On the Translation of the Genetic Code

MARSHALL W. NIRENBERG

National Heart Institute - National Institutes of Health

I would like to take this opportunity to relate something of our knowledge of the genetic language. For some two to four billion years some such language has probably provided the basis for a continuous dialogue between cells and their descendants. Fossil records of bacteria about 3 billion years old have been reported (by Baghoorn-Schopf); the first vertebrate appeared approximately 500 million years ago; and amphibians and mammals about 350 and 180 million years ago, respectively. The presence of bacteria 3 billion years ago may indicate the presence of an operational code at that time, almost surely the code has functioned for more than 500 million years. The remarkable similarity in code-words used in bacterial, amphibian and mammalian replicative processes suggests that most, if not all, forms of life on this planet use almost the same genetic language, and that this language has been used, possibly with few major changes, for at least 500 million years.

It is by virtue of this language that each generation is able to pass to the next generation a library of information which specifies in detail how to make the many kinds of protein catalysts that the cells will need for their development. And although it now seems clear that all, or almost all, forms of life on this planet use virtually the same language, recently a number of "dialects" have been found. I shall describe this

later.

The elucidation of the genetic code has been the subject of much intensive work, particularly in the past four or five years, and I would like to stress at the outset, that this work, and particularly the work with which I have been associated, has been, in a very real sense a collaborative project.* I think this will become evident as I proceed.*

First may I recall briefly - as must have been done frequently in this Symposium - the basic features of the Crick-Watson scheme of protein synthesis. (Fig. 1) Here is shown Fig. 1 schematically the double-stranded DNA, which together with an enzyme, - RNA-polymerase, - catalyses the synthesis of messenger-RNA, using the DNA as a template. Only one strand of the DNA is copied by the RNA-polymerase; and the copying process is sequential, and there are signals, whose exact nature is unkown, which specify the beginning and the end of the messenger-RNA synthesis. The next diagram (Fig. 2), shows schematically, the process of protein synthesis. In the DNA shown here, the different cross-hatchings represent various segments of DNA, each corresponding to a specific protein, or group of proteins. Ribosomes are shown, schematically, attached to the messenger-RNA

*List of collaborators at Bethesda.

where reading, or translation, begins; as soon as one ribosome moves down the messenger-RNA, another becomes attached until the messenger-RNA is virtually covered with ribosomes.

The actual reading takes place by means of s-RNA (soluble-RNA), which carries specific amino-acids and recognizes particular m-RNA code words on the ribosomes. Thus the code-word, or codon, is recognized not by the amino-acids, per-se, but by an adaptor molecule - the s-RNA.

Fig. illustrates, again diagramatically but in more detail, the codon recognition process as exemplified by that most intensively studied organism, E.coli.

The ribosome of E.coli comprises two sub-units: the larger 50S and the smaller 30S. The messenger-RNA lies on the smaller part of the ribosome, and presumably three bases in the messenger-RNA molecule (a 'codon') are recognized by three bases in the s-RNA (an 'anti-codon'). and this latter can then bind at one of two possible binding sites on the larger ribosome sub-unit, a particular amino-acid, (aa). One of these binding sites is for the peptidal s-RNA, so the s-RNA which is attached to the growing (protein) poly-peptide molecule; and the other for the incoming amino-acid s-RNA. Thus three enzymes (the two s-RNA's and) plus (GTP), which supplies the activation energy, are required for the transfer of the growing poly-peptide chain to the next (incoming) amino-acid s-RNA complex.

When this is accomplished the s-RNA required for the previous amino-acid is discarded, and a shift ~~in some way~~ occurs so that the next codon (triplet of bases) on the m-RNA can be recognized by a new s-RNA. In this way the protein synthesis starts at a given place (on the m-RNA), ~~reads~~ groupings of three bases ~~sequentially~~ and with a given polarity.

