told in Chapter 3. There are three possible ways by which 30S and 50S ribosomes may get associated. It can be either protein-protein or RNA-RNA interaction between them, RNA-protein interaction being a distinct third possibility. Since we had pure RNAs in our hand we were curious to know whether the two can associate with each other but they did not. Then we argued, may be they do not have the proper three-dimensional shape as the proteins were removed to have pure RNAs. We also remembered that Nomura could constitute ribosomes from RNAs and proteins only in presence of high $Mg^{++}$ and high salt concentration. We tried that and the two readily associated. Their association was tested by several methods; light scattering was one of those. On decreasing the salt concentration this dissociated, as expected. We were encouraged with the finding and performed some experiments to find out whether this RNA-RNA complex has some biological activity like the intact ribosomes. Our first attempt was unsuccessful. After adding a few proteins the complex demonstrated very feeble but significant biological activity. This has also been discussed earlier. Although we published the results the expert ribosomologists did not take the results seriously. Several years later Harry Noller reported his epoch-making experiment and the outstanding results. He performed the experiment in a reverse way. With the help of a protein-degrading enzyme he systematically

removed the protein from intact ribosomes and tested their biological activity. Even when 90% of the proteins were removed they retained the full biological activity. They, however, did not use *E. coli.* ribosomes as we did and chose ribosomes of an archaea, *Tetrahymena*, which survives under adverse conditions like, high salt, high temperature etc. Noller, however, mentioned about our work in his paper. Actually it was one of the earliest observations in support of the RNA world. Very recently Cech has carried out *in vitro* synthesis of a ribozyme, which can catalyze peptide bond formation. Following the publication of our results I had written a review article entitled 'Primitive ribosomes' in an Indian journal. It drew the attention of Sydney Fox. He wrote a letter criticising some of the points raised. Sydney Fox, an authority on evolution did show earlier that a mixture of amino acids under certain conditions may have very feeble but still significant catalytic activity. He stated at the end of the letter *'Why did you choose the unfortunate title'* (the language is mine as I lost his letter eventually). I had no answer. Noller's subsequent experiment after 7 years and Cech's later experiments provided the answer in retrospect. As mentioned earlier, the mechanism of peptide bond formation is still a mystery but enough evidences have been obtained to vouch for RNA as catalyst.

Other than acting as enzymes, RNAs have more diverse functions. Let me describe some of those to show its functional importance. However, splicing of the largest group of introns, found in nuclear mRNA primary transcripts is mediated by organelles known as 'spliceosomes'. These undergo splicing resulting in the same lariat formation as in case of group II self-splicing introns (not shown). However, these are not self-splicing but require the help of special organelles known as spliceosomes. Those can be visualized under electron microsocope as particles of 40-60 nm diameter with distinctive morphologies. These are complexes of small nuclear RNAs (snRNAs) with proteins. Five snRNAs (U1, U2, U4, U5 and U6) are

involved in this splicing. SnRNAs are small in size (100-200 nucleotides) and their complexes with proteins are known as snRNPs (pronounced as *Snurps*). The function of snRNAs is mainly for proper aligning of introns and exons for the splicing to take place. As shown in Fig.5.6, U1 snRNA aligns at the junction of intron and exon by base pairing and U2 snRNA base pairs with intron helping it to form the lariat type of structure. Base A (circled in Fig.5.6) which bulges out of intron due to non-pairing is thought to be involved in the formation of this type of structure. Other small nuclear RNA protein particles, U2, U4, U5 and U6 then bind near the 3′ splice site to form the spliceosome, which carries out the actual splicing reaction.

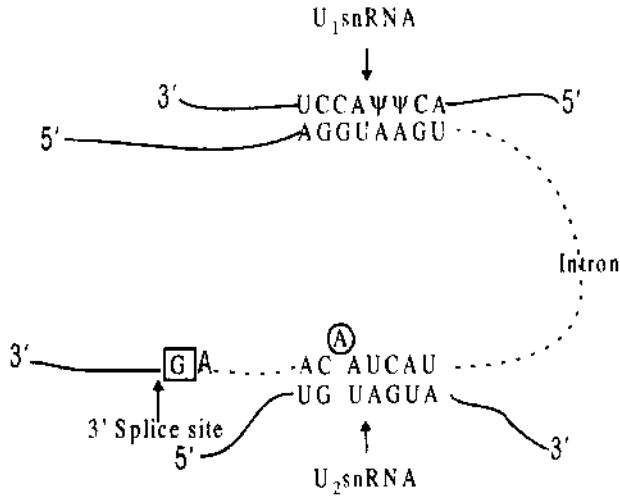



Fig. 5.6 Role of U1 and U2 SnRNAs in the formation of structure for splicing. At the first step of splicing of nuclear mRNA, U1Sn RNA and U2 SnRNA of the two SnRNPs, (see text) bind to the 5′ and 3′ regions respectively of the intron due to their sequence complementarity near the splice site. Then other small nuclear RNA-protein particles U4, U5 and U6 bind and form the 'spliceosome'

As more and more complex physiological phenomena are looked into, new functions for RNAs are come across. Sometime back it was reported that 7S RNA present in signal recognition particles (SRPs) is responsible for protein translocation across the lipid bilayer membrane of the endoplasmic reticulum. It should be mentioned that the proteins, which are to be secreted outside the cell, are synthesised by ribosomes while docked on the outer membrane of the endoplasmic reticulum. The docking is done with the help of signal recognition particle present in the cell (Fig.5.7). The proteins synthesised on such ribosomes have a signal sequence to start with. SRP binds to that sequence and guides the ribosome to the receptor on the endoplasmic reticulum with the help of its own receptor. Thus in the very first step the protein is directly translocated within the endoplasmic reticulum with the help of signal recognition particle (SRP). 7S RNA, which is a part of SRP is known to play a critical role in such process. It occurs in all three kingdoms of living system (bacteria, archaea and

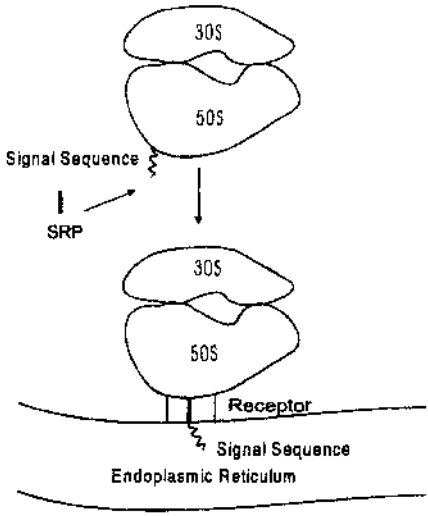



Fig. 5.7 Role of 7S RNA in translocation of newly synthesized protein within endoplasmic reticulum. The proteins which are exported outside the cell are directly transferred at the time of synthesis within the lumen of endoplasmic reticulum. 7S RNA which is 250 nucleotides long is a part of signal recognition particles (SRP) with six different proteins which are responsible for translocation of proteins following the docking of ribosomes on the reticular structure. (see text for discussion).

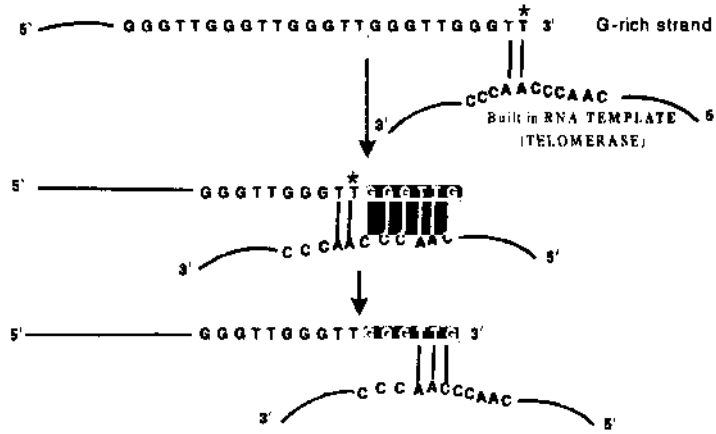



Fig. 5.8 Role of RNA in repair of the telomeric end of chromosomes. As discussed in the text, telomerase which does the repair job has in-built RNA template for the purpose. One of the two strands which overhangs is G-rich, the other strand is C-rich. The overhanging G-rich strand, which is usually damaged is repaired by using in-built C-rich template of the telomerase.

their own structure like that of telomerase. The C-rich strand can be synthesised in the usual way by the regular DNA polymerase but the G-rich strand is left with an overhang, which is subsequently protected by a protein. The mechanism is primarily to protect the ends of the chromosomes although they don't have any message. The main object of the repair process is not to expose the internal message to external damaging agents.

Editing articles, books, newspapers, films etc is a routine job in literary work. Similarly, editing of life processes is also another interesting arena of RNA function. Like editors nature edits some messages with the help of RNA. Three different laboratories while studying mitochondrial genes in the parasite *Trypanosoma brucei* and *Leishmannia tarentolae* ran into a puzzle. The coding in some genes did not correspond strictly to the sequence of amino acids in proteins,

which these organisms synthesise. A more puzzling situation was that the proteins were synthesised in spite of the stop signals present within the reading frame. To settle the issue the mRNA, which directs the synthesis of the protein of interest was isolated and its sequence determined. It was found to be in variance with the genetic code. Eventually it was realised that the genes are being corrected not only at the DNA level but also at the transcriptional (RNA) level. A closer look revealed the fact that the addition or deletion of one or more residues was responsible for restoring the normal reading frame. It appeared that RNA following transcription is being edited. The discovery of a species of RNA known as guide RNA provided the clue to the mechanism of editing. The guide RNA is rich in U and has some bases complementary to the region of mRNA, which is to be edited. Guide RNA associates with mRNA at the site of editing. Its poly U tail which is involved in editing is shown in (Fig 5.9). Just like in splicing, nucleophilic attack of U (at the 3′ terminal) at the

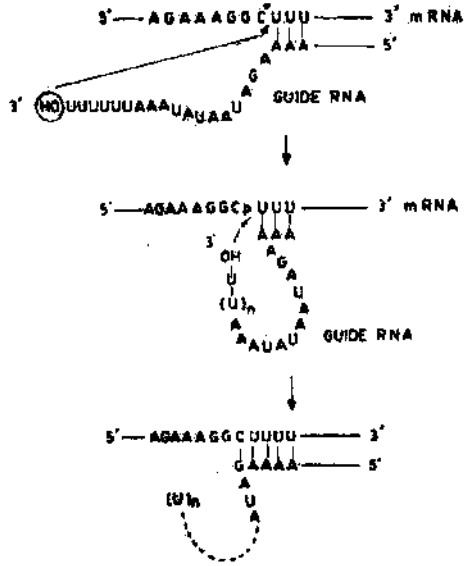



Fig. 5.9 Mechanism of editing (proof reading) of the transcribed genetic message. As discussed in the text, messages transcribed from some genes may be nonsense due to improper reading frame. Editing enzyme has some guide RNA as template. The above message has the sequence CUUU at the site pointed by the arrowhead. An additional U has to be inserted after C to make it meaningful. The guide RNA has a long tail of U at the 3′ end. The terminal U is inserted after C by transesterification reaction.

site of insertion leads to the splitting of RNA. Following insertion of U by nucleophilic attack there is transesterification and joining of the split strand with the insertion of U at that site. Several such U's can be inserted by repetition of the process. The deletion of U may also take place by reversing it. The transesterification is the key to the editing process, as in splicing. The 'RNA-editing' was thus the term coined for this phenomenon. It may throw some light on the early stages of evolution when RNA acted as the genetic message. It could edit the message, whenever necessity arose. The phenomenon may be looked upon as one of the relics of the RNA world.

Another such fossil apparently of the RNA world is more interesting. Some microorganisms are known to contain multiple copy single stranded DNA (abbreviated as msDNA). The single-stranded DNA is fused with a RNA strand (Fig.5.10). At one end, the two are linked through a 2′-3′ phosphodiester bridge which is possible because RNA has an additional free 2′-OH group. At the other end the two are

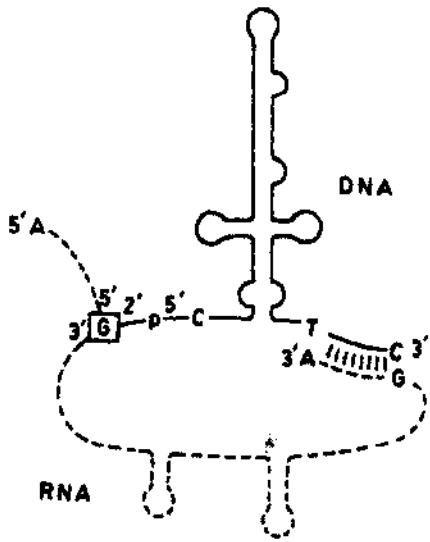



Fig.5.10. Structure of multiple copy single strand DNA (ms DNA) the relic of RNA world. msDNA has two parts, 5′ portion is a single stranded DNA linked to a single stranded RNA (3′ part) by hydrogen bonding to give rise to a circular structure. The link between DNA and RNA is close to G near the 5′ terminal of RNA and is formed by an unusual 2′-5′ phosphodiester bridge in place of normal 3′-5′ diester bridge present in nucleic acid.

joined together through hydrogen bonds between complementary base sequences, giving rise to a circular DNA-RNA hybrid structure with the protruding tail of RNA strand. This is reminiscense of a primer of RNA for DNA synthesis linked to newly synthesised DNA (with 3′, 5′ but not 2′, 5′ phosphodiester bridge). The synthesis of msDNA is also interesting. The two components DNA and RNA are synthesised from two opposite strands (partially overlapping) of DNA, which are most probably transcribed by a retroelement (like reverse transcriptase). The 2′, 5′ bridge between the two is a mystery although there are abundant examples of such linkages in nature, for example, lariat formation in the processing of mRNA. One may speculate that such hybrid molecules are the fossils of the early period during which DNA evolved to act as a carrier of genetic message contained in RNA, instability of RNA being the compelling reason.

