THE MOLECULAR BASIS OF GENE EXPRESSION

R. Hehlmann[1], T. Hunt[2],

[1] Institute of Cancer Research
and Department of Human Genetics and Development,
College of Physicians and Surgeons,
Columbia University, New York, N. Y. 10032

[2] University of Cambridge,
Department of Biochemistry, Tennis Court Road,
Cambridge

Our present understanding of gene expression and transcription is based on experiments performed and models proposed only little more than 10 years ago.(1)

According to this understanding the genetic information is contained in and arranged linearly on nucleic acid molecules (DNA and RNA) in form of a specific base sequence (adenine = A, guanine = G, cytosine = C, thymine = T, and in the case of RNA uracil = U instead of thymine).

Three bases form one triplet which specifically codes (Nirenberg and Matthaei, 1961) for either an amino acid, a starting or a termination signal in protein synthesis. The genetic code summarizes the coding properties of all 64 possible triplets (F. H. C. Crick, 1966,(2)).

In general, DNA is the genetic material. Exceptions are some bacterial-, plant-, and animal viruses, e. g. the RNA-tumor-viruses. The genetic material of cells consists of one or more comparatively long DNA strands of high molecular weight, which are mainly localized in a nucleus.

For proteinsynthesis the genetic information of the DNA must 1) be available in defined units, 2) transcribed into m-RNA, and 3) a regulation mechanism is required which controls the species and amount of proteins to be synthesized at a given time.

The organization of DNA

The operon model by Jacob and Monod (developed for bacterial genes in 1961) proved highly useful for organizing genetic information in defined units. In this model the functional unit of gene expression is the operon, which consists of several structural genes (one gene codes for one protein) and some regulatory sites controlling the same biochemical pathway. The model specifies that the genes of one operon are usually transcribed together to produce one single messenger RNA (m-RNA synthesis or transcription). The enzyme synthesizing the m-RNA is the DNA dependent RNA polymerase or transcriptase. It appears to bind to a „promotor" site of the operon to initiate m-RNA synthesis unless negative control is exerted by a „repressor" molecule which can bind to the „operator" site localized between promotor region and structural genes. In this case, m-RNA synthesis would be initiated by depression of the operon i. e. inactivation of the repressor by an „inducer", which leads to lower affinity of the repressor for the operator site.

THE GENETIC CODE

1st position	2nd position U	2nd position C	2nd position A	2nd position G	3rd position
U	PHE	SER	TYR	CYS	U
	PHE	SER	RYR	CYS	C
	LEU	SER	Ochre (Chain-termination)	(Chain-termination)	A
	LEU	SER	Amber (Chain-termination)	TRP	G
C	LEU	PRO	HIS	ARG	U
	LEU	PRO	HIS	ARG	C
	LEU	PRO	GLUN	ARG	A
	LEU	PRO	GLUN	ARG	G
A	ILEU	THR	ASPN	SER	U
	ILEU	THR	ASPN	SER	C
	ILEU	THR	LYS	ARG	A
	METH	THR	LYS	LRG	G
G	VAL	ALA	ASP	GLY	U
	VAL	ALA	ASP	GLY	C
	VAL	ALA	GLU	GLY	A
	VAL	ALA	GLU	GLY	G

Transcription

The synthesis of m-RNA (= transcription) provides a mediator molecule between the „immobile" DNA in the nucleus and the sites of protein synthesis at the ribosomes in the cytoplasm. m-RNA, which was proven to be complementary in its base sequence to the template DNA (Hall and Spiegelman, 1961), thus represents a true copy of the DNA and contains all genetic information of the corresponding stretch of DNA on a comparatively small and mobile molecule which can be degraded easily after it has fulfilled its function.

Transcription, in addition, is a step in cell metabolism at which the kind and amount of proteins to be synthesized can be regulated.

According to present thinking (4), transcription is initiated by attachment of transcriptase to the promotor site and subsequent binding of two ribonucleotides to the enzyme. Chain elongation occurs by sequential addition of nucleoside monophosphates to the 3′-terminus of the nascent RNA-chain. The mechanism of termination of transcription at the end of the operon possibly involves certain DNA termination sequences but is not yet completely understood. After release from its template DNA and transport to the ribosomes, the m-RNA is then thought to be available for translation and protein synthesis.

Reverse transcriptase

In the context of this workshop it appears useful to briefly define an enzyme which is of importance for the understanding of a number of papers presented here and which, because of its name, caused some confusion as to its relatedness to transcription.

Until three years ago it was believed that the flow of genetic information was only in the direction DNA to RNA and RNA to RNA. However, Temin and Baltimore in 1970, found an enzyme in RNA-tumor viruses, which can synthesize DNA on a single stranded RNA template, and thus reversed this conception.

