

CHEMICAL MODIFICATION OF NUCLEOTIDES AND ITS APPLICATION TO THE INVESTIGATION OF NUCLEIC ACIDS STRUCTURE

N. K. KOCHETKOV

*N. D. Zelinsky Institute of Organic Chemistry, Academy of Sciences, Moscow,
U.S.S.R.*

The chemical modification of biopolymers—alteration of their structure by the action of chemical reagents—is becoming more important as a tool of chemical and biological investigation. As is well-known in protein research this approach has very wide applications in determination of primary structures and enzymatic functions. Studies of the reactivity of nucleic acids and their components began much later, but nonetheless they have produced already some interesting results, and more important developments are believed to take place in the near future. Proteins and nucleic acids are highly specific polymers, and so the most interesting reactions leading to chemical modification would be those which are more or less specific. The specificity of nucleic acids depends on the nature of heterocyclic bases. Much research has been carried out in this field, which is the subject of this paper.

The final aim of chemical modification is the preparation of nucleic acids with changed structure and hence with changed properties, in order to solve structural and functional problems.

The reaction, which can be used for chemical modification should therefore fulfil several requirements: (i) it should be more or less specific, leading if possible to modification of only one heterocyclic base with a high yield; (ii) the transformation should be irreversible during the following chemical or biological investigation of modified nucleic acid; (iii) the conditions of heterocyclic nuclei modification should not be strong enough to cleave phosphodiester linkages.

At present there exist a number of reactions leading to the modification of heterocyclic bases. Most of them have been investigated only on mononucleosides or mononucleotides, but there are several examples of modification on a polymer level.

Table 1 briefly summarizes some of these reactions. It is seen that most of these reactions are not specific since two or more bases are changed. It should be mentioned that in many cases the specificity of the reactions was not completely checked. This group of reactions includes for example such long-known reactions as hydrazinolysis, apurination of DNA, action of nitrous acid, formaldehyde action, etc. In spite of lack of specificity these reactions are used very much for different studies, mainly in biology. The other group of reactions consists of reactions of high and sometimes strict

specificity. It includes the reaction with hydroxylamine, with its *O*-derivatives, with soluble carbodiimide, Girard reagent and so on.

The application of the method, which was investigated in nucleosides and nucleotides, to the modification of polymer nucleic acid is one of the most critical points of whole investigation. Difficulties involved are not only of method, but also the secondary structure of nucleic acid, especially as the

Table 1. Reactions used for nucleic acids chemical modification

	Reactions with bases					Minor bases	Reveisibility of reaction	Secondary structure influence	Phosphodiester bonds destruction
	A	G	C	U	T				
H ⁺	+	+	-	-	-	?	-	-	-
H ₂ NNH ₂	-	-	+	+	+	?	-	-	-
H ₂ NOH (pH 6)	-	-	+	-	-	partly attacked	-	+	-
H ₂ NOH (pH 10)	-	-	-	+	-	-	-	+	+
H ₂ NOR	-	-	+	-	-	-	-	+	-
H ₂ NNHCONH ₂	-	-	+	-	-	-	-	+	-
Cl(CH ₂) ₃ NCH ₂ CONNH ₂	-	-	+	-	-	-	-	+	-
Soluble carbodiimide	-	+	-	+	+	-	±	+	-
CH ₂ = CHCN	-	+	±	±	±	+	±	+	-
Alkylation	±	+	±	±	±	±	-	+	?
Perphthalic acid	+	-	+	-	-	?	+	+	-
Glyoxal	-	+	-	-	-	±	+	?	-
Halogens	-	+	+	+	+	±	-	±	+
HNO ₂	+	+	+	-	-	±	-	+	±
CH ₂ O	+	+	+	-	-	±	+	+	-

fixed one changes position rather seriously because of its influence on reactivity of heterocyclic bases. The presence of very labile phosphodiester bonding also causes trouble. In this connection, the reactions listed in the table have been applied mainly to monomers and there are so far only a few known examples of its application to polymer. They include the reaction of hydroxylamine and its *O*-derivatives and, to some extent, reaction with carbodiimide. One of the most extensively studied reactions is that of hydroxylaminolysis of nucleic acid. This is a very important example of the reactions of nucleic acid with nucleophilic agents and will be discussed in more detail later. Following the work of Schuster who observed the mutagenic effect of hydroxylamine it was proposed that this effect is due to reaction of hydroxylamine with nucleic acid. Later Schuster, Zillig and others, observed that hydroxylamine reacts with uridine and cytidine and is inactive towards purine bases; the reaction with cytosine being fastest at pH 6.0, while that with uracil being fastest at pH 10, as it is seen from Figure 1.