In an actual living cell, even ~~the smallest~~ bacterial cell, ~~innumerable~~ biochemical processes ~~are~~ simultaneously, ~~in process~~ - all part of the cell metabolism. The synthesis of even a single protein is quite an elaborate process involving, inter alia, the transfer of a long DNA message to an m-RNA molecule which has ~~typically~~ sufficient nucleotides (about 1,500) to code ~~some 500~~ amino-acids for ~~the protein poly-peptide chains~~. Moreover, in an actual cell these 1,500 nucleotides will not be arranged in any simple sequence, reflecting the fact that there is a great number of different sequences of amino-acids ~~(of which 20 different varieties are made)~~ which constitute different proteins.

Nonetheless, by a great variety of biochemical and genetic investigations, especially with bacteria and viruses, ~~a great~~ many features of ~~the~~ protein synthesis, including ~~in~~ particular~~ly~~ ~~much~~ information about the code, has been obtained. The work I shall ~~be~~ describ~~ing~~ is, however, characterized by the use of much simpler, in vitro systems, where the essentially chemical features of some of the basic steps in the whole process are studied. The success of these methods, and the concurrence of the results ~~from~~

them with those from in vivo experiments, where both are available will, I hope, demonstrate how a physio-chemical or molecular basis can be found for the basic processes governing such fundamentally biological phenomena as cell metabolism and replication.

The basis for our earlier work on the DNA-RNA code was the use of synthetic messages, (in place, that is, of actual m-RNA) which were randomly oriented sequences of the four code letters, U(racil), C(ytosine), A(denine) G(uanine), the four bases of m-RNA. In this way some characteristic of the code could be determined, in particular the base compositions of the code-words, but not the sequence of the bases in the bases in the words. Thus the problem up to two or three years ago was like that of an anagram: we knew the letters comprising the code-words but not the order of the letters within each word.

It has been well established in several laboratories that if one added a synthetic messenger RNA, in particular polyuridylic acid - a synthetic RNA with entirely U bases - to a suitable mixture of ribosomes, s-RNA's, enzymes, ATP, GTP and amino-acids, that the poly-U would selectively bind phenylalanine s-RNA (i.e., the particular s-RNA associated with the incorporation of the amino-acid phenylalanine in protein) to the ribosomes. My colleague, Philip Leder, and I then speculated how small a message (of the RNA type) would direct the binding of s-RNA to the ribo-

some. Experiment showed that only three bases were needed, that is, very small molecules comprising only the triplet itself would direct the binding of the appropriate amino-acid s-RNA to the ribosomes. This provided a rather simple route towards the determination of the sequence of letters in the RNA code-words.

Our main problem was to devise suitable techniques for synthesizing triplets. At the time we started our work with such triplets, methods had been reported for making some 20 or 25 of the 64 ($=4^3$) triplets which can be constructed from the four nucleotides U, C, A and G. These had been prepared by enzymatic breakdown of RNA, or by chemical synthesis, in the latter case using some of the very elegant techniques devised by Khorana and his associates.

Two general techniques were developed in our laboratory, the first by Leder, Singer and Brimacombe, and the second by Merton Bernfield. The first employed polynucleotide phosphorylase, the enzyme which catalyses polymerisation of nucleotides, i.e. in addition of single nucleotides to di-nucleotides to make trimers, tetramers, pentamers, etc. The second method employed the enzyme pancreatic RNA-ase, which, although normally a breakdown or degradative enzyme, will also catalyze an exchange reaction between polynucleotides and can be used to make triplets with well-defined sequences. Using the methods of

Fig. ?

Khorana and these two enzymatic techniques, it was possible to synthesize amost all of the 64 triplets.

In connection with the use of ~~thism~~ small polynucleotide or "oligonucleotide" molecules such as the trinucleotides, it is important to point out ~~here~~ that any given sequence of nucleotides can exist, when incorporated in actual m‡RNA in three chemically distinct forms, depending on the location of the sequence in the whole messenger molecule. The chemical forms relate to the three positions (a) as an internal codon (trinucleotide) or as one of the other of the terminal groups - so called 3'-terminal codon and 5'-terminal codon.* This is illustrated in Fig. .