Because of its effect opposite to that of the DNA directed RNA synthesizing transcriptase, this enzyme was called „reverse transcriptase". Its operational definition includes four parameters. The DNA-synthesis mediated by a reverse transcriptase 1) must accept single stranded natural RNA as template, 2) must be sensitive to ribonuclease, 3) must depend on the presence of all four deoxyribonucleotides, and 4) the DNA product of the reaction must be complementary in its base sequence to its RNA template. The latter is tested by so called back-hybridization.

At present, there is no indication what so ever that reverse transcriptase has anything to do with transcription of DNA into m-RNA.

Translation

Proteins are linear polymers of amino acids, and the problem in their synthesis is how to put the correct amino acid into the correct position in the chain. There are 20 amino acids which can serve as precursors of proteins; some amino acids – the formation of hydroxyproline from proline in collagen is an example of this. The basic elements of the mechanism of protein synthesis are fairly clear, though detailed understanding of many aspects is still lacking (3). This account starts with a description of the important elements, continues with an account of the cycle of events leading to the synthesis of a protein, and concludes with a description of how and under what circumstances the synthesis of proteins can be regulated.

Ribosomes

Proteins are synthesized on ribosomes; at some point during the process all the elements to be described interact with these particles. They are roughly spherical particles, about 200 Å in diameter with a molecular weight of about 4×10^6. About half the mass is protein, and half RNA. At low Mg^{++} concentrations, or in the presence

of certain proteins, the ribosome dissociates into two unequal subunits which are known by their sedimentation rates, 60S and 40S. The 60S subunit contains two species of RNA with molecular weights of 1.6×10^6 and 1×10^5 (28S and 5S RNA) while the 40S particle contains one RNA molecule with a molecular weight of 8×10^5 (18S RNA). There are probably few, if any overlaps between the protein content of these particles – that is, one can assign every protein of the intact, 80S ribosomes to either the 40S or 60S subunit. There are 50–60 different proteins in ribosomes from mammalian tissues.

Transfer RNA

For every ribosome in the cell there are between 10 and 20 molecules of transfer RNA. This is a heterogeneous class of RNA with an average molecular weight of about 25.000 sedimenting at about 4S; they contain between 75 and 95 nucleotide residues. Despite the heterogenity, these molecules possess several features in common. They all terminate in the sequence –CCA, and each molecule can accept an amino acid bound by an ester linkage to the ribosome of the terminal adenosine residue. The combination of the amino acid with its correct tRNA is catalysed by a series of 20 or perhaps more enzymes which require ATP for activity, performing the overall reaction:

Amino acid + ATP + tRNA = Aminoacyl-tRNA + AMP + PPI

The combination of each amino acid with its cognate tRNA is absolutely specific.

Messenger RNA

The sequence of amino acids in proteins is specified by messenger RNA. Ribosomes attach to mRNA at one end, at a strictly specified site, and proceed along the mRNA until they reach a termination signal, and release the protein they have been assembling. The translation of the sequence of nucleotides in the mRNA is accomplished by aminoacyl tRNA, which binds specifically to codons in the mRNA by classical base-pairing. This automatically brings the amino acid at the other end of the tRNA into the correct register so that a peptide bond can form between it and the growing polypeptide chain. Messenger RNA itself has proved an elusive entity, mainly because there is very little of it compared to the tRNA and ribosomal RNA, and also because it is exceedingly heterogeneous, both as regards size (which must be roughly proportional to the size of the protein it specifies) and composition, which is also a function of the protein it codes for. However, it is now possible to prepare specific mRNA – examples are the mRNA for globin, immunoglobins, viral proteins of several species, silk fibroin, and so on – the list lengthens almost daily at the time of writing.

„Factors"

Besides the proteins of the ribosome, several enzymes which are only loosely or transiently associated with the ribosomes catalyse the process of protein synthesis. There seem to be at least 6 or 7 identifiable activities, though some would extend the list. There appear to be at least three, and possibly 6 factors involved in initiation, two in the assembly, and two for termination of the chains. Some of these proteins are well-characterized and highly purified; others are not.