It is worth mentioning here that a sensational development is taking place in the RNA area, which may have far reaching consequences. Messenger RNA is the only type of RNA which are coding RNAs derived from DNA. Other RNAs, such as tRNAs, rRNAs, snRNAs etc., though transcribed from DNA, are noncoding RNAs. A large number of RNAs designated as micro RNAs (miRNAs), which are small RNAs, about 22 nucleotides in length, also belong to noncoding RNAs. These have been discovered only recently in plants and animals and have regulatory functions in development. For example, genes lin-4 and let-7 of the worm *C.elegans* which regulate developmental timing of the worm act as repressors of the target genes lin-14, lin-28 and lin-41. The miRNAs, produced from the precursor RNAs transcribed from the regulatory genes, do not interfere with messenger RNA production nor do they degrade messenger RNA but temporarily stop the translation of the message at the ribosomal level. These miRNAs are derived from preRNAs about 70 nucleotides long which are transported to the cytoplasm and converted to mature miRNAs by an enzyme called Dicer (an RNaseIII). Gene silencing mechanism

originally discovered in plants and known to act as defence mechanism against viruses operates through another class of small RNAs designated as siRNAs (small interference RNAs). Those eventually lead to the degradation of messenger RNAs. Thus their mechanism of action is different from that of micro RNAs. However, it has been recently observed that a class of miRNAs may act in the same way as siRNAs do by degrading mRNAs. Thus micro RNAs may act in both ways repressing mRNA at the translational level and mediating cleavage of mRNA. The discovery of the latter has led to the opening of a new chapter in Molecular Biology.

To come back to the original discussion, the existence of the RNA world before the DNA world does not mean the end of the story. There are sufficient grounds to assume that RNA may not be the starter. The life might have started in the preRNA world and initially in a different fashion than both DNA and RNA world would have had. The famous experiment of Stanley Miller, student of H.C. Urey, showed that there were plenty of opportunities for life to be generated in the primordial soup. The composition of the planets in our solar system, galaxies, stars etc. clearly indicates the presence of $CO_2$, $NH_3$ and even $H_2O$ in the early days of earth. Miller continuously refluxed a mixture of chemicals like $CO_2$, ammonia, methane etc. in water exposing the vapour to electric discharge (simulating the thunderstorm in primordial atmosphere) and could show the formation of methane, traces of sugar and even some simple amino acids which are the building blocks of the living system. These clearly indicated that polypeptides or protein-like molecules could have been present in the world when life evolved. Stanley Miller, then a 23-year-old graduate student of H.C. Urey at the University of Chicago published his work in 1953, the same year in which Watson and Crick published their seminal paper on structure of DNA. Miller might have been inspired by the concept of Oparin on the origin of life on earth. Herold C Urey, a Nobel Laureate in Chemistry at the University of Chicago proposed earlier that the

atmosphere of the earth in the early time was reducing and rich in gases such as hydrogen, methane and ammonia, which are abundant on Saturn, Jupitor and Uranus. It was originally Urey's Oparin-like concept that inspired Miller to conduct the famous 1953 experiment. The experts are, however, divided in their opinion about the type of atmosphere, which prevailed on the primordial earth. An alternative view has been developed about the origin of life. According to that hydrothermal vents below the sea bed was the site of origin. Sulphur provided by sulphite of iron and other metals was the supplier of food and also acted as source of energy. This is also not universally agreed upon.

RNA might have even evolved a little after protein. If this is true, then the existence of only RNA in RNA world would be ruled out. If we think that RNA and protein originated almost at the same time and evolved simultaneously, a situation similar to that depicted in (Fig.5.11) is most likely. In the prebiotic world ribose and bases might have originated independently and in combination with phosphate could give rise to ribonucleotides which in turn gave rise to deoxyribonucleotides, as discussed already. Ribonucleotides polymerized to RNA-like material (oligonucleotides) and eventually RNA of the modern world evolved. Similarly, amino acids synthesised in the prebiotic soup gave rise to peptides and eventually proteins, as existing today. There was every chance that oligonucleotides and peptides during their evolution recognized and communicated with one another and shared some of the functions of the primitive cells. Under such a situation RNA had the genetic information and also acted as a catalyst to some extent. Some proteins were better catalysts and carried out more diverse functions. They finally took over more and more of the catalytic functions as we come across them in the present day world. The DNA evolved following RNA using the deoxyoligonucleotides, which in turn were formed from ribonucleotides. DNA slowly took over the main function of RNA

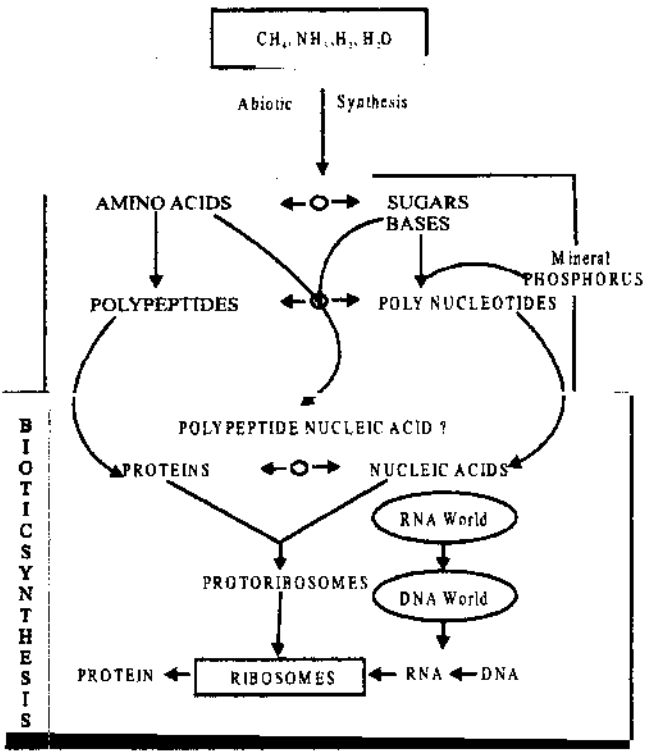



Fig. 5.11 A model of coupled evolution of RNA and protein in the primitive world before the ushering in of the RNA world followed by DNA world. It is assumed that as soon as the amino acids as well as building blocks of nucleic acids became available polypeptides and polynucleotides started evolving independently with continuous information exchange between them. Originally RNA carried out multiple functions including storage of genetic information. Later the responsibility of carrying genetic message was passed on to DNA to switch over to DNA world where proteins and nucleic acids of both types started playing the predominant role.

i.e.,carrying the genetic information. This choice was most likely as DNA is more stable than RNA, as already mentioned. The latter was confined to other functions; specially transfer of the message from DNA for the synthesis of proteins by the translational system in which RNA has diverse and important roles to play. The primitive life form can be also analysed from another angle. Ribosomes, the protein synthesising machinery of the cell, are ubiquitous and evolved extensively starting from bacteria and ending in the most complex organisms but the basic structure has remained the same. Protoribosomes (merely a simplified RNA-protein particles) may be imagined as the precursor as shown in (Fig 5.11). The communication between polypeptides and polynucleotides is thus assumed as the basic feature behind the origin of life.

However, it is being speculated by Orgel and others that RNA world was preceded by another world known as PNA world. PNA means peptide nucleic acid. It consists of a peptide (polyamide) backbone to which nucleobases are attached (structure not shown). These molecules have been shown to pair with DNA and RNA according to Watson-Crick base pairing rule. This structure hints at the interesting possibility of PNA acting as the genetic material. However, there is not much supporting evidence beyond this. Another argument has been advanced by Stuart Kaufman, an expert working on the origin of life that the start of life was not with any genetic material but the simple metabolic (chemical) reactions linked to each other through a network. These reactions were thought to augment each other and work in a cyclical fashion. Thousands of such metabolic (rather chemical) reactions are known today. Kaufman's idea is that very simple molecules, which evolved in the prebiotic world carried out the basic metabolic reactions and also learnt how to duplicate as if they were enclosed within a cell. Once this system was established the genetic system evolved later to sustain life. Most of these ideas have remained as purely speculative. Until and unless concrete

evidences are found, the origin of life, even the existence of RNA world would remain as purely speculative.

There is no controversy that the knowledge about prebiotic era of the earth, origin of life and subsequent development have remained mysterious as it is not easy to extrapolate the present day information back to those days. There are no alternative ways than to speculate and the speculation may turn out to be absolutely wrong. One such example may be cited here and that may be a good lesson. In 1982 the research submarine Alvin was guided by the pilot to a site where a hot spot 3 km beneath the Pacific Ocean existed. On drilling there, a grayish-white fluid spiraled out like smoke from the Earth's interior through a deep-sea hydrothermal chimney. A methane producing microbe isolated from the collected material was named *Methanococcus janneaschi* after Holger Janneaschi, the leader of the expedition. The microbe was recognised as a member of the archaea, a primitive form of life. At that time very little was known about that organism, but 14 years later its total genome (1.66 million bases) was sequenced and a reasonable amount of information about the archeon was collected. Since the archaea were found to thrive in hot springs and bacteria cannot generally survive under such conditions it was thought that the latter are more primitive although appeared to be close to bacteria. Hence the name 'archaea' was coined. But later the mistake was realised. Carl Woese, one of the experts on archaea, later showed by ribosomal DNA analysis that archaea may not be very primitive. They have some characteristics of eubacteria and some of eukaryotes. The living kingdom was earlier grouped under two broad domains, prokaryotes and eukaryotes. With the realisation that archaea belonged to a third domain the whole living system has now been divided into three kingdoms, bacteria, archaea and eukaryotes. The reason for citing the above example is to emphasise that it will take quite sometime to know unequivocally whether life started as a metabolic network or through PNA world or RNA world

or DNA world. It is, however, quite likely that RNA world existed before the DNA world and today's slave RNA created its master DNA in its own image. If it is true, RNA will at least turn out as an earlier player of the music of life.

## References

1. *Origin of life* (2nd edition). A. J. Oparin , translated by S. Margolis (from Russian to English), Dover NewYork (1953).
2. *Primitive ribosomes.* D.P.Burma, *Journal of Scientific and Industrial Research* (India) Vol.43,p 388, (1984).
3. *RNA world and ribosomes.* D.P.Burma, *Current Science* (India) Vol. 63,p 550 (1992).
4. *The RNA world*, ed. By R.F. Gesteland, T. Cech and J.F. Atkins, Cold Spring Harbor Monograph Series, 37 (1999).
5. *RNA world,* W. Gilbert, *Nature*, Vol. 271,p501 (1978)

*****

# Chapter - 6
## When the tune goes wrong
### (Cancer of the cells and its challenges)

*Music sounds horrible when the tune goes wrong. Similarly our life becomes miserable when cancer takes over. There is hardly any difference between normal and cancerous growth except that the former is a regulated one while the latter unregulated. Though cancer has not been completely conquered, a lot of information about the normal cells has been obtained through the study of cancerous ones. This chapter records the cause and effect situations at the DNA level, which lead to cancer. The most amazing information we have today is that some genes (protooncogenes), which are essential for normal living, turn into cancer-producing genes (oncogenes). Therefore life-long all of us are exposed to the potential danger of cancer. The cure has still remained a challenge.*

In 1955 when I visited the National Institute of health at Bethesda on my way to join the University of Wisconsin at Madison, Bernard Horecker, an enzyme chemist of international reputation and Herman Kalckar, a Danish Biochemist visiting his laboratory advised me to apply for a fellowship at the Jane Coffin Childs Memorial Fund for Medical Research to work under their joint supervision. The funding organisation directed me to meet Dr. Charles Huggins, Director of the Ben May Cancer Institute at Chicago. The laboratory building was a sky scrapper and on the top floor was Huggins's office, which looked like a large living room. I was somewhat scared but he was

very friendly and cordial. I explained my plan of work to him with Horecker and Kalcker, which I had submitted earlier. Before bidding good bye he said, 'You are a bright young man; I wish, you get the Nobel Prize some day.' I was excited and as soon as I came down I conveyed the message to one of my friends, who was then working in the same Institute. He was my colleague from Bose Institute. My good friend said, 'Don't be so excited, he tells that to everyone as he himself is aspiring for the prize for quite sometime' The Nobel Committee really took a long time to award him in 1966 for his contribution to cancer chemotherapy. I had met him in 1956, 10 years earlier.

After the detour let me now come to the actual topic of this chapter. In spite of colossal amount of research carried out on cancer and its therapy[1] during the last century it has still remained a challenging problem for mankind. We neither fully understand the etiology of cancer nor have learnt ways to combat different types of cancer. There is no doubt that the longevity of cancer patients has increased significantly due to modern therapy but survival chances has still remained slim. Although we have learnt a lot about human biology due to studies on cancer yet we still do not know all the causes of cancer. Once in a while, we come across a sensational discovery of some 'wonder drug' of cancer and in most cases it happens that the drug becomes short-lived; it is either successful in a special type of cancer or only of limited success in a few cases. It is interesting to recall an event while I was working at the University of Wisconsin, Madison. Charles Heidelberger of the Medical School did hit upon a wonder drug, which would cure cancer completely. One of my classmates from the University College of Science at Calcutta was working in his laboratory at that time. Heildelberger was present in the 1964 meeting on Nucleic Acids organised by P.M. Bhargava at Hyderabad. My classmate would not let me know what the drug was. Naturally, it was kept as a trade secret. But on a large-scale hospital-based trials it was not very successful although Heidelberger's own

uncle was completely cured. Later we came to know that it was a very simple analogue of thymine, one of the bases of DNA, where methyl group had been replaced with fluorine and is thus known as fluorouracil. Apparently it interferes with the synthesis of DNA in cancer cells. But it interferes with that in normal cells as well. In spite of that fluorouracil is still used as an anticancer drug. Its success is, however, limited and unpredictable. A good number of anticancer drugs with partial success were hit upon subsequently.

In general, we understand that cancer is related to some abnormality. All the cells of our body happily divide when they have to, stop division when there is no necessity to do so and thus maintain the proper physiological state of health. But once a cell loses this property the desirable state of survival is lost. When the cells go on dividing without any regulation and do not know when to stop that leads to a cancerous growth, a state which is called malignancy or neoplasia. The basic question is, how does it happen? How does a normal cell start behaving abnormally? What is the signal? Who signals it to be abnormal? In a sense, what goes wrong? Without going into details at this stage let us discuss about 'contact inhibition'as observed in normal cells. When we have a wound, the cells in the damaged area receive some kind of signal to repair it, so they start dividing and fill up that area. But as soon as the wound is repaired the cells stop dividing. This is known as contact inhibition. Cells in contact communicate with each other and advise mutually to stop the division. The matter seems to be rather simple although it is not. The contact inhibition is part of regulated cell growth. The loss of such inhibition is naturally expected to give rise to malignancy. Still we are far from pinpointing the signal. Researchers[2] are working very hard to understand it. At this stage it is better to clarify the difference between a benign tumor and a malignant tumor. Benign tumor is infrequently a localized disease, which is not as dreadful as neoplasia. Cancer cells detached from the primary tumor may translocate to distant sites and

grow as secondary tumor. This process is known as metastasis. The capacity of malignant cells to grow at distant sites is the basis of neoplasia. The mechanism of metastasis is not fully known. The tumor is a mosaic of cells having different properties with respect to cell surface receptors, morphology, enzymes and also the capacity to metastasise. Once metastasis is established treatment of cancer becomes the most difficult job.