Fig.

All of the evidence ~~to date~~ suggests that the biological characteristics of codon recognition may in some, perhaps in many, ~~cases~~ be influenced by the particular position of the codon in the m‡RNA (or equivalently in the DNA). Thus each of the 64 triplets referred to above may exist in three effectively different structural forms.

The significance of these "secondary" chemical features is indicated by experiments, in vitro, with the oligonucleotides,

*The helical RNA (or DNA) has a definite sense or direction - with a definite "beginning" and a definite "ending". 3' and 5' refer to features of the chemical structure at these respective terminals.

and specifically by studying the influence of various (phosphorylating) substitutions on either the 3' or 5' terminal hydroxyl groups of the sugar in the trinucleotides. Thus Fig.

Fig.

shows the binding of phenylalanine sRNA to ribosomes as a function of the concentration of the trinucleotide. A simple triplet, UUU, has an activity* shown by (a). If one adds a phosphate to the 5' hydroxyl group of the sugar the activity is greatly increased, i.e., the binding or template effectiveness of the trinucleotide is greatly enhanced, (b). A phosphate attached to the 3' terminal lowers the template effectiveness, (c). Recently, Fritz Rotman prepared some analogues of UUU with a methyl group attached to the 5' phosphate, and also with a methyl group attached at both terminals, i.e. both 5' and 3' phosphate. The methyl group at the 3' phosphate terminal greatly reduced the template effectiveness. A triplet with 2'; 3' cyclic phosphate shows very little template activity.

It seems possible that significant terminal variations of this sort may occur in different biological circumstances, and that these may possibly regulate the template activity of the codons. For example, the terminal hydroxyls of the sugars (ribose) may

*The binding of the sRNA to the ribosome is determined by techniques in which a radioactive tracer is incorporated in the sRNA, so that the radioactivity associated finally with the ribosome complex is a measure of this binding. It is in that the term activity in figure denotes the effectiveness of the binding.

be modified in such a manner. Certainly a substitution at the 5'-terminus may be important because this could furnish a signal which specifies the attachment and/or the detachment of the ribosome from the message, (m-RNA or substitute). Recently Mitra and Hurwitz, and also Stent, have shown that, in vitro at least, messenger-RNA contains a triphosphate attached to the terminal hydroxyl; and although it is not clear what physiological function this triphosphate serves, it is highly plausible that it may in some way specify the initiation of reading the message. It could also determine the first (three letter) word to be read, phase the reading, and, perhaps affect the susceptibility to enzymes that could attack the termini of the messenger-RNA.

Internal codons may also be modified by these secondary chemical changes; the 2' hydroxyl or the base could be modified and such changes may be relevant to the punctuation of the message. The possibility also cannot be excluded that the codon recognition process is in some instances affected by the particular neighbors of that codon on the message.

Also, it should also be pointed out that there could possibly be a difference between internal initiation and termination (i.e., initiation or termination of polypeptide sequence (protein) by a codon internally located in the message) and terminal initiation and termination (the same process effected by terminal codons). Consider the situation where the messenger-RNA appears

to contain the information for the assembly of more than one protein, (or more than one polypeptide chain of a protein). If one starts to read (from the left in Fig.) the codon for

Fig.

the terminal initiation, one then reads in the message until one reaches the word that says "Stop", and then ~~there will be~~ an unknown mechanism ~~for~~ starting the second message at an interval position. It seems quite plausible, although not known, that these terminal and internal initiation and termination mechanisms could be different --possibly different codons.