The Ribosome Cycle

Immediately after a ribosome has finished making a protein, it leaves the mRNA, drops the finished protein, and dissociates into its subunits. The first identifiable step in the next cycle of synthesis is the binding of methionyl-tRNA$_f$ to the 40S subunit, catalysed by an initiation factor (or two) and GTP. This tRNA is unique in several respects, and is the only known tRNA which can bind to ribosomes in the absence of mRNA. It is used only for starting new protein chains; another methionyl-tRNA exists to put methionine into internal positions in the chain. Having bound the initiator tRNA, the mRNA is bound to the 40S/met-tRNA$_f$ complex. This probably requires a protein or proteins, and following the correct binding of the message, the 60S subunit joins on. The next amino acid specified by the mRNA is now brought into position, attached to its tRNA, and catalysed by a protein whose function is to carry tRNA from the synthetases to the ribosomes; this process appears to involve GTP binding and probably hydrolysis also. Having bound correctly, the first amino acid forms a peptide bond to the second, and the ester linkage between the methionine and the tRNA is also broken. The enzyme which catalyses the peptide bond formation is an integral part of the 60S subunit, and cannot be removed. The situation is now that a dipeptide, bound to tRNA, is located in the same place as that tRNA was located when it first entered the ribosome; it is now necessary to move the peptidyl-tRNA into the location that was first occupied by the met-tRNA$_f$, so that there is a space for the next aminoacyl-tRNA to enter the ribosome, and so that the next triplet codon on the mRNA is properly positioned. The movement, called translocation, is catalysed by another enzyme, called EF II; GTP hydrolysis is again involved. After translocation, the scene is set for another round of the cycle to occur, and the process continues with EF I catalysed binding of aminoacyl-tRNA, peptide bond formation and translocation until a termination signal in the mRNA is reached, at which point of protein termination factor enters the ribosome instead of a tRNA, and translocation probably results the release of the completed protein.

The Ribosome Cycle in Eukaryotes

1. Native 40S subunit + Met-tRNA$_f$ + GTP → 40S/met-tRNA$_f$/GTP
2. 40S/met-tRNA$_f$/GTP + mRNA → 40S/mRNA/met-tRNA$_f$(GTP?)
3. 40S/mRNA/met-tRNA$_f$ + native 60S subunit → 80S/mRNA/met-tRNA$_f$
4. 80S/mRNA/met-tRNA$_f$ + val-tRNA $\xrightarrow{EFI\ GTP}$ 80S/mRNA/met. val-tRNA
5. Translocation of met. val-tRNA from A site to P site, catalysed by EFII and GTP
6. Repeat steps 4 and 5 until termination codon is reached, when R factor binds, and combined peptide bond formation to ? water and translocation release nascent chain from the ribosomes and the ribosomes from the mRNA. At this point, the ribosomes exist as subunits, and have a choice of either repeating the cycle or becoming relatively inactive 80S ribosomes.

The details of these reactions are in most cases quite obscure, particularly the movement of the tRNA and mRNA with respect to the ribosome. Many of the ideas about the mechanism have come from taking the components apart by means of washing ribosomes with strong salt solutions, and adding back wash fractions to see what is needed to restore activity. Also important, however, are a range of inhibitors which affect various aspects of the cycle; for example, aurinetricarboxylic acid, pactamycin, and edeine all inhibit various phases of the initiation process; sparsomycin, and anisomycin inhibit peptide bond formation; diphtheria toxin (+ NAD) and cycloheximide inhibit translocation; and puromycin mimics the end of aminoacyl-tRNA closely enough that it causes the premature release of nascent peptides from the ribosome. This list is not exhaustive; the importance of these inhibitors is that they may cause the ribosomes to accumulate at one point in the cycle, which reduces their heterogeneity and makes the study of intermediates in the cycle a possibility.

The Control of Protein Synthesis

There are very few well-understood examples of control of protein synthesis. It is an article of faith that ribosomes are indiscriminate translators of whatever mRNA happens to come their way, so that most of the control of the type of protein synthesized in cells is made at the level of transcription of DNA into RNA, or the subsequent processing of the RNA or its transport to the cytoplasm. In the case of infection by viruses like Vaccinia, VSV, polio, and other viruses which replicate in the cytoplasm, it seems that viral RNA is poured into the cytoplasm and takes over; some believe that they also subvert the ribosomes, so that they can only read viral messages; but there is little evidence that this is in fact the case. There is one situation in which it seems that protein synthesis can be turned on and off during the cell-cycle, in the synthesis of histones. Histone synthesis occurs only during the S-phase of the cell-cycle, and is strongly inhibited by inhibitors of DNA synthesis. The basis for this control is unknown. Other case of control of protein synthesis seem to be more general – a kind of overall quantity control. This occurs during amino-acid starvation, at elevated temperatures, during mitosis, in serum deprivation and various other conditions which are sub-optimal for cell-growth. In these conditions, protein synthesis is inhibited non specifically at the level of initiation. The basis of this control is not understood.

References

1. J. D. Watson, Molecular Biology of the Gene, Benjamin Inc. 1970
2. Cold Spring Harbor Symposia on Quantitative Biology, Volume XXXI (1966)
3. Cold Spring Harbor Symposia on Quantitative Biology, Volume XXXIV (1969)
4. M. Chamberlain, in Cold Spring Harbor Symposia on Quantitative Biology, Volume XXXV (1970), 851–887,