After we came out of the Stone Age we became more and more dependent not only on metals but also on a wide variety of chemicals. The list of chemicals has lengthened dramatically. Initially, we were depending on naturally occurring ones, which were and still are abundant. However, when we learnt how to synthesise them and also the new ones, which do not occur in nature, their creation and subsequent use knew no bounds. It took us long time to realise that many of those cause cancer, or in other words, they may turn out not only toxic but also cancer-inducing. These chemicals do not necessarily belong to a particular group but are of wide variety in nature. Hydrocarbons containing polycyclic rings fused together are potential carcinogens. Their tumor producing capacity has been demonstrated by applying the chemicals in various doses on the skins of experimental animals and observing whether that leads to tumor formation. Modern civilisation is very much dependent on 'plastic ware'. We use those randomly in various forms. But many of us are not aware that some of those materials may slowly leach in water in very low amounts in spite of their apparent insolubility in water and we are constantly exposed to those harmful materials, which definitely have the probability of causing cancer at least in some individuals.

'Cigarette smoking is injurious to health'. We see the statutory warning in the advertisements of various brands of cigarettes. Some chain smokers also argue that they smoke heavily but still do not develop cancer. It is difficult to explain why. But the statistics is there to help us out. There is no doubt that cigarette smokers are much

more prone to cancer than non-smokers. Some non-smokers also have the chance of developing cancer if they spend more time in the vicinity of the smokers (passive smoking). People living in countries, which do not strictly follow the rules and regulations regarding exhausts from cars, trucks, factories etc. are constantly exposed to potential carcinogens. It is no wonder that cancer cases are increasing due to pollution of the atmosphere, primarily from the industrial wastes and random use of a variety of chemicals.

However, chemicals are not the only agents that cause cancer. Sunrays are good for our health; sunbathing is a popular pasttime among the Westerners. In this part of the world (near the equator) we are lucky to enjoy sun rays more than others. However, ultraviolet component of the sunrays may induce cancer. Ozone layer which acts as a canopy of the atmosphere protects us from the ultraviolet light. Due to extensive use of fluorinated compounds as refrigerants the ozone layer is being gradually depleted and pose a great risk to our health. In the Western countries the use of such refrigerants has been banned. X-ray is now-a-days almost routinely used for check up of many diseases. But exposure to X-ray (even at a low dose) for a long period of time causes cancer. Therefore radiologists work behind lead screens to prevent the routine exposure to X-rays. Following the discovery of radioactive isotopes people handling those without taking special care developed cancer. Madam Curie died of cancer without knowing the dangerous effects of radioactivity. All the nuclear plants and nuclear devices have to safeguard the health of the workers. Cherbanoyl mishap in the erstwhile U.S.S.R still looms heavily in the world scenario. The suffering was greater for the people who survived the direct devastating effect of the atomic bombs dropped on Hiroshima and Nagasaki due to the exposure to radioactivity. Japanese scientist Okazaki (after whom 'Okazaki fragments' produced during DNA synthesis are named, as mentioned in Chapter 1) was one such victim.

There is another etiology of cancer, well established in animals but not in human. Infections with some viruses (not all) are known to induce cancer. For example, *Rous* sarcoma virus which causes sarcoma (tumor of muscle) in chicken is an RNA-containing virus. Viruses, as we already know, may have either DNA or RNA as the genetic material. Polyoma virus which is also a causative agent of cancer has its genetic material as double-stranded circular DNA. By now a large number of cancer producing viruses have been discovered. Many of those can be tested in cell culture. Under tissue culture condition cells can proliferate in a suitable medium and normally form monolayer sticking to the bottom of the culture plate. The sticking to the surface is one of the characteristic properties of normal cells. As soon as the total surface of the plate is covered and thus a monolayer culture is formed they stop dividing due to contact inhibition. However, in case of cancer cells unregulated growth takes place which results in multilayers due to loss of contact inhibition. Some tumor cells can be grown in suspension culture like bacteria. The best example is leukemic cell line. It is also possible to maintain some tumor cell lines for a long time, for example He La cells derived from the uterus of an African lady Henrietta Lacks who had uterine cancer are maintained even today although she died long ago. Her cells have become immortal. This tumor cell line has been maintained by serial culturing. Thus the cancer cells unlike normal ones may be immortal. Although viruses have been shown to be the causative agents of many animal tumors strong evidences are not available in the case of humans. However, normally grown human cells in culture can be made malignant by infection with animal viruses. This is an indirect evidence to assume that some human tumors may have the etiology of viral infection. For example, Hepatitis B and C viruses are most probably associated with hepatocellular carcinoma. Similarly, papiloma viruses are linked to anogenital cancer and skin cancer. Etiologies of some cancer are mentioned in (Fig.6.1). Lately it has been realised that genetic predisposition is one of the important elements for the

susceptibility to the disease. Ageing also contributes significantly to the development of cancer. In this connection it is interesting to recall the observations made by Mel Greaves[3] in his book 'Cancer, the evolutionary legacy'. The author argues why the old paradigms of infectious diseases or genetic disorders have proved fruitless, when trying to account for the complex and elusive puzzle that is cancer. Rather he suggests that by looking at cancer in the evolutionary context, we can answer some of the big questions in cancer that concern almost all of us. Drawing on both ancient and most

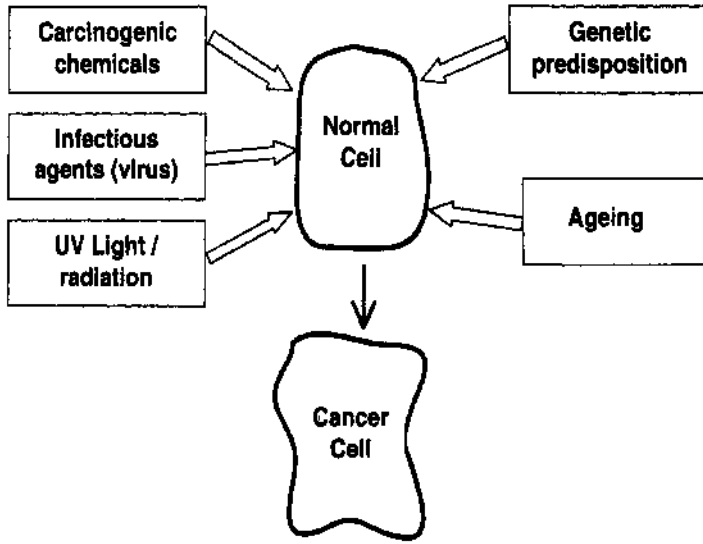



Fig 6.1 A variety of causes of cancer. Of several causative agents, the main three are the following : (a) chemical carcinogens (b) various types of radiations, specially ultraviolet light and (c) tumor viruses. Besides the above-mentioned external agents oncogenes which are parts and parcel of genome are also responsible for cancer. Genetic predisposition and ageing as well are contributory factors. As a whole, oncology is rather as complicated as physiology.

evolutionary legacies he shows how human development has changed the rules of evolutionary games,trapping us in nature-nurture mismatch.

So far we have been involved only in a generalised discussion and have not raised the issue at the molecular level[4]. If we have to do that then the genetic material or DNA has to be naturally brought into the picture. It is easy to say that but not so easy to explain. We are aware that DNA molecules have two types of information, one is the structural information of proteins and RNAs and the other, functional information through which the gene expression is regulated. This knowledge we acquired originally from Jacob and Monod's operon concept although it is valid mostly for prokaryotes (Chapter 3). In case of eukaryotes genes are hardly organised in operon, they are individually regulated. Anyway, what has become evident from various studies at the molecular level is that if something goes wrong in the regulatory functions of DNA which control the expression of the genes, the neoplasia starts. Just like the mutations of bases of DNA are responsible for the structural changes of RNAs followed by changes in proteins, the mutations in the regulatory elements are likely to end up in the disturbance of regulation. For example, disturbance in the production of the factors for cell division results in unregulated cell growth. The cell division is intimately connected to the replication of DNA which starts duplicating when the cell plans to divide. As soon as it is duplicated and the cell division is complete the duplication halts. Depending on the type of signal it receives, the cell gets ready to divide again. The cell cycle is actually the reflection of various stages of DNA duplication followed by pauses as shown in (Fig 6.2). One of the major events is the duplication of DNA (or synthesis of DNA), which takes place in the S phase. This is preceded and followed by G1 and G2 phases respectively. In the G1 phase the cell prepares itself for the synthesis of DNA, which takes place in S phase. G2 phase is the period during which the cell gets ready for mitosis, which

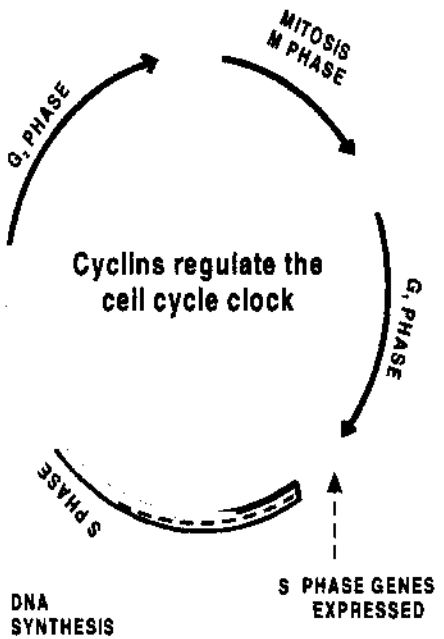



Fig 6.2 Phases of cell division cycle. There are different phases in cell division cycle as described in the text.

takes place during the next phase (M phase). During the M phase the duplicated DNA has to be divided for passing on equally to the daughter cells. At every stage special factors (known as cell division factors) are necessary to carry out the various functions. Besides the regular mechanisms for DNA synthesis there are repair mechanisms which become most effective after the synthesis of DNA so that the inherent defects of the synthesized DNA are practically removed. Thus the phosphorylation and dephosphorylation of the factors are essential to regulate their activities. If this gets disturbed the cell division is expected to go on at an uncontrolled rate.

At this stage let us discuss another etiology of cancer. As already

mentioned, among the environmental factors ultraviolet light is one of the strong causative agents. In this case there is no doubt that DNA is the direct target. Ultraviolet light is known to cause damage directly to DNA, as a result aberrant DNA is produced. To cite an example, two thymine bases present side by side in DNA become linked to each other by crosslinking induced by ultraviolet light and form a dimer ( Fig.6.3.). Carbons 5 and 6 of each are involved in the linkage to form a cyclobutane ring. Due to this a kink is produced in DNA. Such DNA cannot faithfully duplicate and any base can be inserted in place of thymine resulting in mutations leading to cancer. There are, however, repair mechanisms to chop off the dimers and replace those with right type of bases but there are also chances of other bases to get in their place during replication and thus to cause mutation. If there is excessive damage produced no proper repair is possible. Even simple aberration of DNA structure may cause cancer. Excessive use

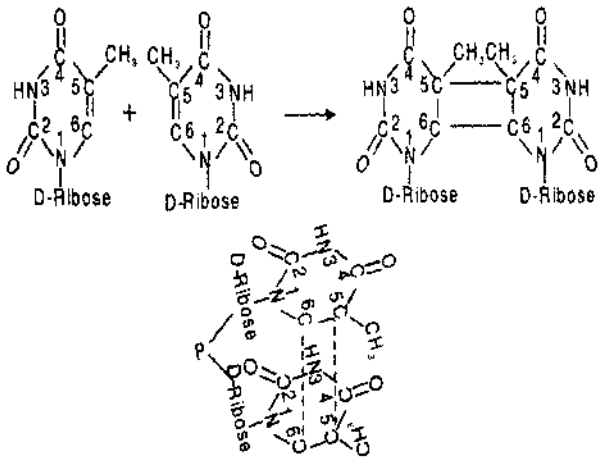



Fig.6.3 Formation of thymine dimer by ultraviolet light. When there are two thymine residues next to each other in DNA, the ultraviolet light sometimes induces the linkage between the two to form thymine dimer.

of ethidium bromide, which has a similar planar structure like purines and pyrimidines leads to deleterious effect. Due to the planar structure it can intercalate between base pairs of DNA and deform the DNA structure (Fig 6.4). During replication altered DNA template produces mutant DNA molecules with wrong bases inserted, leading to cancer.

In case of chemical carcinogens a similar situation is observed. For example, many carcinogens are alkylating agents. Those introduce methyl (an alkyl) group into purines and pyrimidines, the primary site of attack being N-7 position of guanine and to a lesser extent the N-3 position of adenine. Such chemical changes may lead to anomalous base pairing during replication, for instance, guanine instead of pairing with cytosine may pair with thymine, thus

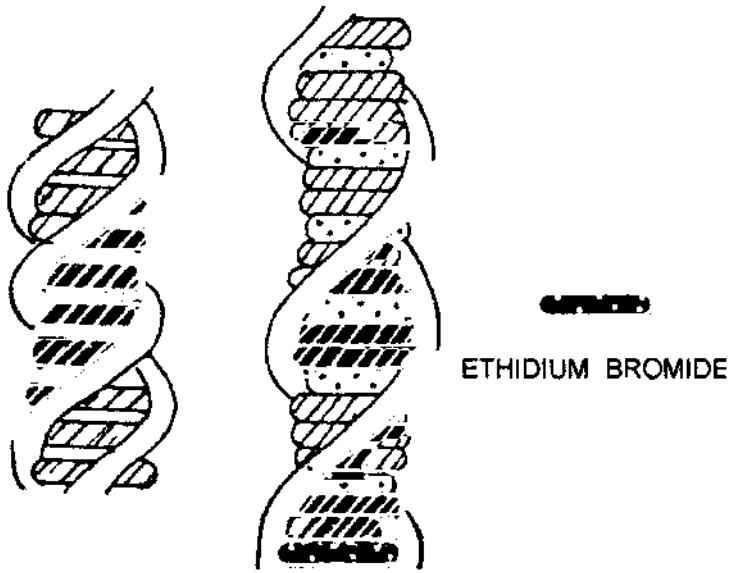



Fig.6.4 Distortion of the structure of DNA due to intercalation of ethidium bromide. As discussed in the text, aberrant DNA structure leads to the insertion of wrong bases during the synthesis.

undesirable mutations will be introduced in DNA. This change (mutation) is expected to sustain during subsequent cell divisions, ultimately leading to cancer.