Another feature of codon recognition concerns the degeneracy of the code, or the existence of synonyms, i.e., different codons which code the same amino-acid in the polypeptide sequence. With the appropriate oligonucleotides, one can examine, in vitro, the effectiveness of different synonym messages in binding the particular amino-acid s-RNA's to the ribosomes. The results of such are illustrated in Fig. . For example,

Fig.

phenylalanine s-RNA responded to both the oligonucleotides UUU and UUC, but UUC was slightly more active than UUU. Similarly lysine-s-RNA responded to both AAA and AAG but ~~here there is quite a marked~~ difference in the template activity ~~between~~ of the two synonyms. The first of these degeneracies, that between the (smaller) pyrimidine bases C and U when ~~they~~ occur as the third let-

ter of the codon, is universal throughout the code. The second case, the degeneracy of the (large) purine bases A and G in third place, occurs in all but two or three words (c.f. Fig.)

We turn now from these refinements and detailed features of the triplet-binding method to the actual results obtained by this procedure. Since the triplets have a well-defined sequence of nucleotides, and there are 64 possible triplets; and we have synthesized 63 of these and determined the amino-acids which they code. The results are summarised in Fig.

Fig.

The asterisks indicate base compositions of codons which were determined by directing protein synthesis in E.coli extracts with synthetic randomly-ordered polynucleotides. It is clear that there is a very close correspondence with the results of the earlier work. It is interesting to notice the types of synonyms which occur (some of which have already been mentioned). Thus glutamic acid corresponds to the codons GAA and GAG - an example of A=G degeneracy in the third place. Likewise Aspartic acid and GAU, GAC corresponding to U=C degeneracy in the third place. Another type of degeneracy is illustrated by Threonine which is coded by AC and any of the four U, C, A, G in third place. Methionine, on the other hand, is one of the rare cases (tryptophan may be another) in which there is no third place de-

generacy AUG ~~codes for~~ but AUA codes for Isoleucine.

This degeneracy of the code can have many consequences. One of the ~~more~~ obvious is the possibility of a great deal of "silent" mutation, that is on one of the ~~code-words~~, or groups of synonymous code-words, ~~there~~ may be conver~~sion of a base in the third position~~ to another base without resulting in an amino-acid replacement. Another obvious conclusion is that amino-acids which are very similar chemically, such as the dicarboxylic acids (aspartic acid and glutamic acid), have closely related codons. This may reflect the evolution of the code, but whether or not this is so, one consequence would certainly be that when an error in replication does occur, usually the first two bases are read correctly and the third one incorrectly. And very often the result of an error in reading will be the substitution in a protein of a chemically related amino-acid. Thus the general picture of the code is that it is quite conservative-- in the sense that it usually minimizes error or the consequences of error. The various patterns of synonym codons are summarized in Fig. . (N-formylmethionine sRNA shown here is the initiator).

Fig.

In addition to the codons for the specific amino-acids, ~~there as has been mentioned earlier,~~ some code-words ~~which~~ appear to serve special functions ("punctuation" etc.). For example, the recent work of Brenner, Garen and Zinder, and of others,

indicates that UAA and UAG may indicate the end of a message - although the precise mechanism for punctuation is unknown. UUG, CUG, AUG and in some cases GUG may specify the initiation of a message. Our recent studies, and also those of Clark and Marker in England, have indicated that these codons - at least when in terminal positions - are recognized by formylmethionine and this may serve as an initiator of protein synthesis. Some possible special function codons are listed in Fig. .

Fig.

Sanger first observed ~~in E.coli~~ that one of the two s-RNA ~~species associated with methionine could accept~~ a formyl group; that is the amino group of ~~the~~ methionine, after the methionine ~~was linked to the s-RNA could be formylated.~~ The ~~work~~ of Capecchi and colleagues, and of Zinder, ~~have~~ suggested that this may specify initiation of message translation. And as I mentioned already, UUG, AUG, CUG and to some extent GUG are recognized by formylmet~~hionine~~-tRNA; also that UAA and UAG may serve as terminators. It also appears likely that the words AG - ~~with~~ ending in U, C, A or G may also serve as special function words; but ~~if so~~ these functions have not ~~so far~~ been found. The present situation in this field is a most interesting one, in that the necessary tools for deciphering the special function words ~~are to hand~~, and it should soon be possible to understand more about the mechanism of these special words and the role they play in protein synthesis.