It follows that specificity of the reaction even in a narrow range of pH is not sufficient and attempts to obtain RNA with a strictly definite type of modification using these data were unsuccessful. However, it seemed possible to produce from this basis a highly specific method of modification leading either to breakdown of uracil or to alteration of cytosine nuclei. Such a

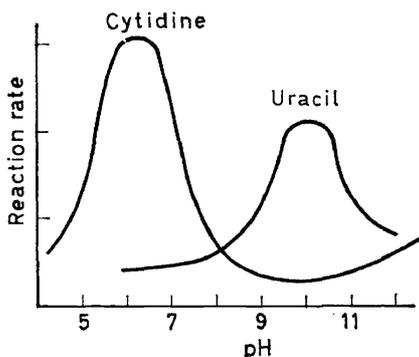
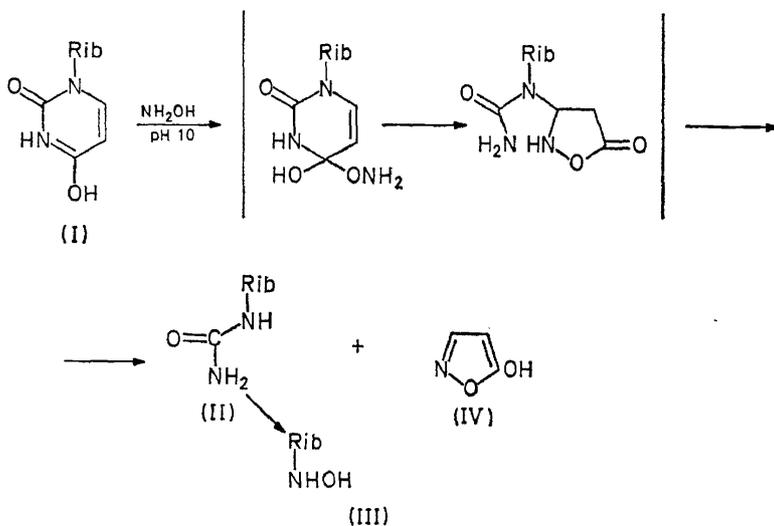


Figure 1. The rate of cytidine and uridine modification

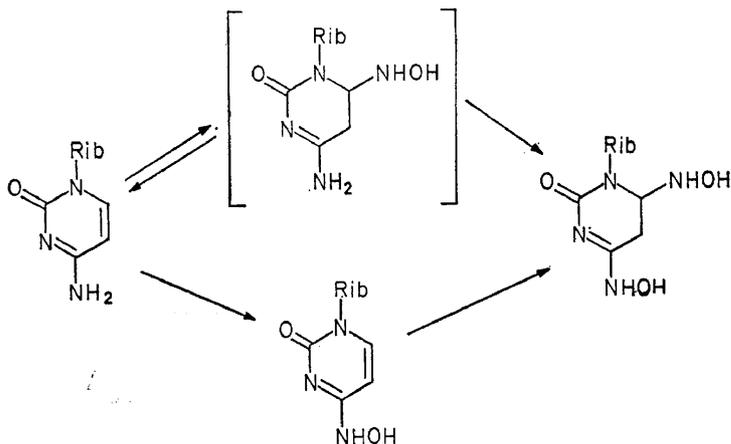
modified RNA could be useful for solving different chemical and biological questions which will be discussed later. The first aim of our work was the investigation of the uridine unit modification.

Uracil nucleus reacts with hydroxylamine according to the most probable scheme:



As can be seen, the uracil nucleus is attacked by hydroxylamine at C₄ with subsequent conversion of the resulting intermediate into ribosylurea (II) and isoxazolone (IV). Subsequent action of hydroxylamine converts ribosylurea into ribosyloxime (III). Hence in general the uridine residue could be transformed during hydroxylaminolysis of RNA into ribosylurea, ribosyloxime or free ribose, and the presence of all three types of residues is possible after modification.

At the same time the possible transformation of cytidine links should be taken into account. The most probable scheme of reaction is as follows.



Hydroxylamine molecule adds to the double bond of cytosine leading to non-aromatic 5,6-dihydro-6-hydroxylamino derivative. The second molecule of hydroxylamine attacks the carbon atom C₄ with formation of 4-oxime of 6-hydroxyamino-5,6-dihydrouracil. Uraciloxime, being also N₄-hydroxycytosine, can also be formed directly from cytosine by replacement of amino group.

As noted before the following use of modified RNA in chemical and biochemical experiments requires the preparation of the polynucleotide with a known structure of modified unit and with a known degree of modification which can be changed if necessary.

