Avery Gateway Dedication September 29, 1965

Dr. Robert W. Holley

Dr. Bronk, Dr. McCarty, ladies and gentlemen. It is a great honor to participate in this program this afternoon. It seems to me that the dedication of a memorial gate is especially fitting, that it's a particularly appropriate symbol, a particularly appropriate way of commemorating a man whose work has opened up such a vast world of new ideas and experimental opportunities. It's difficult for me to think of any piece of work that has been more important than that of Avery, MacLeod, and McCarty. There are very few experiments that even come close to equalling the consequences that this work has had in biochemistry, genetics, and biology in general. I think that my viewpoint is a little different than the earlier speakers partly because I am a little younger and partly because of my ignorance as an organic chemist for a while my knowledge of nucleic acids is post Avery, MacLeod, and McCarty. At the time when I really became interested in nucleic acids this was accepted, this was the truth, this was just background, something in history that everyone accepted. This makes me, as Dr. McCarty has indicated, a member of the very large group of workers who have come into the area of nucleic acids and related subjects after the period when the significance became clear. And I believe it is as a representative of this group that I speak this afternoon. There isn't time to consider all of the ways in which this work has been of tremendous significance. Pick just one area, that of the structure of nucleic acids, and indicate the present status in the area with particular reference to RNA, since that's the area in which the most of the work has gone recently. I want to describe very briefly the recent technical work on the structure of an RNA. The work I'm going to describe took a number of years, involved

several people. Those particularly involved were Gene Apgar, B. P. Doctor, George Everett, James Madison, Mark Marquise, Susan Merrill (?), John Penswick and Odis and Mere. The nucleic acids that we have worked with on structure have been the transfer RNAs, the relatively low molecular weight ribonucleic acids that carry specific amino acids to the site of protein synthesis. As obtained from cells, one gets a mixture then of transfer RNAs carrying all of the different amino acids, that is, at least 20 different RNAs. And the first problem, one that took us about half the time, actually, was to separate the mixture obtained from the cells. And the first slide indicates the results, using a technique -- counter-current distribution -- that needs no description here at The Rockefeller Institute. This shows the separation that is obtained in counter current distribution, starting with bulk yeast transfer-RNA. The solid line indicates the spreading of the material in the apparatus, the dash lines indicate positions of the different individual ribonucleic acids. By picking the most active fractions, putting them back in the apparatus, redistributing, it was possible to obtain three of these relatively homogeneous. The next slide indicates the results with the alanine transfer-RNA, using the counter-current distribution technique. The solid line indicates RNA, the dash line indicates activity for alanine, the dotted line is a calculated theoretical distribution curve. It is material of this sort, then, that was used in structural studies. Although this material had activity for no other amino acid, the possibility existed that it was a mixture of molecular species, so it was a gamble to work on its structure. We could be certain that if this was a mixture of molecular species, eventually we would find it was an insoluble problem. There seemed to be no alternative but to try. By increasing the scale of procedures we eventually obtained tens of milligrams rather than, as you see here, a peak

of 2 milligrams; so we have tens of milligrams to work with. Even so, the preparation was a continuing chore and we were always working with the minimum amount of material. In determining the chemical structure of this nucleic acid, we utilized about 1 gram of purified RNA, obtained from about 200 grams of bulk yeast RNA, and that, in turn, from about 300 pounds of commercial baker's yeast. The approach in determining the nucleotide sequence was to use degradated enzymes, ribonucleases, to cleave the RNA chain. The next slide summarizes the fragments obtained using pancreatic ribonuclease. For the nonspecialists here, a nucleic acid can be considered as a message written in English letters, if you please. The transfer-RNAs are the smallest biologically active nucleic acids known; they contain roughly 80 units, so what we are trying to determine is a sentence of 80 letters. The approach, then, is to break the sentence into words and determine the nature of the words, hoping to eventually be able to put the words back together into the sentence. These, then, would correspond to the words obtained by breaking the nucleic acid wherever pyrimidine nucleotides occur, c is cytidine, nicy is pseudouridine, u-uridine, a-adenosine and so forth. All of these fragments or words were obtained in approximately 80 to 95% yield, with the exception of two of them, actually one shown on here, the methyl-i pseudouridine, which was obtained in only about 50% of the theoretical amount. As we'll see this was obtained in a different word so we know it actually exists. The approach was to identify each one of these fragments, these words, and determine a sequence of letters. Composition alone was sufficient to determine sequence of the letters of the words containing only two letters, because this enzyme cleaves next to the pyrimidine, the c or the u. In the case of a larger word three nucleotides, one knows that pyrimidine is at the right, as we're writing on here, the question