It has already been mentioned that cancer-producing viruses in animals were known for quite some time. Generally these are referred to as tumor viruses. Some of the tumor-producing viruses contain RNA instead of DNA as their genomes. Those are known as retroviruses. Most retroviruses do not kill their host cells. Some, however, have an additional gene that can cause the cells to become cancerous. It was a dilemma, how RNA can cause tumor in the animals which have DNA as the genetic material. The belief was found to be true later as these bring to the hosts some foreign genes responsible for the formation of tumor. How is this possible? As early as 1962, that is two years following the discovery of DNA-dependent RNA polymerase (Chapter 3), Howard Temin of the University of Wisconsin predicted that there must be an enzyme to synthesize DNA on the RNA template. Most of the molecular biologists did not believe that because Crick's central dogma DNA→RNA→ Protein was deep-rooted in their mind. How can RNA produce DNA? Message flow must be unidirectional. But in 1970 Temin himself and David Baltimore independently discovered RNA-dependent DNA polymerase which catalyses the synthesis of DNA on RNA template. They shared the Nobel Prize in physiology or medicine in 1975 for this achievement. That discovery solved the key problem, how do the RNA-viruses carry the genetic information for the host in the form of RNA and later how the message is converted to DNA. These viruses are called retroviruses and the enzyme was therefore called a retro enzyme as it acts in the reverse direction ('retro' means 'backward'). These retroviruses cause not only cancer but other diseases as well, for example, AIDS (acquired immuno deficiency syndrome) which we know as another deadly disease caused by retrovirus. RNA viruses of animals and plants after entering their hosts (unlike bacterial viruses, the intact viruses enter) copy their RNA

genome in the form of DNA with the help of retro polymerase. The DNA may express itself or get integrated into the host genome and remain dormant temporarily. It does express itself whenever the opportunity arises. This dormant state known as lysogenic state (as in case of temperate DNA-containing bacterial viruses) is important for their opportunistic living. The discovery of the RNA-dependant DNA polymerase led to the revision of the central dogma by Crick himself who put a double-headed arrow (as shown below) between RNA and DNA indicating that message can flow in either direction. This is known as revised central dogma.

$$\text{DNA} \rightleftharpoons \text{RNA} \rightarrow \text{Protein}$$

The reverse process is naturally called reverse transcription. However, it looks like, there is very little opportunity of putting a reverse arrow between RNA and protein. The question of prion, a protein responsible for transmissible encepalopathies (brain diseases) where protein seems to be the genetic material remained open for quite sometime. We know now that prion has no so-called genetic information.

The study of the viral genes responsible for production of cancer led to the discovery of 'Oncogenes.' The word 'Onco' means 'Cancer'. Oncology is thus the discipline involved in the study of cancer[4]. The genes present in a virus responsible for the production of cancer became widely known as "Oncogenes". Some of the oncogenes are related to one another and thus belong to one family, for example, *ras* family of oncogenes causing *Rous* sarcoma in mouse. *Rous* sarcoma virus was named after Peyton Rous who studied chicken tumors. These oncogenes were first discovered in *Harvey and Kirsten* murine sarcoma viruses. Later, they were shown to be present in a slightly variant form (designated as protooncogenes) in mammalian cells as well. Factually, protooncogenes are counterpart of oncogenes in normal cells. If the former are mutated to oncogenes under certain circumstances or the mutated form (oncogene) is brought in through

a virus, cancer is induced. Thus protooncogenes and oncogenes are like two sides of a coin. On the basis of this it is widely believed (rather well established) that all of us carry the cancer genes (protooncogenes) which are not only harmless but are required for normal functions of the body. In case they are mutated to oncogenes due to internal or external factors carcinoma ensues. It is generally agreed upon that our body has surveillance mechanism to detect the aberrant genes and either correct them or destroy them as early as possible. However, if the mechanism of surveillance and destruction of aberrant genes fail, cancer is induced. These findings have led to the extensive investigations into the structure and function of *ras* genes. At least three genes (*Ha-, Ki- and N-ras*) have been identified in mammals and found to code for the closely related proteins. It has been further shown that organisms as widely divergent as *Drosophila* (fruitfly), *Dictyostelium* (mold) and *Sachharomyces* (yeast) possess similar genes. Apparently these genes are ubiquitous.

Let us divert a little before proceeding further. As already emphasized, the key to understanding the onset of cancer is to know how the perfectly regulated growth of normal cells becomes unregulated. Naturally, the mechanism of regulation has to be known. The cells communicate with the outside world with the help of receptors embedded in their membrane as well as present in cytosol. These are mostly proteins and some of them are trans-membrane proteins. They transverse the lipid bilayers of the cell membrane and thus remain anchored therein. Their extracellular part binds to the ligands which are growth factors, hormones, neurotransmitters etc, in a specific manner and transmit the signal inside the cell with the help of intracellular component. The cell acts in a very specific way in response to the signal received from outside. Naturally, the receptor molecules have very specific recognition sites to bind to specific ligands. In photosynthetic tissues light is the specific signal, which acts through the photoreceptors. The responses to light, hormones

etc., are in the form of a cascade of chemical reactions. Several pathways through cyclic AMP, inositol etc., are primarily responsible for initiating these reactions and naturally they are branded as second messengers. The receptors can be of various types but two of them are very important. These are G-protein- linked receptors and the growth factor receptors that signal through protein kinase-mediated phosphorylation. It should be pointed out at this stage that the protein phosphorylation and dephosphorylation are two important events in such regulation. All types of regulators act practically as switches and turn on and off the cascade of reactions involved in regulation, as and when it is required. Protein kinases are responsible for phosphorylating proteins. There are numerous such kinases and each one is usually very specific for its substrate which is called the target protein. The addition of phosphate as catalysed by kinases is usually at the serine or threonine or tyrosine residue. The protein thus becomes activated and carries out its function. This is practically 'switching on' of the system. Similarly, the phosphatases are the enzymes, which dephosphorylate the proteins and make them inactive, the process being known as 'switching off'. It is interesting to recall at this stage that some synthetic chemicals like phorbol esters act as tumor promoters. These compounds activate a specific protein, protein kinase C. Those mimic cellular diacylglycerol as secondary messenges. However, due to their very slow metabolism unlike diacylglycerol the 'switch on' situation remains constantly on and thus promotes tumors.

Let us now come back to oncogene products. These are grouped under G-family of proteins. There are a wide variety of proteins carrying out different types of functions but come under the same family (G-family) due to their binding to and requirement of GTP for their functions. These proteins have inherent GTPase activity which leads to the hydrolysis of GTP to GDP and $P_i$ (Fig.6.5). Polymerisation of tubulin to microtubules (network for segregation

of chromosomes during cell division) requires GTP. We have already discussed (Chapter. 2) that GTP is required for protein synthesis and EF-Tu as well as EF-G are GTP-binding proteins. Both have inherent GTPase activity and hydrolyse GTP to carry out their functions.

The products of *ras* genes may be distant relatives of G proteins (Fig.6.5). Like their counterparts *ras* gene products are membrane-bound proteins. Oncogenic mutations also do not change their association with the plasma membrane. The ras gene codes for p21 protein, so called as it has molecular weight of 21 KD. It binds GTP and is present in small amount in normal cells. p21 protein has very low GTPase activity and hydrolyses GTP to GDP at an extremely slow rate. Thus in many respects p21 protein is analogous to G

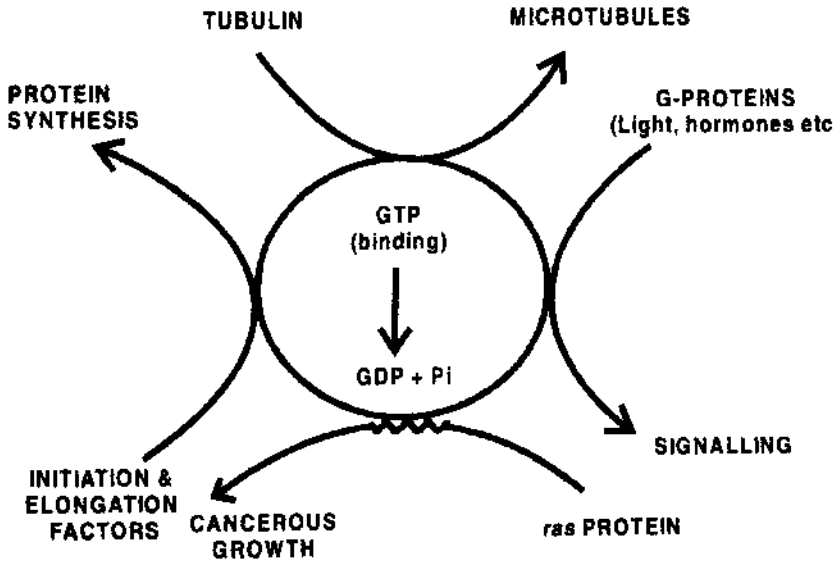



Fig. 6.5 G family of proteins carrying out various types of functions. These have the common property of binding GTP and also have an inherent GTPase (GTP hydrolysing) activity which is responsible for their functions.

proteins. However, due to the slow rate of hydrolysis once the process is started it is not stopped resulting in uncontrolled growth.

Transmembrane G proteins are rather complex in nature and so far it has not been possible to determine their three-dimensional structure by X-ray crystallography. However, *ras* protein is of a comparatively smaller size but has sequence similarity to G proteins, in general, specially with respect to GTP-binding domain. Since the crystal structure of *ras* protein has been determined and the structure of its GTP binding domain is known (Plate 11 Fig 6.6 ), it has been possible to deduce the GTP-binding domain of G proteins, in general. As per model building this domain (shown in the figure) undergoes conformational change during hydrolysis of GTP to GDP and Pi and thus acts as the switch regulating the activity of these proteins.

As already mentioned, a large number of hormones like insulin, glucagon, epinephrine etc., are essential elements for the maintenance, replication and growth of the eukaryotic cells. These are usually required in small amounts and normally do not enter the cells. There are receptors for hormones and many other signalling agents, which recognize them specifically and send the signals inside the cells with the help of proteins leading to a series of metabolic reactions. As mentioned earlier, some of them act as trigger for cyclic AMP production. The latter acts as the messenger initiating the series of events which finally lead to the enhancement of specific metabolic processes linked to the primary functions of the cells. Even light is capable of triggering similar events ending in the production of sugar from carbon dioxide and water. Calcium ion ($Ca^{++}$) also acts as a messenger and is known as second messenger. Other types of messengers are known as well. These processes of transmission of the signal from the outside to the inside of the cell is generally known as signal transduction. All the hormones do not necessarily act in the same way, for example, steroid and thyroid hormones bind to some serum proteins and are carried directly to the target tissue. Those

**Table 6.1**

***Ras* and *Ras*-like Gene Products**

| Source | Product | No. of Amino Acids |
|---|---|---|
| **Retrovirus Oncogenes (ras genes)** | | |
| *Harvey* Sarcoma | P21 proteins | 185-189 |
| *Kirsten* Sarcoma | ,, | ,, |
| *Neuroplasma* | ,, | ,, |
| *Human Brain* | ? | 380 |
| ***Saccharomyces* (Yeast I)** | | |
| *RAS I* gene | SCI | 300 |
| *RAS II* gene | SCII | 322 |
| *YPP2* (or *yPTI* gene) | YPT proteins | 206 |
| ***Escherichia coli*** | | |
| *Lep A* gene | Lep A proteins | 598 |
| *Era* gene | Era protein | 310 |

cross the cell membrane barrier and bind to specific receptor proteins of the nucleus and exert their effect at the genetic level. In this way, those directly stimulate the transcription of certain messages leading to the enhancement of production of some proteins including enzymes. Some of the toxins responsible for the deleterious action of pathogenic bacteria like *Cholera bacilli* act through the transmission of wrong signal from outside, for example, release of water from the gut leading to severe diarrhea.The Table 6.1 enlists a few *Ras* and *Ras*-like gene products and their sources. It is clear from the data that even bacteria and yeast contain similar products like *ras* gene products of viruses.

Another breakthrough in cancer research occurred with the discovery of tumor suppressor genes[5]. In those days, cell fusion

technique was already available. It was possible to fuse different types of cells to give rise to hybrid cells. In one such experiment normal cells were fused with transformed (cancerous) ones and it was expected that the hybrid cells so formed would be malignant. When a cell becomes malignant its characteristics change significantly. As already mentioned, its growth characteristics also change drastically. Normal cells ordinarily grow on the surface in mono layers due to contact inhibition whereas transformed cells grow randomly in multiple layers, one over the other as the contact inhibition is lost. From their shape and size it can be surmised that the cells have been transformed. Surprisingly, hybrid cells formed by the fusion of normal and malignant cells behaved normally contrary to the expectation (Fig.6.7). It was further observed that if the hybrid cells are propagated through several generations, those sometimes are reconverted to transformed cells. It was already known that hybrid cells sometimes lose one or more chromosomes on repeated culture. So it was suspected that there are some gene or genes present in normal cells, which suppress tumor formation. The loss of such gene(s) in course of repeated culture might have been responsible for to regain malignancy. This suspicion was strengthened by the observation that different types of tumor suppression are linked to different chromosomes, or in other words, the phenomenon must be gene-associated.

Eventually, quite a number of tumor suppressor genes were identified. Retinoblastoma gene was one of the earliest and most extensively studied. Most of these genes were connected with the modification of transcription factors, as discussed earlier. One of the best examples of regulatory proteins is the family of signalling proteins, which belong to the G family, as already discussed. They act as switches, in association with GTP they become active and signal the hormones etc., to switch on the metabolic pathway as a cascade reaction. Activation results in the hydrolysis of GTP and switching off the pathway. This is extremely important for the normal cells. If

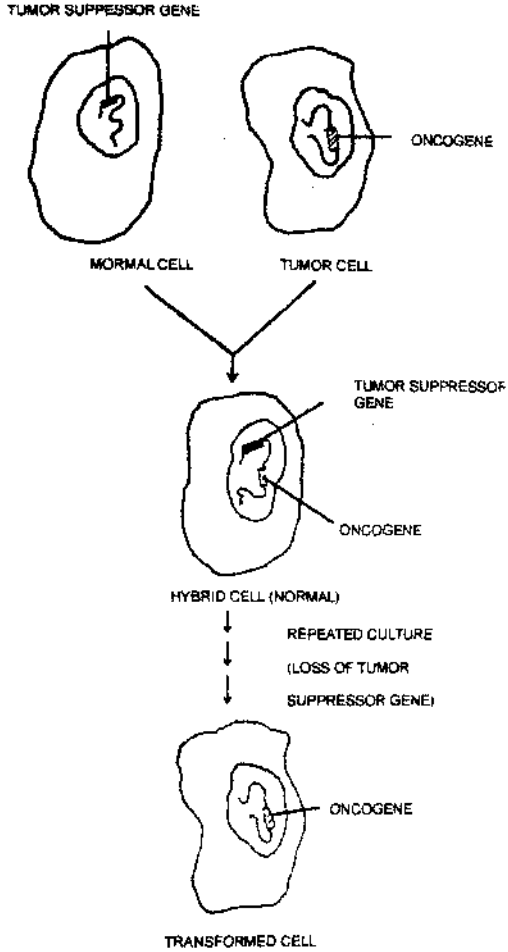



Fig. 6.7 Experiment to demonstrate the presence of tumor suppressor gene in normal cell. When the normal cells and tumor cells are cultured together they give rise to somatic cell hybrids, which behave normally showing loss of the tumor producing genes. But on repeated culture the hybrid cells are reconverted to tumor cells. Actually tumor suppressor genes present in normal cells suppress the expression of tumor genes. During repeated culture suppressor genes are lost. Therefore the tumor genes start expressing again.

there is mutation due to which GTP is not hydrolysed the signaling will continue in an unwarranted fashion, resulting in aberrant growth. The same is true for transcription factors, which are also turned on and off as per the need of the cell. Actually phosphorylation and dephosphorylation of these proteins form the main regulatory mechanism.