~~I would like to~~ turn ~~now to~~ a variation of the triplet-binding method, ~~which throws~~ further ~~light on~~ the coding mechanism. D. Hatfield has recently prepared some radioactive triplets, (in the earlier experiments it was the s-RNA which contained the radioactive tracer), and has studied the binding of these triplets to the ribosomes in the presence of the amino-acid s-RNA. Fig. shows both the binding of the triplet and of the s-RNA (here phenylalanine s-RNA) to the ribosome.

Fig.

~~As can be seen~~, in the presence of the appropriate triplet polynucleotide phenylalanine s-RNA binds to the ribosome; in the absence of the s-RNA very little triplet binds to the ribosome. Because of this, in the presence of the s-RNA both the triplet polynucleotide and the phenylalanine s-RNA bind to the ribosome at approximately the same rate. Thus the complex on the ribosome may well be a one-to-one association of triplet and s-RNA.

This technique provides a very simple and quite sensitive method for detecting codon recognition by s-RNA which is not acylated* with amino-acids. Thus some special function words may not be recongized by activating enzymes, s-RNA's, which are

---

*

not acylated, and this method would provide a relatively simple route towards detecting such recognition.

We have also made investigations (in collaboration with B. P. Docter and Walter Reed) with purified s-RNA fractions, i.e., media containing essentially only a single type of s-RNA, derived from E. coli fractions. We find that Tyrosine-s-RNA recognizes both UAC and UAU, which again exemplifies the C=U degeneracy in the third place. (There are two types of Tyrosine-s-RNA, differing in

; both types recognize UAC and UAU.) Similarly, Valine-s-RNA recognizes both GUA and GUG (G=A degeneracy) but the GUG to a much lesser extent than GUA. The E. coli fraction leucine-1-s-RNA and leucine-2-s-RNA both recognize the leucine codons (UUA, UUG, CUU, CUC, CUA, CUG). Recently, however, J. A. Carbon has reported that in mammalian liver one species of leucine-s-RNA preferentially recognizes AAG, and the other preferentially recognizes AAA. There are also types of leucine-s-RNA which recognize CUG, and others which recognize UUG.

The major variant of methionine-s-RNA which, as mentioned previously, will accept a formyl group recognizes UUG and CUG, but a less prominent methionine-s-RNA recognizes AUG preferentially. Likewise there is a Tryptophan s-RNA which recognizes UGG, CGG and to a smaller extent AGG. The pattern here is clear: a close relationship between U, C and A in the first place of

the coding triplet. R. Holley, working with purified fractions of yeast s-RNA, found alanine-s-RNA recognized GCU, GCC and GCA -- again the group U, C, or A but now in the third place of the coding triplet. It should also be pointed out prominent leucine-s-RNA binds to ribosomes very weakly in response to the nucleotide triplets; it is possible that this type of weak recognition involves only two of the three nucleotide basis in the triplet.

This work with pure fractions, such as ~~this~~ alanine-s-RNA prepared from yeast, can afford some further insight into the mechanism of codon recognition. This is especially ~~so~~ in this case since Holley and his collaborators have recently reported the sequence of bases in the alanine-s-RNA. ~~In~~ Fig. ~~is~~ shown the ~~variation of~~ binding of alanine-s-RNA to ribosomes ~~with~~ concentration of the s-RNA.

Fig.

The dotted line represents 100% binding, i.e., all the available s-RNA is bound to ribosome. This fraction of s-RNA, which Holley supplied to us, was estimated to be ~~greater~~ than 95% pure; and yet this s-RNA recognized quite well at least three of the alanine codons -- GCU, GCC, and GCA. It ~~did not~~ respond ~~or~~ only very slightly to GCG. (On the other hand, with unfractionated E. coli s-RNA, alanine-s-RNA responded quite well to GCG -- indeed this was the best alanine-s-RNA codon, and the response to GUU, GCC

and GCA was relatively weak.) Since the yeast extracted s-RNA fraction was ~~of high purity~~, the results strongly suggest that a single molecule of s-RNA can recognize alternatively at least three of the four alanine synonyms.