then is whether this one, for example, is AGC or GAC. By complete degradation with an enzyme called snake venom phosphodiesterase one obtains the letter at the left end as a nucleoside; this one gave A, for example, and that gives the sequence of a trinucleotide. You'll note a number of unusual letters, complicated abbreviations for methylguanylic, dihydrouridylic, inosinic, and so forth; these are characteristic of the transfer ribonucleic acids. The larger word, such as one composed of 4 letters, it's easy enough to get the two ends; one has the question of the order of the letters in between. This can be obtained, in this case anyway, by degradation with ribonuclease T1 that cleaves next to G, so we had AC together and we knew the sequence as GGAC. One of these fragments is listed without a phosphate--the p refers to phosphate--that's cytidine, that comes from the right end of the molecule that would correspond to the period at the end of the sentence. As obtained from commerical baker's yeast transfer-RNAs mostly lack the terminal adenylic acid unit and terminate in cytidylic acid; that is the end of the molecule. There is an extra p on this word; it's the only one that has a p on the left. That would correspond to the capital letter at the beginning of the sentence, that's the left end of the molecule. So we know this is at the left end of the molecule. To determine the order of letters on anything as long as 8 nucleotides required a new method, which we developed as shown on the next slide. What we did was treat with venom phosphodiesterase to give a mixture, in which successive letters had been removed from the right end of this molecule. The presence of a phosphate on this end interferes with the enzyme so that was removed first, by alkaline phosphatase. Now, this mixture can be separated, as shown on the next slide. Where this is the starting material, this is after removal of a phosphate and then removal of one, two, three,

four, five nucleotide units. Now simply analysis for the terminal group, which can be obtained by alkaline hydrolysis gives us UGAGA corresponding to the sequence of those four, five groups. That gave us a set of letters, a set of words, corresponding to breaking sentence wherever a C or a U occurred. Another enzyme was used, ribonuclease T1, to break the sentence wherever G occurs. These are shown on the next slide. And you'll see each one of these ends in a G, with the exception of this one piece. That, then, is the right end of the molecule. This is where the period is again. And there is the left end as a guanosine diphosphate. Again, the identities of all of these fragments were determined. Nucleotide composition is analysis for the letters; it was sufficient for all of these. We had to determine the sequences of the others. The methods I've described were sufficient for determining these words for all except 3 of these. And a method indicated on the next slide was used for those. The method I describe with partial degradation is not good for the last 3 nucleotides. But micrococcal nuclease was shown by Dr. Laskowski, Sr., to leave the last 2 nucleotides untouched, as a dinucleotide is a combination of 2 letters. So that these remaining three sequences were completed using this enzyme. That gave us two sets of words corresponding to the sentence and now we wanted to put them together. Because ~~of the presence of~~ these minor nucleotides occur only once in the molecule, we had overlaps between the two sets of words, so we could put these together to some extent and this is shown on the next slide. Putting all the information together, we could conclude there were 77 letters in the sentence and they could be grouped to give us 14 sequences. Our problem

now was to arrange these 14 corresponding to the intact molecule. We tried various approaches. the approach that worked was to get large fragments that break the sentence into just a few pieces. The next slide shows results obtained by John Penswick who did this. very brief treatment of the RNA with ribonuclease T1, zero degrees, in the presence of magnesia. The mixture was chromatographed on diethylaminoethyl cellulose in the presence of 7 molar urea. This correspondes to the intact RNA untouched by the enzyme. This is still active. These two peaks were inactive, and they look superficially like the molecule has been cleaved in two. Each one of these could be a mixture, however. This had to be determined. Each of these peaks was purified and then analyzed, as shown on the next slide, by complete degradation to the words we'd already identified, using ribonuclease T1. This pattern corresponds to those words obtained with ribonuclease T1 where, with two or three exceptions, each one of these peaks represents a different word. If the molecule is split into two pieces then words should be found in one piece or the other and not in both, and that is what is observed. One and two are dinucleotides, there in fraction 2. Peak 3 is monoguanylic acid that would be expected to be in both parts. Four is in 2, 5 and 1, 6 and 7 in 2, 8 and 2. Peak 9 is actually a mixture of the two ends of the molecule, one end is found in 1, the other end in 2, and so on. So we had the words separated into two groups. it's like a jigzaw puzzle; if you could separate the pieces that go together into two groups rather than having them all together. If this worked to get two large pieces -- it seemed reasonable it might work to get several. as shown on the next slide, that was possible. This is again the pattern on complete digestion to the words that I had on the slide before. This is brief digestion with the enzyme; many of those peaks are