Another group of proteins known as cell division factors are also involved in such regulatory action[4]. Each and every cell has the capacity to grow and divide when the necessity arises and stops cell division as soon as proper signal (like contact inhibition) is obtained. As repeatedly emphasised, unregulated cell division is branded as cancer. It has already been pointed out that several cell division factors including some of the transcription factors control the cell cycle as depicted earlier in Fig 6.2. During G1 phase, the cell starts preparing for cell division. This is followed by the preparation for DNA synthesis which takes place in S phase. After the DNA synthesis, the cell has to get ready (G2 phase) for mitotic division for passing on the chromosomes to two daughter cells following division, which takes place during the M phase. Several cell division factors like cyclins and Cdks are involved in different phases. These are mostly phosphorylating and dephosphorylating proteins and act in unison with the transcription factors. The retinoblastoma (Rb) gene is a tumor suppressor gene. Loss of this function leads to tumor genesis. There is a long list of such tumor suppressor genes. The retinoblastoma protein production by tumor suppressor gene plays an important role in different stages depending on its state of phosphorylation, one of its important functions is to promote the genes to be active in S phase. As far as cancer induction is concerned, $G_1$ pause (see below) is important for the correction of errors in DNA. This may be treated as the surveillance mechanism not to allow the cell to become cancerous. If the pause is shortened cancerous situation has a better chance to thrive. It should be mentioned in this connection that p53, a DNA-binding transcription factor has been identified as a tumor suppressor protein. During the normal cell division cycle (Fig 6.2) there are pauses at each and every stage (G1, S, G2) which act as check points when damages to DNA are checked and repaired. There are two mechanisms for this, one is p53-independent, the other is p53-dependent. The latter results in removal of damaged cells from the population. A failure to do this will lead to the formation of cancerous cells. The transcription

factor p53 is one of the cells'main defense against tumorogenesis, carrying out many functions to stop cells becoming cancerous including halting cell duplication, repairing damaged DNA and triggering apoptosis (cell death).

The final evidence for the existence of tumor suppressor genes came from the study of the cancer cases, which were thought to be hereditary. One of the best examples is breast cancer. For long time it was suspected from family studies that it is a heredity-linked disease. The first breast cancer gene was identified on chromosome 13 and designated as BRCA2. The locations of some tumor genes on different chromosomes are shown in Table 6.2. This information initially helped in human genome sequence program, an international collaborative venture to map and sequence all the human genes located on 23 pairs of chromosomes composed of 3 billion base pairs. Eventually, the locations of quite a few cancer genes were determined.

**Table 6.2**

**Location of Tumor Genes on Chromosomes**

| Genes | Chromosomes | Locations* | Functions |
|---|---|---|---|
| MSH | 2 | p15-18 | Familial, Colon, Cancer |
| SCLC | 3 | p21.1-21.3 | Lung, Cancer |
| CDKN2 | 9 | p21 | Malignant, Melanoma |
| MENZA | 10 | p11.2 | Multiple, Endocrine, (type 2A) |
| H-RAS | 11 | p15 | RAS Oncogene |
| BRCA2 | 12 | p14 | Retino-Blastoma |
| BRCA1 | 17 | p21 | Breast Cancer |

* p represents the longer arm of chromosomes from the centromere (central constricted part of chromosome), the number indicates arbitrary posititon of the gene from the centre.

Although a large number of weapons[6] (Fig. 6.8) are now available in the arsenal for the war against cancer it still remains as one of the most threatening diseases of today. Surgical removal of the affected organ (like breast) at the early stage before the start of metastasis (spread of cancer to other organs of the body) is one of the effective procedures. In some cases, this is not effective as metastasis starts quite early in them. Radiation treatment (radiotherapy) is one of the efficient ways where the sole objective is to destroy the cancer cells. However, the normal cells surrounding the cancerous ones may become cancerous due to excessive radiation exposure (in spite of the special protective measures taken by the radiologists). The third treatment which is extensively followed at the early and late stages is chemotherapy. The classical example of fluorouracil has already been mentioned. Aminopterin is a valuable drug in the treatment of acute leukemia (blood cancer) and choriocarcinoma (originating at the placental site during pregnancy). This is an antifolate drug. Folic acid (tetrahydrofolate form) is essential for the synthesis of thymine (essential component of DNA) from uracil. These drugs are thus expected to interfere with the synthesis of DNA of the carcinoma cells. However, excessive use interferes with the synthesis of DNA in the normal cells as well. Several such cancer drugs like mitomycin C block DNA synthesis of not only cancer cells but also of the normal cells. So the choice of dose in chemotherapy is extremely important. The use of higher doses is now possible with the help of gene therapy. Sometimes combination therapy of radiation and drugs is also used even after surgical operation. In a few cases simple treatments are sometimes dramatically effective, in other cases not. In general, it can be said that longevity of the cancer patients has been increased significantly but we are still far from finding a method which will guarantee the eradication of cancer. On the basis of spectacular success in human genome sequencing and the concomitant wealth of information revealed at the molecular level through varied approaches, it has been possible to develop ways to tackle the deadly disease. One of them is

immunotherapy (Fig. 6.8). The enhanced antigenic capacity of tumors is being utilized to prepare antibodies against tumors. The monoclonal antibodies, which are highly specific, are naturally the best choice. Similarly, the use of antisense oligonudeotides which associate with specific messenger RNAs responsible for the continuous production of cell division factors and silencing them is another new approach. The development of newer vaccines against tumors is also in progress in various laboratories. Although gene therapy is at an early stage of development, the cancer specialists are looking upon it as a futuristic approach of great potentiality. One specific approach may be cited

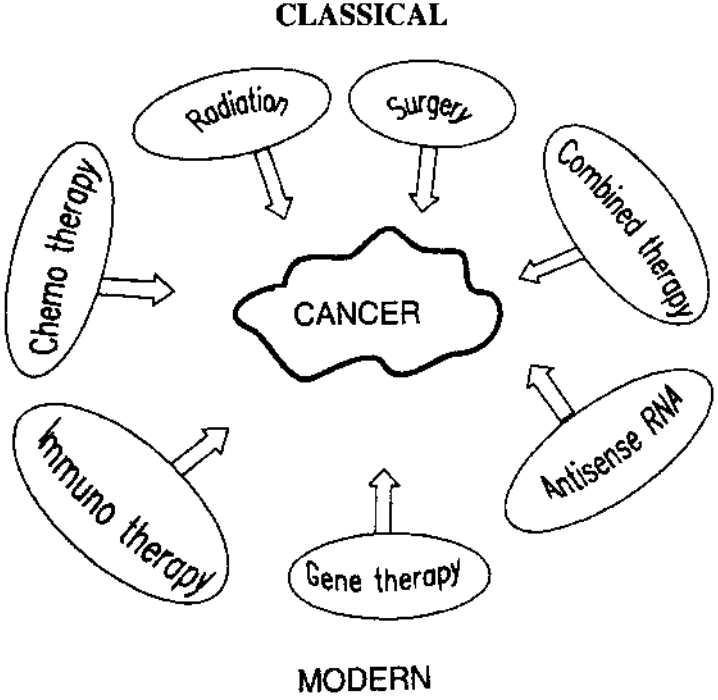



Fig. 6.8 Various approaches to cancer therapy. The details have been described in the text.

here. Bacterial genes known as *mdr* (multiple drug resistance) do not allow the foreign materials (like drug) to enter the cells. The transfer of *mdr* genes through bone marrow cells prevents the entry of the chemotherapeutic drugs to the normal cells and thus permit the use of higher doses to destroy the tumor cells. Another interesting but entirely different approach is worth mentioning here. The products of *ras* oncogenes are attached to isoprenyl group through the SH group of cysteine. These groups help the *ras* proteins to anchor to the membrane and exert their activity. The transforming property of the protein is lost on blocking the isoprenyl group. So this could be another novel way of preventing some types of cancer. Unfortunately, the variety of efforts made so far extend the life of the patients, but the ultimate cure is still a dream. However, tremendous amount of knowledge has been acquired in the area, yet a much more intensive as well as extensive approach has to be made to reach the final goal which seems a long way ahead. As already mentioned, the dislodging of the cells that are capable of metastasis, from a benign tumour leads to neoplasia. If the mechanism of metastsis becomes known and can be counteracted neoplasia could be nipped in the bud.

It is true that the acquired knowledge about the mechanism of transformation of a normal cell to a cancerous one will help us to understand the mechanism of regulation of the biochemical pathways dictated by the regulatory genes in the normal cell. Molecular biology has been enriched to a great extent from such studies on oncology. That was mentioned by St. Gyorgi in one of his lectures, which I had the opportunity to attend. He, however, said 'Cancer helps lot more people than it kills.' The hint was naturally at the huge amount of funding available to work in the cancer area and the benefit to a large number of scientists, clinicians, technicians etc., engaged in cancer research. A good number of philanthropists have created trusts for cancer research and donated liberal amount of fund to the universities and institutions. Public funds are also available through liberal grants.

That has not only increased our knowledge about the physiological state of our health but also provided new approaches to combat cancer though most of it remains unconquered. Considerable effort has been made and lot of public money has already been spent on cancer research and more is expected to be done in the future. Therefore, researches, evaluation studies, clincial diagnosis, therapy etc., are feeding a large number of persons. Scientists in general, clinicians in particular as well as diagnostic experts earn their livelihood through cancer research.

A final word about 'Cancer' may not be quite palatable at the end. All of us are potentially carrying cancer genes (protooncogenes) which are essential for normal physiology and we are thus prone to cancer. It is absolutely clear from the above discussion that cancer is primarily due to various types of damages done to DNA resulting in the conversion of protooncogenus to oncogenes. However, our body has also a good number of DNA repair mechanisms including DNA recombination, which correct the copying errors (during DNA synthesis) as well as other damages done. Surveillance is constantly carried out by the enzymes to find out and correct the errors. A lot of us escape the horrors of cancer due to the surveillance mechanism. However, our environment is being increasingly polluted with various types of carcinogens including the use of radioactive materials which are the products of civilisation leading to an endangered society. The bulk of error is constantly increasing and nature is not expected to cope with the increasing burden. Until and unless we are conscious about it and do something to avoid the calamity any amount of funding on cancer research will be miserably inadequate to solve the problem. Therefore we should be prepared for a 'doomed' future for our innocent children unless we mend our present day living conditions.

## References

1. *Cancer surveys*, Imperial Cancer Research Fund, Cold Spring Harbor Laboratory Press (Yearly).
2. *The Biological basis of cancer*. Ed. R.G. McKinnell, R. A. Perchment, A.O. Perantoni and G, B. Pierce, Cambridge University Press (1998).
3. *Cancer: The evolutionary legacy*. Mel Greaves, Oxford University Press, (2000)
4. *Molecular Oncology*. Ed. J. M. Bishop and R. A. Weinberg, Scientific American Incorporated, NewYork (1996).
5. *Insight on 'Cancer'*, *Nature* Volume 411, 335-395 (2001).
6. *Cancer Medicine*. Volumes I & 2 Ed. J. F. Harland, E. Frei, R. C. Bast, D. W. Kufe, D. L. Morton and R. R. Weicheselbaum, Lea and Febiger, London (1993).

*****

# Chapter - 7
# Darwin's Flute
## (Theory of Evolution : Its Molecular Aspects)

*Charles Darwin as well as Alfred Russel Wallace were not aware of DNA and the Mendelian principles even though both were discovered during their lifetime. In fact, whatever has been discussed so far in the last six chapters were completely unknown to them. But the theory of evolution as conceived by Darwin and presented in the 'Origin of Species'[1] in an abbreviated form is the gospel truth and has been thoroughly vindicated by the knowledge acquired subsequent to his death. Life has the music but Darwin played the flute, the tune of which matched with that of life. Attempt has been made in this chapter to narrate the history of his realisation of the origin of one species from another. However, the main object of this last chapter is to show how DNA has been used as a molecular marker to strengthen what has been learnt through the painstaking efforts of the classical biologists. Darwin's flute plays the music of life through Molecular Biology.*

*"Nothing in biology makes sense except in the light of evolution"*
***-Theodosius Dobzhansky***

'The HMS Beagle'[2]was a ten-gun brig (Plate 12, Fig 7.1a) which was commissioned by the British Admiralty on a surveying voyage around the world. Its captain Fritz Roy was looking for a naturalist to be recruited as a member of the team to take advantage of the trip to study 'nature' in all its aspects, during the expedition. Here comes 22

year old Charles Darwin, a geologist by training, under the internationally known geologist Sir Charles Lyell. Fritz Roy was pleased to appoint him. Without any prior knowledge he was sowing the seed of the greatest discovery of the nineteenth century in the field of biology. But the most regrettable incident was the suicide of Fritz Roy as he thought that he unknowingly gave shelter to Darwin during the voyage and thus developed intimate friendship with him. He was a strong believer in God and the Bible and therefore could not tolerate Darwin's anti-religion theory. *The Beagle* left the shore of England through Davenport on December 27, 1831 and took almost 5 years to complete the trip through hundreds of ordeal and returned to Felmouth on October 2, 1836. It visited Brazil, Argentina, Tierra del Fuego, Chile, Peru, Tahiti, the Galapagos Islands, New Zealand, Australia and other countries and islands on the way (Plate 12, Fig.7.1b). The geologist turned naturalist had plenty of opportunity to satisfy his curiosity about the origin of various species of plants and animals. He satisfied himself undoubtedly but he was not sure as to whether everyone else would be convinced. That was the reason why he waited for almost 30 years to make his views public which led to the ' Origin of species' in 1859. Even at that time he was not prepared to publish voluminous findings of his painstaking research. When he became aware of similar views of Alfred Russel Wallace he took the message as his defeat and was reluctant to publish his treatise. But eventually he was persuaded by Lyell as well as the famous botanist Joseph Hooker to present the work on the same platform with Wallace at the Linnean Society. None of them, however, personally presented the work. Wallace respected him a lot and agreed with him in many situations but had some differences of opinion in specific cases, one of the examples being sexual selection. The spirit of science, which was imbibed in Darwin, had hardly any parallel and his calmness was a clear sign of that especially when the clergymen repeatedly provoked him. It was a lesson to be learnt not

only by the scientists but also the human society as a whole. The fights, quarrels and the untoward incidents around great discoveries, which are made today, are the eye-openers and reflect the uglier side of the human character. However, good sense must prevail eventually.