Optimal pH value (10.0) and concentration of hydroxylamine (10 M) leading to specific uracil modification with minimal effect on cytosine nuclei were found during model experiments on hydroxylaminolysis of corresponding nucleosides. The kinetics of uracil nuclei modification directly on RNA were then investigated.

The course of hydroxylaminolysis should be followed by appropriate methods of control of separate stages of process. As can be seen from the above reaction schemes, isoxazolone (IV) formation reflected the conversion of uridine units into ribosylurea residues and the rate of isoxazolone formation may serve for estimation of the rate of transformation I → II (*Figure 2*). This is true for the conditions chosen since in the I → II transformation the limiting step is the first intermolecular one.

The conversion ribosylurea → ribosylhydroxylamine results in incorporation of hydroxylamine into the polynucleotide chain, but the modification of cytidine also leads to incorporation of two moles of hydroxylamine as shown above.

Using data on kinetics of isoxazolone and bound hydroxylamine accumulation (*Figure 3*) we succeeded in calculating the rate constants for different steps of the uridine transformation. These were the constant for conversion to ribosylurea, the constant for conversion of ribosylurea into ribosylxime and rate constant for cytidine modification. We determined these constants of the main reactions for optimal conditions within temperature range 0–40°. The correctness of the assumptions on the mechanism and that of

CHEMICAL MODIFICATION OF NUCLEOTIDES

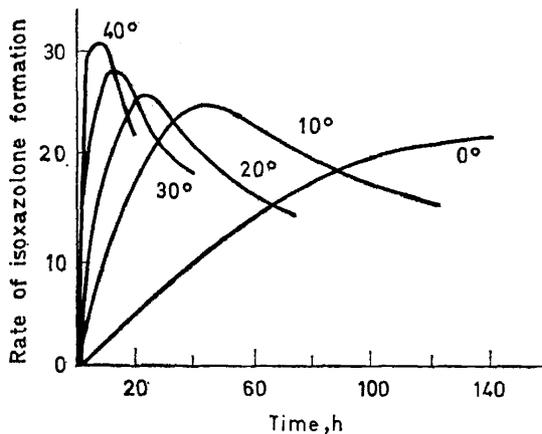


Figure 2. The isoxazolone accumulation during the reaction of RNA with 10 M NH_2OH at pH 10.0 (in moles/100 moles of nucleotides)

methods of analysis was confirmed by direct analysis of modified RNA samples: the experimental and calculated data were in good agreement. On the basis of results presented we could elaborate the method of preparation of deuridilic RNA without uridine units; these units may be converted to ribosylurea or if necessary to ribosyloxime. We succeeded in finding conditions for quantitative transformation of uridine units into both types. At 0°

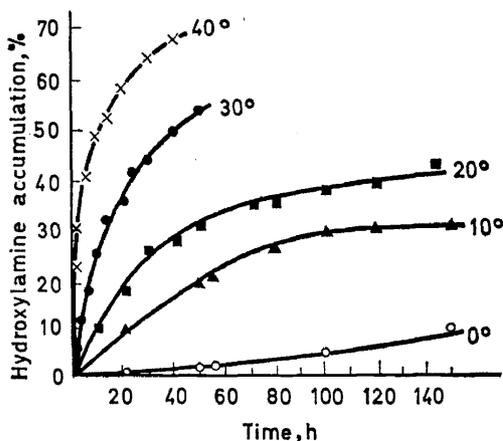


Figure 3. Accumulation of bound hydroxylamine during the reaction of RNA with 10 M NH_2OH at pH 10.0 (in moles/100 moles of nucleotides)

hydroxylaminolysis leads mainly to ribosylurea, while at 10°, as can be seen from Figure 4, conversion into ribosyloxime takes place. In fact, action of 5 M solution of hydroxylamine on high molecular RNA at 0° results in the formation of deuridilic RNA-I, containing ribosylurea residues instead of all uridine residues (Figure 5), the alteration of cytidine residues being only about 3 per cent. The action of 10 M hydroxylamine on ribosomal RNA at 10° gave rise to deuridilic RNA-II, containing ribosyloxime residues instead of uridine (Figure 6).