missing, and instead one has many large fragments containing combinations of words. Each one of these now could be analyzed, put back through the procedure to find which words were present in each. I'll run through that with just two of them, peak 16 and peak 18, and summarize the rest of the results. Next slide shows: taking peak 16, retreating with the enzyme to break to the words that we already knew, we obtained peak 12 and peak 9, traces of impurities. The identities are shown on the next slide. ~~Next slide please.~~ Peak 9 is the end of the molecule, the one corresponding to the period; peak 12 can go then only to the left of it, and we have two of those words put together. The next slide. Peak 18 turned out to be a mixture of things. They were separated by re-chromatography at 55°, so we had three large pieces to analyze, and these are shown on the following slides. Next slide. 18A contained 1 and 15. These are shown on the next slide. These two words - we knew that I was present in the sequence IGC. This fragment then could not go to the right here; there is no G to put in between; it must be to the left and we know that sequence. The next slide. 18B had three components; these are shown on the next slide. We knew already that the AG words were in this sequence; this word was not present in the fragment, therefore that has to go to the left, and we know that sequence. It's just putting pieces together one after another. The next slide. Finally, peak 18C gave us these four pieces, as shown on the following slide. We knew that end of the molecule that accounts for the PGP, the two G's; it accounts for the only C present, establishes that CG was there. We knew that u-methyl-G was in this sequence; since there was no additional C, the UG has to go to the left; we know that much. We have one G left. If that is put there or there we create a GGU sequence, which was not present in the molecule; that would show up in the pancreatic ribonuclease digest, therefore

the G has to go to the right of the methyl-G, and we know that sequence. The next slide summarizes all of the data. These are the four that I've just indicated. We had additional large fragments corresponding to these five sequences. ~~On~~ And the process~~es~~ is just like putting a jigsaw puzzle together; put one more piece on, another two pieces on, eventually you have only a piece or two left, and it is very easy to drop them into place. These correspond then to the two halves that were obtained. We know the period is over here, the capital letter over there. They can only go together in one way to give us the nucleotide sequence of the alanine transfer RNA isolated from yeast. Given the structure in these terms we're interested in questions of biological interest. One is, where in this molecule is the specificity for interaction with the messenger-RNA? It is the part called the anticodon that corresponds to alanine. Our indications would be it's not proved that it is the IGC sequence in the middle. One would like to know what parts of the molecule interact with the ribosome when the protein is formed, and we would suggest this region with the ~~Tseudeu-rydyllic, psychodyllic~~ T-pseudo-uridylic, cytidylic, and the region with the ~~dyhydrouse~~ dihydral-U's (?) may well interact in that way. One would like to know what this looks like in space. It isn't just stretched out as a line in solution. How is it twisted around? If one looks at sequences, you find that there are ~~very,~~ only short regions that are complementary in the Wat~~ts~~son and Crick sense. It is not possible to make long double-stranded regions; various ways one can have short regions are shown in the next slide. Where an attempt is made to line up one long double-stranded region, one with two double-stranded regions, the one with three looked very poor, the one with four looks quite good. My personal preference at this point is something like this. Through the cooperation of Robert Harte of the Americal Society of Biological Chemists, we've gotten enough of the

Society's models, the Adams models, that they are producing, to build a model of this structure. I'd hoped to be able to have a slide for it but time ran out. What is clear from the model is that when you build this structure you have something that is still very flexible. The different parts here can be twisted around in many ways. The way that appeals to us is to twist this part back over that one -- it's sort of like shaking hands -- the two parts fit together, that would give a region for interaction with the ribosome. This would be the anticodon for interaction with the messenger-RNA, and this is where the amino acid is carried. This is a very brief review of present status of structure of ribonucleic acids, the reason for working with transfer-RNAs is because they are relatively small, 77 nucleotides, viral RNAs are much larger, and it's of course a long way to a transforming DNA. However, considering the rate of progress in the field of structure of nucleic acids -- many laboratories all over the world making rapid progress -- I think it's a safe prediction that in perhaps 10 years we will be able to write out the complete chemical structure of something with transforming activity.

Thank you.