The whole sequence of the nucleotides in this alanine s-RNA are shown in Fig.

Fig.

The alanine amino-acid ~~in~~ linked to the terminal adenosine, and this is shown in the diagram in only one of the suggested possible conformations. There are several single-stranded regions of the s-RNA of possible interest. ~~There is a~~ The sequence: G, T, Ψ U, C (Ψ U is an isomer of U) ~~which sequence~~ has been found in virtually every s-RNA that has been examined. Another interesting sequence is the C, G, G, surrounded by two dihydrouridylic acids. A third is the IGC region (I=inosine) right in the middle of the s-RNA molecule. These latter two regions of interest are shown in more detail in Fig.

Fig.

If ~~these two~~ the triplets CGG and IGC were really the s-RNA anti-codons, that is, the nucleotide groups which recognized the nucleotide-triplet codons for alanine, recognition would be by parallel* pairing between C and U; and the G would then have to recognize

---

* [illegible]

U, C and A. If, however, base pairing were according to the Watson-Crick hydrogen-bonding, or anti-parallel scheme, C would pair with G, G with C and the inosine I in this position would base-pair with one of U, C or A, but not G. This latter is the pattern observed for the alanine code; and Crick has recently proposed a detailed mechanism which would permit hydrogen-bonding between I and U or C or A.

This mechanism, by which I recognizes U, C or A in the anti-codon - codon pairing, termed the "wobble" by Crick, involves a movement, at the end position of the triplet, of either the s-RNA or the m~~essenger~~-RNA on the ribosome. All the experimental results are, I believe, in accord with this type of recognition mechanism. The table shows the base-sequences in the

Table

s-RNA anti-codon and the corresponding base-sequences in the m~~essenger~~-RNA codon. Thus inosine in an end position in s-RNA can recognize by ~~alternate~~ base pairing U, C or A; a G in the end position of s-RNA could similarly recognize ~~alternately~~ C or U, and A could recognize U, C or G, and U could recognize by alternate pairing A or G. We would also predict on this model that a ribothymidylic acid-s-RNA would pair also A and G, (perhaps the interaction with A would be stronger than for a uridylic acid in s-RNA); that a ψU in s-RNA might recognize alternately A, G or U - a pattern that has been noticed rather

often with s-RNA.

Another possibility is that dihydrouridylic acid would not base pair (~~with the expected~~ complementary A), so that the interaction with the messenger would be a weak interaction; but it is also quite possible that a U or C in a terminal position would not greatly inhibit the interaction. A metal group on a 2'-hydroxyl deoxyribose (sugar) might also result in a weaker interaction, and furthermore, by permitting a greater freedom of motion on the ribosome, such a modification might result in greater ambiguity, i.e., lower specificity of the coding.

These results with infrequently occurring (or "trace") bases, and particularly those with Inosine, ~~rather~~ strongly suggest that s-RNA may be modified enzymatically, after it is released from the DNA template (where it is assembled in the cell). Since the level of "trace" bases is quite high in an actual cell, it seems likely that there exists a whole spectrum of intermediates, s-RNA's in various stages of successive modification. The consequences of this are ~~rather~~ easy to visualize. For example, if an adenine(A) in s-RNA is de-aminalid and so converted into an inosine(I), the A which would normally recognize the Uridylic acid base in the message would now be replaced by something (the I) which can recognize U, C or A. Similar interconversions would result from the de-amination of a C or the conver-

sion of G to I. It is possible, although perhaps rather premature to speculate, that this type of interconversion plays an important biological role.