Charles Darwin had both insight and foresight to present such a revolutionary concept on 'evolution' in his time. His '*Origin of* Species', a masterpiece in biology can be compared with '*Principia*' of Newton in physical sciences. What Darwin prophetically visualised was the mode of origin of different species of plants and animals. Even in a single species each individual develops a specific morphological feature. Some time, one or more of those help in the survival as one feature may be more helpful than the other. A woodpecker with a stronger and sharper beak will survive better than the other with a weak and blunt one. The survival of the fittest is the general rule. A finch with a convenient shape of beak, which is helpful in cracking the type of nut available in the area, would survive better. A flower with an attractive colour will invite more insects, which would help in its pollination. In the competition, a flower with somewhat less bright colour would be handicapped. This type of competition between different members of the same species would lead to the survival of the fittest. Changes in morphological as well as other features are taking place constantly as a natural process, the selection is also going on simultaneously. This is positive selection (outcome of natural selection) according to Darwin. Thus even a single species may have different varieties. This process is however, slow; in fact extremely slow. When a number of changes take place in a particular species a new species with different features may evolve over a long time which may sometimes spread over millions of years. Darwin could rightly visualise that the large number of species of plants and animals we see today are the results of natural selection and evolution of one species from another. A new species cannot be created abruptly, however powerful and resourceful the source could

be. But it cannot be denied that today natural selection can be mimicked in a test tube putting selection pressure as in the case of ribozymes (Chapter 4).

Natural selection, survival of the fittest, sexual selection etc., are the basics of Darwinian theory but Darwin was ignorant of genetics, a subject which developed almost half a century later, no question of molecular biology which was born almost 100 years after that. Incidentally the foundation of Mendelian genetics was laid during his life time but he was not aware of it. Let us forget about sexual selection which raised bitter controversy between Darwin and Wallace and by and large has remained controversial even today. But natural selection and the survival of the fittest have been accepted even by some anti-Darwinians as well. That one species has been derived from another species by a very slow change, spread over millions of years, is the backbone of Darwinism. While Darwin was still alive Gregor Mendel, the Austrian Monk, was crossing his garden peas and sowing the seed of genetics and thereby establishing the hereditary principles (Chapter 1). Unfortunately, he died unknown not only to Darwin but also to the main stream biologists. Others took almost 35 years or so to rediscover his work, which eventually provided the solid foundation of the Darwinian theory of the origin of the species.

In retrospect it appears that Darwin's theory of evolution of species is the most important discovery of life sciences followed by the determination of the structure of DNA by Watson and Crick in 1953. However, life sciences had a severe blow due to the negligence of the fantastic experiments of Mendel. Luckily they were rediscovered, although after a long time. Mendel himself was absolutely certain of the principles of hereditary transmission of characters which he laid down. He must have died as a very disappointed man as the world did not take cognizance of such an astonishing discovery. Although there was tremendous onslaught on Darwin's theory when it was first propounded, it was fully recognised during his life time. In spite of all

the preliminary setbacks coupled with his ill health he must have died a fully satisfied person. But he himself would have been amazed to see the far-reaching consequences of his fantastic theory today. History was cruel to Mendel but not to Darwin.

One had to wait for almost 100 years to see the dawn of the molecular basis of evolution and that came only when DNA was finally established as the genetic material. Historically speaking, Miescher while isolating 'nuclein' in 1868 (within Darwin's lifetime) was not aware that it was the 'genetic material' although he believed that the newly isolated material resided in the nucleus which was known as the central organelle and recognised as the nerve centre of the cell. Only when Crick and Watson conceived the structure of the gene in 1953 did the matter more or less settle at the molecular level. However, that DNA is the hereditary material was already established through the famous experiments of Avery, McLeod and McCarty on the one hand and Hershey and Chase on the other. These historical events have already been recorded earlier and will be written in golden letters for several millennia to come.

But the knowledge about the basic structure of DNA was not enough to make it the tool to study the molecular basis of evolution. As DNA was characterised more and more both at the physical and chemical levels phylogenetic approaches were made to correlate the physio-chemical properties of DNA with the origin of species. Just like the phenotypic characters, other aspects of different species such as geographical distribution, ecological background etc. were the powerful markers for the study of evolution. The characteristic features of DNA were also turning out as the molecular markers of evolution.

Before we proceed further it is better to pause for a while to understand phylogenetic relationship through the construction of phylogenetic trees. The latter show the relationship between different species, whether those are closely or distantly related, whether two or more species have common ancestors etc. A typical example is shown in (Fig 7.2). Several features are obvious from the tree. Bacteria (prokaryotes) are distinct from

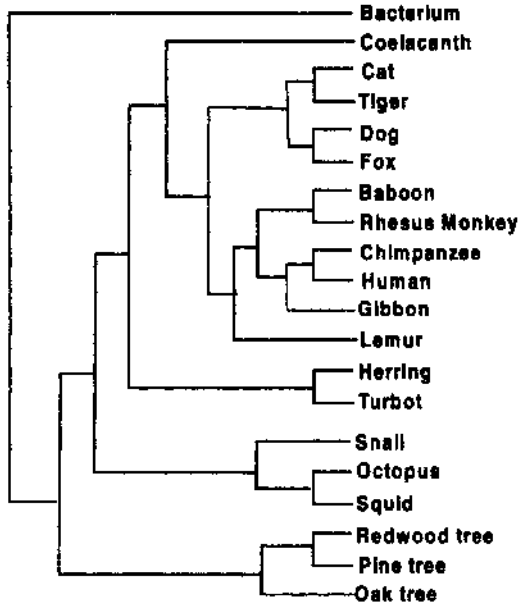



Fig. 7.2 Universal phylogenetic tree. The family tree represents the relationship between 20 different organisms to show whether they are closely related to one another or whether two or more organisms have common ancestors.

the plants and animals (eukaryotes). Plants form a separate group, animals do the same. Again the animals known as mammals (for obvious reasons) are clustered together, fish forms a separate cluster and so on and so forth. If we consider a pair, like cat and tiger, they are very close and have a common ancestor. Humans and chimpanzee are closest to one another.

The concept of the tree was originally from Darwin and it depicted in 'Origin of species' as shown in (Plate 13, Fig.7.3). It will be worth quoting from Darwin himself about the tree as shown in the *Origin of Species*[1]

*'The affinities of all the beings of the same class have some times been represented by a great tree. I believe this simile largely speaks the truth. The green and budding twigs may represent existing species; and those produced during each former year may represent the long succession of extinct species. At each period of growth all the growing twigs have tried to branch out on* all sides, *and to overtop and kill the surrounding twigs and branches, in the same manner as species and groups of species have tried to overmaster other species in the great battle for life. The limbs divided into great branches, and these into lesser and lesser branches, were themselves once, when the tree was small, budding twigs; and this connexion of the former and present buds by ramifying branches may well represent the classification of all extinct and living species in groups subordinate to groups. Of the many twigs which flourished when the tree was a mere bush, only two or three, now grown into great branches, yet survive and bear all the other branches; so with the species which lived during long-past geological periods, very few now have living and modified descendants. From the first growth of the tree, many a limb and branch has decayed and dropped off; and these lost branches of various sizes may represent those whole orders, families, and genera which have now no living representatives; and which are known to us only from having been found in a fossil state. As we here and there see a thin straggling branch springing from a fork low down in a tree, and which by some chance has been favoured and is still alive on its summit, so we occasionally see an animal like the Ornithorhynchus or Lepidosiren, which in some small degree connects by its affinities two large branches of life, and which has apparently been saved from fatal competition by having inhabited a protected station. As buds give rise by growth to fresh buds, and these, if vigorous, branch out and overtop on all*

*sides many a feebler branch, so by generation I believe it has been with the great Tree of Life, which fills with its dead and broken branches the crust of the earth, and covers the surface with its ever branching and beautiful ramifications'.*

It may be mentioned here that the kind of tree, which Darwin drew to explain the origin of a variety of species, is the first phylogenetic tree (Plate 13, Fig.7.3). When the protein sequencing method developed the sequences of proteins like haemoglobin, cytochrome C etc., were used to build the phylogenetic trees. The gene sequences were also derived indirectly from the genetic code. However, when the DNA sequencing method developed the sequences could be directly used as molecular markers. Molecular evolutionists are deeply indebted to Frederick Sanger who for the first time sequenced a protein and subsequently DNA. It was no wonder that the Nobel Prize was first awarded to him for doing the Herculean job of sequencing a small protein like insulin which took him almost 10 years to complete. He was awarded the second Nobel Prize for sequencing of DNA. Both the methods later became routine in biology due to his pioneering work. This has already been discussed in detail earlier. However, the use of proteins as molecular markers of evolution came from the seed experiment of Pauling on sickle cell haemoglobin followed by that of Ingram who showed by finger printing of peptides the difference between a normal hemoglobin (adult hemoglobin) and sickle cell hemoglobin.

The protein molecules, which were used earlier as molecular markers of evolution, are haemoglobin, cytochromes, enzymes like carbonic anhydrase etc. Changes in amino acid sequences were used to follow the mutation in different species. A typical example is shown in Fig.7.4. The data have been taken from a publication of W.M. Fitch and E. Margolish. The authors used cytochrome C as the molecular marker. The amino acid sequences of cytochrome C from a number of organisms starting from yeast (like *Candida, Saccharomyces* etc.,)

to man were compared. The amino acids which were replaced with other amino acids during the evolution of different organisms were identified. It was calculated from the genetic codes (Chapter 3) how many mutations in the DNA were involved in producing these changes. The mutational data were used to produce a matrix to find out the genetic distances between the different organisms. These distances in turn were used to build the phylogenetic tree as shown in Fig. 7.4. The trees built in this way were in good agreement with the trees constructed with morphological markers although there were controversies in some cases.

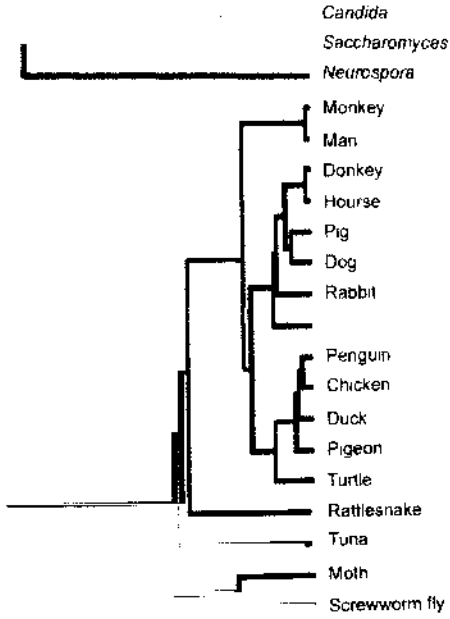



Fig. 7.4 An early phylogenetic tree constructed on the basis of molecular data. The tree has been constructed by the Phylip programme from the calculated nucleotide replacement data of Fitch and Margolish (*Science*, vol. 155, p279, 1967) on the basis of derived gene sequences of cytochrome C present in several organisms.

This is, however, an indirect reflection of genetic changes. The replacement of an amino acid is due to change in the codon and that, in turn, may be due to the changes in one or more of the three bases (of the codon). It should be remembered that an amino acid may have more than one code, so in some cases the change of a base (specially the third one known as 'Wobble` base) may not affect it at all. But sometimes, when only one letter of the codon changes, one amino acid is replaced with another but the function of the protein may remain unaltered if the particular change does not affect its function. The change of one acidic amino acid with another acidic one or a hydrophobic amino acid with another hydrophobic one may not necessarily result in the change of function. This is referred to as 'neutral evolution`, which plays a significant role in Darwinian concept of evolution. This was first pointed out by M. Kimura who developed the neutral theory of evolution, which still remains controversial. However, the trees built with the help of protein sequences for example, cytochrome C are not far from the universal tree already built with the help of phenotypic characters. However, the molecular markers are much more effective in building the tree of closely related organisms. Such trees built with different protein markers agree with each other remarkably well. The term 'molecular clock' was coined in this connection indicating the rate of mutation (in terms of amino acid codon change) is fairly constant like that of a clock. This concept is being challenged these days because the rates may not be exactly the same in all cases. Sometimes those may differ quite significantly. Until and unless the difference in rates is taken into account, two trees built with different proteins may not tally with each other. This was a handicap in the case of establishment of the relationship among the three primary kingdoms eubacteria, archaebacteria and eukaryotes. The clock speeds for different proteins are shown in Table 7.1 It is clear from the table that the clock runs 600 times faster in case of fibrinopeptide in comparison to histone H4.

Before DNA sequencing methods became available attempt was made to use some characteristic features of DNA as molecular markers. For example, the base composition of DNA is reflected in its melting temperature depending on its GC content. However, there was limited scope of using such method. Sol Spiegelman's method of hybridisation between different DNAs also came into practice. The extent of hybridisation between two types of DNA naturally reflects the extent of compositional similarity of DNAs of two different organisms. DNA of human and chimpanzee hybridise to the extent of 98% indicating that they are phylogenetically close. It was already known that the latter has one pair of chromosomes extra than the former having 23 pairs due to either fusion or fission of two chromosomes. Another technique came into practice after the restriction enzymes were discovered. Different restriction enzymes cut the DNA molecules at specific sites depending on the nucleotide sequences. If two DNA molecules have similar sequences the chances are bright that a number of restriction enzymes may produce DNA fragments of similar sizes

**Table 7.1**

**The rate of evolution of different protein genes**

| Proteins | $K_{aa}$* (Rate X $10^9$ yrs ) |
|---|---|
| Fibrinopeptide | 6.30 |
| Pancreatic RNase | 2.10 |
| Lysozyme | 2.00 |
| α-globin | 1.20 |
| Myoglobin | 0.89 |
| Insulin | 0.44 |
| Cytochrome C | 0.30 |
| Histone H 4 | 0.01 |

* the rate $K_{aa}$ is the probability of a single amino acid substitution per year.

(electrophoretic movement in a gel gives the idea of the sizes). So the extent of similarity can be judged from the gel pattern and this information can be used to build phylogenetic trees. There is no doubt that lot of preliminary information on phylogenetic relationship turned out through such studies. However, these methods were of limited use.

Although extensive work was carried out in the building of the phylogenetic trees with the help of protein markers the major breakthrough in determining the molecular basis of evolution took place with the introduction of the DNA sequencing method. The one, which is primarily used these days, is that introduced by the Master of protein sequencing, Frederick Sanger and known as dideoxy technique. It is no wonder that Sanger won the second Nobel Prize for this outstanding work and the DNA of his choice was that of a bacterial virus $\phi$x174. This is being repeatedly mentioned, as it is difficult to find a parallel example of this type of success. The other method known as chemical method was introduced by Maxam and Gilbert. Gilbert shared the Nobel Prize with Sanger in 1980 for this achievement. The latter method is known as chemical method as it depends on the specific reagents which are responsible for the cleavage of phosphodiester bonds between specific bases. Sanger's technique, however, is based on the enzymatic method coupled to the termination of DNA synthesis. The latter is now-a-days used extensively for certain obvious advantages, one of them being that the replicative form (RF) of the M13 DNA is a very good vector for cloning of genes and it is rather easy to sequence the single stranded mature phage DNA with the insert. The sequencing of double-stranded DNA is also no problem these days.

As the DNA sequencing method became almost routine in various laboratories, a large number of molecular biologists started sequencing various genes as well as total genomes. Prior to that physical mapping of DNA (the so-called restriction mapping) came into wide practice.

As already mentioned, DNA has its specific restriction site(s) and locations of the restriction site(s) help to characterise the DNA to a great extent and compare those. As mentioned earlier, attempt was made to use restriction enzymes for phylogenetic studies. It was rather easy to sequence restriction fragments due to their small sizes and build the sequence of the genome through the fitting in of the data into the structure of the whole genome. Thus physical mapping was of considerable help in sequencing, specially when the question of handling larger size genomes arose. At the critical juncture an exemplary international collaboration was introduced through the Gene Data Bank (including the Protein Data Bank as well). Whatever sequencing data became available in a particular laboratory those were made available through the data bank to the international community. That became a great help, especially in the studies related to evolution. With the availability of more and more data the study of evolution with DNA as molecular marker naturally proceeded at a rapid stride. However, the method of alignment of the sequences is rather complicated. Sequences may not perfectly align all the time. The first job is to align them by adjustment, creating gap/junctions by deleting or inserting bases. For better alignment intervening sequences between the aligned bases have to be sometimes ignored. An example is given below.

T T A G T C * G A A G T C
T T * G T C A G A A C T C

It has been possible to align the two sequences by creating one gap each in both the sequences. Such alignment is not easy by visual inspection. Naturally one has to depend on algorithms. Various computer-based programmes are available for the purpose. Once the proper alignment is done one can count the number of base changes between the two. Once the base changes have been determined several methods can be tried to derive the phylogenetic relationship. The two methods which are most generally used these days are known as

homology and distance method. In a rather simplistic way it can be said that the number of identical characters is a measure of homology and the nucleotide replacement counts indicate distance. A third method called parsimony is also extensively used. The literal meaning of parsimony is 'excessive economy or stringiness'. In a word, parsimony means in evolutionary language the minimum change that is to be made for the switch over from one species to another which will be explained below.

We are taking a simple example (from Kendrew's *Encyclopedia of Molecular Biology*[3]) to show how the phylogenetic tree is built from the DNA sequences. A case of 4 animals (monkey, horse, kangaroo and opossum) are presented below. As such we know kangaroo and opossum which belong to the class marsupials are close to one other. Parts of the sequences of the four animals are presented. Note that no gap/junction has to be created for the alignment of the sequences.

| | 10 | 20 |
|---|---|---|
| Monkey | TGAACTCAAGCACCAAAAAGGAAGACT | |
| Horse | TAGGCTCAAGCACCAACATGGCATACT | |
| Kangaroo | TAAGCCATAGCGACAAACTAGCCGTCT | |
| Opossum | TAAGCCATAGCGACAAACAAGCCTATT | |

**Table 7.2**

**Differences in Nucleotides**

| | Monkey | Horse | Kangaroo | Opossum |
|---|---|---|---|---|
| Monkey | 0 | | | |
| Horse | 7 | 0 | | |
| Kangaroo | 13 | 12 | 0 | |
| Opossum | 13 | 12 | 4 | 0 |

If we count the differences in nucleotides (A, T, G, C) pair wise between the four sequences the following picture will emerge (Table 7.2)

The matrix constructed in this way clearly shows that kangaroo and opossum are close to one another. Similarly, monkey and horse are close to one another but distant from the other two. It is possible to build the phylogenetic tree from these data but one should take into account long stretches of DNA to have a proper tree built.

A matrix of homology (Table7.3) can be built in the following way. Between horse and monkey there are differences of 7 nucleotides out of 27. So percent homology is [(27-7)/27] X 100 = 74%. It can be easily calculated that monkey and horse have 74% homology whereas kangaroo and opossum have even more (85%). However, homology between monkey (or horse) and kangaroo (or opossum) is comparatively much less which is also expected. It is obvious that homology is more when distance is less. The two are inversely related.

**Table 7.3**

**Homology Matrix**

| | Monkey | Horse | Kangaroo | Opossum |
|---|---|---|---|---|
| Monkey | 100 | 74 | 52 | 52 |
| Horse | | 100 | 56 | 56 |
| Kangaroo | | | 100 | 85 |
| Opossum | | | | 100 |

The parsimony method is quite different from homology/distance method. Let us take for example the character (base) at the 6th position (bold letters) of each of the four sequences. Both monkey and horse have T in that position whereas kangaroo and opossum have C in the same position. On the basis of this information only (forgetting about

others) it can be surmised that the first two must be close to one another, similarly the other two are close to one another. If this is correct, let us see which minimum change will help to deduce the route of evolution. One example is shown in (Fig.7.5). By changing T to C (single step) we can connect all the four. The ancestor of horse and monkey will have T in this position and the ancestor of the other two will have C in the same position. All the changes along the sequence have to agree to this relationship as shown above. If there is maximum possible agreement that will lead to the most parsimonious tree, or in other words, that will be the actual phylogenetic tree or at least the most probable one.

It is better to explain now some of the terminologies used in building a phylogenetic tree. Operational taxonomic units (in short OTU) mean contemporary sequences of proteins or DNAs which are used for building the phylogenetic trees as described below. A tree (Fig. 7.6) consists of nodes (exterior as well as interior). A, B ,C, D are the external nodes whereas G and H are the internal nodes. External nodes naturally represent OTUs whose sequences have been used to

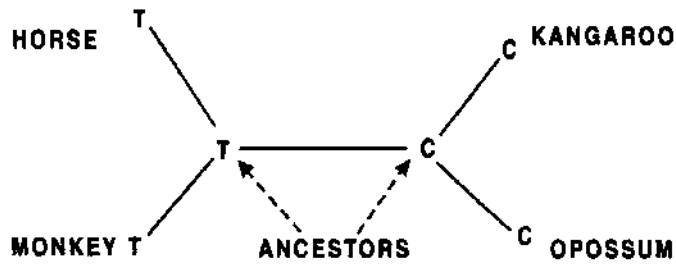



Fig. 7.5 Phylogenetic relationship of four animals as derived from homology matrix based on DNA sequences. Parsimony method has been used. According to this, horse and monkey had a common ancestor. Similarly Kangaroo and Opossum had a common ancestor. The two common ancestors are related to one another (see text for discussion). This is, however, a very simplified example.

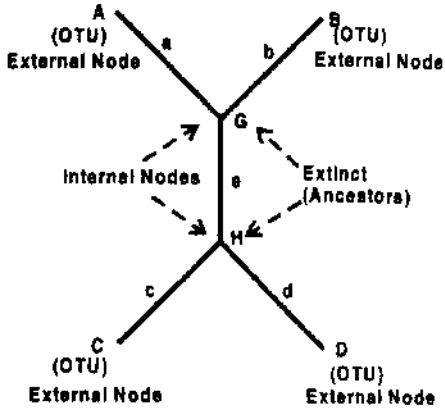



Fig.7.6 Terminology for construction of phylogenetic trees. Operational taxonomic units (OTUs) are the various species (extant organisms) for which the tree is constructed. OTUs (A, B, C, D) are naturally at the end of the branches and are referred to as "External nodes". Internal nodes (G, H) are the points from which the branches bifurcate and naturally represent ancestors (extinct organisms). The branch lengths are a,b,c,d,&e.

build the phylogenetic tree. Internal nodes represent the ancestors, for example, G is the ancestor of A and B. Similarly H is the ancestor of C and D. Two ancestors G and H must have had a common ancestor which is not known as the tree is an unrooted one . The ancestors are all extinct. According to the Darwinian theory, the two species which are close to one another must have a common ancestor which is also extinct. It is interesting to note that the tree depicted by Darwin in 'Origin of Species' to explain how the varieties (species) originate is the forerunner of all the trees we build today. The small letters a, b, c, d indicate links or branch lengths and are dependent on the non-matching sites between the aligned sequences. Those represent the mutation-based distances, as already discussed. Tree length is the total number of non-matching sites i.e., sum of the branch lengths including the distances between the ancestors. From the given sequence data a large number of possible trees may be constructed but the one with the minimum length is accepted as the most likely tree. This is naturally the tree derived through minimum evolution. For details regarding the construction of phylogenetic trees from molecular markers, the beginners may consult John Avise's book[4].

A few words about the universal tree of life will be useful at this stage. Originally, it was well accepted that the living world may be divided into two kingdoms, prokaryotes (or eubacteria) and eukaryotes (plant and animal kingdoms). As we already know, the cells of the former have no nucleus, so the genetic material is not partitioned whereas in the latter, chromosomes (genetic material) are encased in the nucleus and thus isolated from the cytoplasm. When the microorganisms which can survive under very harsh conditions like extreme temperature, high salinity, presence of unusual chemicals like sulphur etc., were discovered those were thought to be more primitive than bacteria and therefore branded as *archaea* (rather archaebacteria). Later Woese extensively used ribosomal DNAs as molecular markers for phylogeny and after studying a variety of organisms including

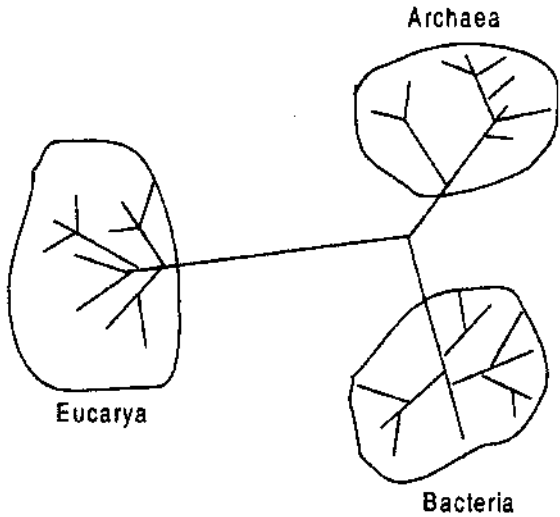



Fig. 7.7. Division of universal tree of life into three broad kingdoms. Divisions are (a) Bacteria (b) Archaea and (c) Eucarya. The details have been described in the text.

*archaea* came to the conclusion that archaebacteria evolved as a separate branch. They have properties mostly common to bacteria but some characteristics as that of eukaryotes. Therefore, in the universal tree of life *archaea* forms an independent kingdom. It is generally agreed upon that the universal tree of life can be divided into three broad kingdoms, which are bacteria, archaea and eukaryotes (Fig.7.7). However, it should be remembered that some thermophylic and halophylic bacteria do not belong to Archaea.

Let us now concentrate on the route of evolution of DNA itself, as the key in understanding evolution lies in the solution of this problem. As the organisms developed more and more the DNA content increased with the increase in complexity of the organism but not quite proportionately. Obviously, as the system became complicated more and more information became necessary to operate the system. So the increase in content of DNA is expected but there are certain

**Table 7.4**

**Genomic size and code content**

| Organisms | Genomic size (base pairs X $10^9$) | Coding DNA (%) |
|---|---|---|
| Bacterium*(E.coli)* | 0.004 | 100 |
| Yeast *(Saccharomyces)* | 0.009 | 70 |
| Nematode (*Caenorhabditis*) | 0.09 | 25 |
| Fruitfly (*Drosophila*) | 0.18 | 33 |
| Newt *(Triturus)* | 19.0 | 1.5-4.5 |
| Human (*Homo sapiens*) | 3.5 | 9-27 |
| Lungfish (*Protopterus*) | 140.0 | 0.4-1.2 |
| Flowering plant (*Arabidopsis*) | 0.2 | 31 |
| Flowering plant (*Fritilaria*) | 130.0 | 0.02 |

anomalous situations, which are not understood. DNA present in simple unicellular organisms, as bacteria, has on the average $4x10^6$ base pairs (Table 7.4). The DNA content of yeast which is basically a unicellular organism and belongs to eukaryotes having well-defined nucleus, has more than double the amount of that of E. coli. DNA content of *C. elegans*, a nematode, is 10 times that of yeast, and that of the fruit fly *Drosophila* is much more ($1.8x10^8$bp). The human DNA contains approximately $3.5x10^9$ bp. The DNA content of lungfish and the flowering plants (*fritilaria*) is huge, $140x10^9$bp and 130 $x10^9$bp respectively. However, the DNA content of another flowering plant (Arabidopsis) is comparatively lesser ($2x10^8$ bp).