~~There has certainly been~~ a great deal of work recently ~~which~~ suggests that ~~it~~ is possible in actual cells, ~~in some way to~~ modify the specificity of codon recognition, ~~and this is certainly something which could have~~ profound biological consequences. An example of this is the effect of the antibiotic streptomycin. ~~It has been shown, by~~ Davis, Gilbert and Gorini showed that streptomycin will bind on to the 30-s part of the ribosome (the small subunit), and all the available evidence suggests this binding of ~~the~~ streptomycin to the ribosome may in some way distort the topography of the codon recognition site so that greater ambiguity in codon recognition results. This may be one mechanism ~~for~~ the greater degree of error in protein synthesis, ~~although, of course, this may not be the only reason to account for the~~ action of streptomycin on bacterial cells.

There are other examples, in addition to streptomycin, of the modification of the specificity of codon recognition. A recent, comparative study ~~has been made, by~~ R. Marshall and T. Kaske made of the specificity of codon recognition with s-RNA from amphibian, Xenopus laevus, liver, from guinea pig liver and from E. coli. E. coli arginine-s-RNA does not recognize AGG and recognizes CGG only very slightly, whereas for both amphibian and mammalian

AGG and CCG (see Fig.)

Fig.

The contrast between alanine-sRNA's from yeast, mentioned earlier, and E.coli is also shown in this diagram.

In both amphibian liver and guinea-pig liver GCG is a very active codon, whereas in the amphibian liver GCG has no activity for alanine-sRNA. This contrasts with the activity for E.coli alanine-sRNA. In all species tested, AAA is recognized (by Lysine-sRNA), whereas AAG has only slight activity in E.coli although it is a very active codon in higher organisms. Sinne-sRNA recognition of UCG and of AGU and AGC is also variable, as indicated. Threonene recognition of ACG is likewise variable. We have, however, found no differences in the codon recogniton of sRNA's corresponding to aspartic acid, cytene, glutamic acid, histadine phenylamine, proline, tryptzine and valine.

~~I might mention~~, a somewhat different type of sRNA modification, in vivo, ~~which~~ we have studied in collaboration with N. Sueoka, ~~It is~~ observed ~~on~~ infection of bacterial E.coli cells with the virus, T2-phage, ~~that within one minute after infection~~ an enzyme (protein) is synthesized by the bacteria which modifies a pre-existing Leucine-sRNA component. (This sRNA is necessary for the biochemical machinery of the host bacterium but not for the virus.) The modification was such that it was technically possible to purify the modified sRNA and test it

for codon recognition. We found that it recognizes only poly-UG but it does not recognize any triplet. We have tested all the UG-triplets. One also finds that together with the modification of the s-RNA, there is a cessation of protein synthesis by the bacterial host. We do not understand the mechanism of this "turning-off", but we think it likely that enzyme produced by T-2 infection so modifies the leucine-sRNA component as to interfere with the host protein synthesis, and it does this without preventing the protein synthesis by the phage. This is a very subtle way of subverting the metabolism of a cell so that viral proteins can be synthesized in a large amount. This is a problem we are now investigating.

I trust I have shown, by the examples I have briefly sketched, how some features of the complex machinery for protein synthesis in cells can be studied by means of relatively much simpler systems, in vitro. Thus it has been established that the same sequences of three nucleotide bases code the same amino-acids throughout the whole range of organisms, from bacteria to mammalian livers. And this universal code has been explored by molecular biochemistry in vitro.

However, we have seen that there are secondary features, such as the relative responses to different synonym codons, and the subtle modifications of the sRNA's which can be of great importance in actual, complex living organisms. Features such as

may play important biological roles; by selectively controlling the rate of protein synthesis they may be an important factor in the general process of cell differentiation. These are certainly problems for the future.

Finally, I ~~would draw attention to the fact~~ that even, in vitro, at its simplest, the whole detailed process of coding in protein synthesis - involving DNA-m-RNA-s-RNA-ribosomes, activation enzymes, ATP, etc. is far from fully understood. Even the basic underlying questions - why, for example, does a triplet code of this sort exist, why should not phenylalanine instead of alanine correspond to GCU and GCC? Is there a basic chemical reason for this, or is it to some degree a matter of (historical) chance? My personal belief is that there is an underlying meaning for this and that it will be found.