There is another anomalous situation regarding the DNA content of the various organisms as shown in the (Fig. 7.8). The DNA content of the genome is usually fairly constant among the prokaryotes and lower eukaryotes like fungi, algae etc., but varies widely ($10^8$-$10^{11}$) among some higher eukaryotes, specially, flowering plants, where there is 1000 fold variability, almost three orders of magnitude between minimum and maximum values. This is also true for some insects, fish, amphibians and some other species. This is known as c-DNA paradox. Even today it is not known why the DNA content varies so much when the complexities of the related species are more or less in the same order of magnitude. This becomes more perplexing when the information content of the DNA is taken into account. The indexing could be done by the amount of coding DNA on percent basis (Table 7.4). Practically all of the DNA of a bacterium is coding DNA. Its entire genome has the information (100%), yeast has 70% and *Drosophila* 33%. It becomes obvious that the information content (percent wise) decreases with the increase in the amount of DNA. Again there is more anomaly among the different flowering plants. *Arabidopsis*, one of the flowering plants has reasonably high coding DNA (31%), whereas *Fritilaria* has only 0.02%. In the last case it appears that major amount of DNA has no information. Contrary to

that, humans have only 1% of coding DNA. The question is whether it has really no information or we are not aware of it. We know by now that the informative DNA is not constituted of coding region only, it has some other functions as well. For example, it has to bind to several factors known as transcription factors for regulating the expression of the genes. So far our present day knowledge indicates that this part would be compratevly lesser than the coding region. Again, coding regions (exons) are separated by noncoding regions (introns). Introns may or may not have some functions we do not know about. The words 'selfish DNA' or 'junk DNA' are often used

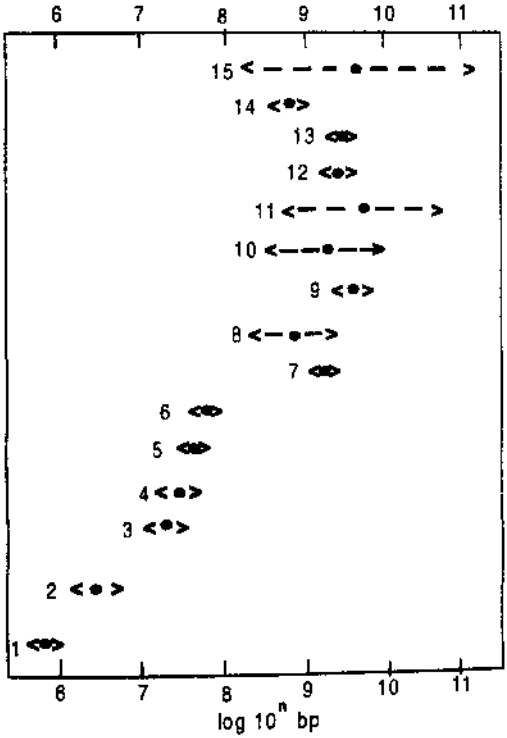



Fig. 7.8 C-value paradox. DNA content of an organism is generally related to the complexity of an organism. This is usually true for simple eukaryotes but quite variable in case of higher organisms. The details have been described in the text. · Mean value ,< > Range of values.

1.Mycoplasma 2.Bacteria 3 Fungi 4.Algae 5 Molds 6. Worms 7 Mollusks 8 Insects 9. Cartilaginous fish 10. Bony fish 11. Amphibians 12. Reptiles 13. Mammals 14. Birds and 15. Flowering plants.

to refer to those regions of the genome, which have apparently no function. 'Junk DNA' naturally means that it is of no use and 'selfish DNA' indicates that it is not interested in doing any work for others, other than perpetuating itself along with other useful genes by replication. None of the two terms makes us understand the need for creation of such useless material through evolution. It is possible that the evolution of DNA is not necessarily always on demand. Duplication, triplication, multiplication followed by mutations are natural processes. It may also be possible that the conversion of non-coding DNA into coding message or any other type of message is a subsequent phenomenon. May be, selfish DNA or junk DNA is anxiously waiting to be converted to a meaningful message in future through the evolutionary processes. This is, however, not generally agreed upon. Alternatively, during multiplication of genes, many genes become functionless due to inversion, fragmentation, recombination and numerous other processes. We cannot rule out the argument that the so-called junk DNA is there to absorb mutations which occurs randomly at different sites at a constant rate. Occurrence of pseudogenes is rather very common.

The most interesting development during evolution of DNA is the switch over from a prokaryotic cell to an eukaryotic one. DNA occurs without any encasement in prokaryotes where as in the eukaryotes it remains isolated in the nucleus, without any direct contact with the cytoplasm. Before we consider this fact we have to understand the problem of development of the genomic material. Whatever might be the primordial genetic material either DNA or RNA (see the earlier chapter), it existed in small sizes and contained very limited genetic information due to its small size e.g.,in prokaryotes. There are plenty of evidences in the literature to indicate that a large number of genes evolved by duplication, triplication etc. A gene may occur more than once in a genome, there may be so many copies of a particular gene. There may be some differences between

them due to mutation. But through the evolutionary process over millions of years a gene may change so much that it may carry out different (sometimes related and sometimes entirely unrelated) functions. For example, the dehydrogenases like alcohol dehydrogenase, glucose-6-phosphate dehydrogenase, lactic dehydrogenase etc., must have evolved from a single ancestral gene because all of them have practically the same cofactor (NAD)-binding site. Naturally, all of them carry out similar functions but the substrates are quite different. Similar cases are known of which the best example of duplication of genes is of myoglobin and the variety of β-chain of haemoglobin, as ascertained from the early work of Fitch and Margolish. These chains originated from an ancestor by duplication at various stages. The ancestor of modern haemoglobin genes duplicated about 380 million years ago. One of the genes evolved into α-chain of haemoglobin and the other diverged into different chains like α and β chains. The first oxygen binding molecule produced was myoglobin and α chain of haemoglobin. With the evolution of other chains haemoglobin (a tetramer) became more efficient in oxygen binding. Another duplication occurred about 150 million years ago during the evolution of higher primates. One chain continued as α-chain and the other evolved into β-chain. Besides the duplication a variety of other changes also followed but gene duplication was the primary event. So it can be assumed that the original genome of primitive organisms was simple enough and more and more genes might have evolved through millions of years by multiplication followed by addition, alteration or mutation to give rise to other genes as shown in Fig. 7.9. There are several ways through which genomes can evolve. Three most probable ones are: (1) Multiplication followed by mutation. A single gene like A can multiply and through mutation can generate other closely related genes like B and C. Many of the closely related genes originate in this way. (2) Symbiosis, a procedure by which a variety of genomes by duplication (or multiplication in general) may be assembled, then encased by a membrane (nuclear

membrane) and survives by feeding each other with different genetic messages (information). The higher organisms, which have more than one chromosome, might have been derived in this way. (3) Epigenesis, this process may lead to differentiation. A single genome (derived by multiplication followed by mutation) may eventually give rise to differentiations due to mutation in different genes in specific ways. The above are mostly speculations without experimental evidences.

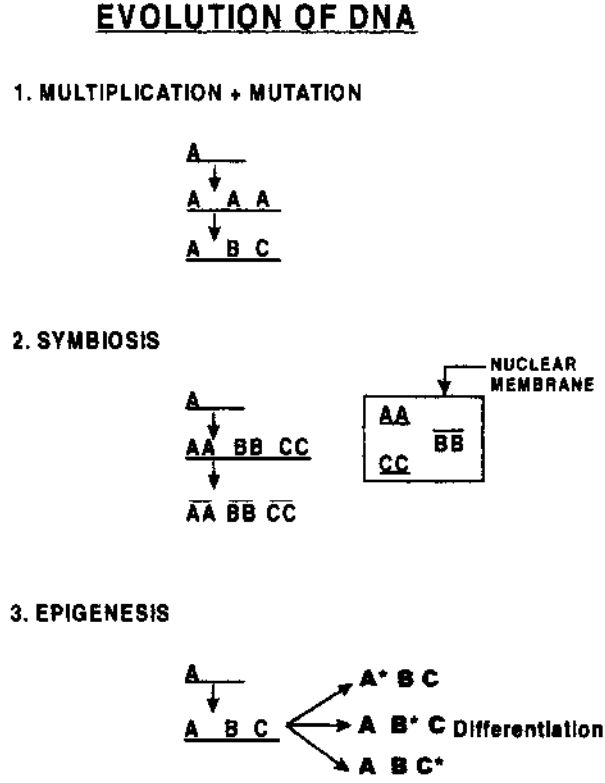



Fig. 7.9 Evolution of genomes. There are several ways through which genomes can evolve. The details are described in the text (The figure is adopted from a review written by Swathmary and Myrad Smith in *Nature*, vol.374, p227, 1995)

Therefore, the major event in the evolution of DNA is the multiplication followed by mutation, as depicted in Fig. 7.9. It may be interesting to note at this stage that the smallest parasitic microorganism, *Mycoplasma* has about 500 genes and it was calculated that it can survive with only half the number of genes since some of them are repeat genes and some have common functions. In contrast, human genome is assumed to have 50,000 - 100,000 genes.

However, the most amazing aspect of evolution of genomic DNA is its fragmentation followed by encasement within the nucleus. Generally the microorganisms have a single chromosome (or simply a DNA molecule) which is circular. This was perhaps necessary to protect against the damage to the ends. In the eukarytoes, DNA molecules are linearised and the method of repair of damaged telomeric ends evolved. Naturally, this dramatic step led to the evolution of eukaryotes from the prokaryotes. Actually, one of the primary events in the evolution is the origin of the nucleus. It has been assumed that the fragmentation of the genomic material took place when it grew in size but the fragments collaborated with each other as in symbiosis. The next step was the encasement of different fragments within the nuclear membrane. Since the fragmented genomes have to depend on each other for replication, function etc., their survival and operation may be equated to endosymbiosis as this symbiosis is an internal event ('endo' means within). Before we proceed further it may not be out of place to point out the evolutionary advantage of this fragmentation. As the genome grows larger in size it becomes rather difficult for the cell to handle it and maintain the structure, replication, transcription etc. So the fragmented genomes became atmost a rule in case of eukaryotes. The problem was not very acute in prokaryotes where the genomic sizes are reasonably small to handle although some bacteria are known to have more than one chromosome. This part of the evolution, the transition of prokaryotes to eukaryotes is hardly understood on the basis of the

present day knowledge of molecular biology.

Various evidences have clearly established that the organelles mitochondria, chloroplasts etc., have been generated from the outside invading organisms. These organelles have their own genomes but not the complete information for their replication, function etc. They are dependent on the nuclear genomes for their functions (specially replication) as their own genomes have not the requisite information, complete in all respects. One may be tempted to compare these organelles with the viruses, which also do not have the complete information for their own replication and have to borrow information from the host. There is however, a big difference between the two, one being helpful and the other destructive for the host. Some viruses (for example, lysogens) may coexist with the host but they usually do not provide any helpful information to it. The organelles, however, not only borrow some genetic information from the host but also supply some to them, which is a clear instance of symbiosis. More than that the cells carry out very important functions with their help for example, energy production by mitochondria, photosynthesis by chloroplasts and so on. This further strengthens the concept of symbiosis. Some preliminary evidences are available to support the assumption that the nucleus also might have originated from the invading microorganism(s). Thus it is a foreign body for eukaryotes. However, more direct and strong evidences are necessary to establish this hypothesis.

A few words about chromosomal organization may not be out of place here. Although DNA is the only information carrying material in chromosomes its organization is fully dependent on a number of proteins, specially histones, sometimes spermins. The organization in turn plays very important roles at different stages like duplication during cell division and partition of genetic material among the daughter cells, expression of genes and shut off with precise tuning depending on the necessity etc. Although chromosome number is

expected to increase with the size of DNA but that rule is not always followed. As mentioned before, the number of chromosomes increased by fission of the existing chromosomes in some cases. Similarly, the number may decrease by the fusion of chromosome (as evident in case of human and chimpangee). Thus fission and fusion of chromosomes are part of evolutionary processes. Susuno Ohno argued quite sometime ago that genomes were extensively duplicated in evolution. Although it was not generally agreed upon but is now accepted more or less as an axiom.

Another aspect of evolution of DNA is extremely important for multicellular organisms, specially for the evolution from lower eukaryotes to higher eukaryotes. A simple example is that lung cells of any animal have the same DNA as the liver cells but carry out different functions. Even the structural organisations of the different organs harboring the same DNA are different. How is it possible? Naturally, the expressions of different genes from the same genome take place under different environments of lung and liver or, in other words, there must be differential expression of various genes in different organelles. Some are induced or stimulated in one organ and remain dormant or expressed to less extent in another organ. Some genes are switched on and some switched off under different circumstances. There must be differences in the regulation of gene expression as a result of the information content in the individual genes. Naturally the regulatory genes play a very important role in differentiation. This must be programmed in the early processes of differentiation of the cells. As the cells are predetermined to be part of specific organs, DNA is also predetermined to express the desired genes in the individual organs. The term 'epigenesis' which was in routine use for geological features was coined by C.H. Waddington to indicate the effect of cellular environment to regulate the genes for differentiation. More recently the term is used to describe mechanisms, which can control gene expression without themselves being under

immediate control of genes. Although, the basic mechanism is vaguely understood epigenesis may account partly for complicated and diverse functions of highly developed eukaryotes which would be the crucial steps in evolution. In the meantime, a number of regulatory genes involved in development have been identified in the worm *C. elegans* and the fruitfly *Drosophila.* Very recently a large number of micro RNAs have been discovered in plants and animals (as discussed in Chapter 5). These are involved in development and may throw more light on the mechanism of regulation of gene expression at various stages.

Finally, another crucial step in evolution which is in the realm of sociobiology is now in progress and that is mainly confined to human society. It should, however, be pointed out that sociobiology is prevalent even among the lower animals like ants, bees etc. This is much more evident among the primates. However, the human brain, mainly its nervous system is under tremendous selection pressure from the acquired knowledge, which is developing at a very rapid stride. Psychiatry is no more a philosophical exercise but part of social science, the understanding of which is essential to cope with the developing human society. This phase of evolution in the human is perhaps the ultimate for Darwinian evolution.

In summary, although the Darwinian evolution has been well accepted, in general, lot of intricacies have cropped up in that connection and require urgent attention. But basically, Darwin's concept of 'Natural selection' has withstood the test of time.

## References

1. *Origin of species (by means of natural selection or preservation of forward races in the struggle for life).* Charles Darwin, Avenel Books, 1979.
2. *Darwin and the Beagle.* Alan Moorehead, Harper and Row, New York, 1969.

3. *Encyclopedia of Molecular Biology.* John C. Kendrew, Editor-in-Chief, Blackwell Sciences. London. 1994.

4. *Molecular markers, natural history and evolution.* John C. Avise, Chapman and Hall, New York, 1994.

5. *Evolution*, Th. Dobzhansky, F.J. Ayala, G.L. Stebbins and J. W.Valentine, W.H.Freeman and Company, 1976.

*****

# Epilogue

After going through the seven chapters of the book one may wonder 'Where is the music? To justify the title may I quote from the Webster's dictionary 'Music is the science or art of ordering tones of sounds in succession, in combination and in temporal relationships to produce a composition having variety and continuity'. *Through the seven chapters I have attempted to order scientific facts to compose the 'Music of Life'.*

# Index