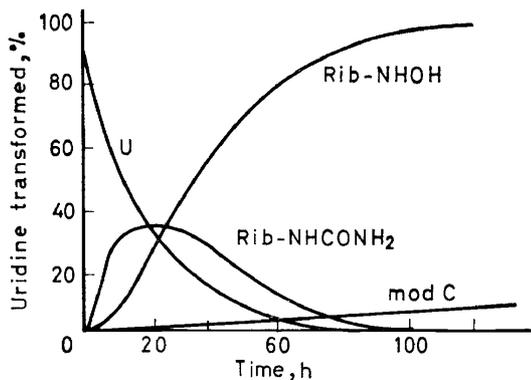


Figure 4. Concentration of uridine, ribosylurea, ribosylhydroxylamine and modified cytosine residues in RNA during the modification with 10 M NH_2OH at pH 10.0

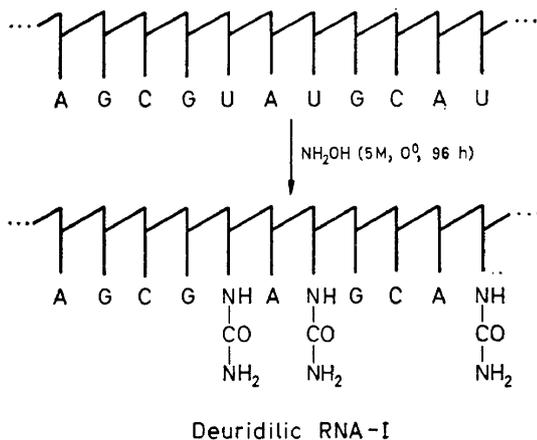


Figure 5

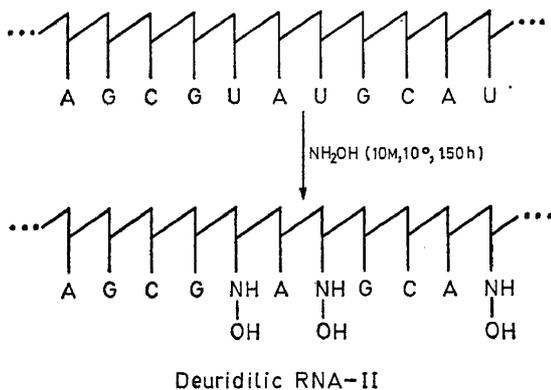


Figure 6

CHEMICAL MODIFICATION OF NUCLEOTIDES

As we suggested, these deuridilic RNA are of some interest for the study of RNA primary structure since with them it seems possible to develop a new highly selective degradation method. The stability of *N*-glycosidic bonds in ribosulreay residues of deuridilic RNA-I has practically no difference from that in uridine, and therefore any attempts to perform the selective chemical splitting of this modified biopolymer at the sites of former uridine units were unsuccessful. However, the modified uridine units in deuridilic RNA-I are not attacked by pyrimidyl-ribonuclease. It was proved first of all by lack of rupture of polyribosylurea-phosphate (obtained upon action of hydroxylamine on poly-U) after incubation with pyrimidyl-RNase at 20° for 24 hours, as can be seen from *Figure 7*.

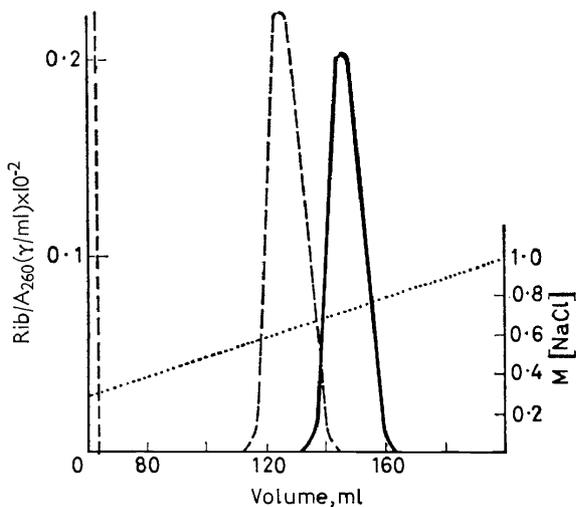


Figure 7

After this, the destruction of RNA itself was investigated and found to proceed very smoothly and specifically. The specificity of modified RNA rupture was confirmed by identification of only cytidilic acid in the mononucleotide fraction and by complete absence of ribosylurea phosphate. Finally, the analysis of terminal groups in oligonucleotides formed after digestion of modified RNA with RNA'se was performed. As can be seen in *Figure 8*, following splitting of these oligonucleotides with phosphomonoesterase and subsequent alkaline hydrolysis only cytidine was detected in nucleoside fraction. The ribosylurea was present only as a monophosphate. This means that splitting of deuridilic RNA-II took place exceptionally at the cytidine. This fact demonstrates the stability of phosphodiester linkage in ribosylurea-phosphate which is obviously of the same order as in usual nucleotide; no additional weakening in β -position to ribosylurea moiety takes place. It is in full agreement with the existence of glycosylurea entirely in cyclic form with no tautomeric transition into acyclic derivative.

Preparation of deuridilic RNA-II, containing ribosylloxime residues instead of uridine, opened another possibility for selective fragmentation of

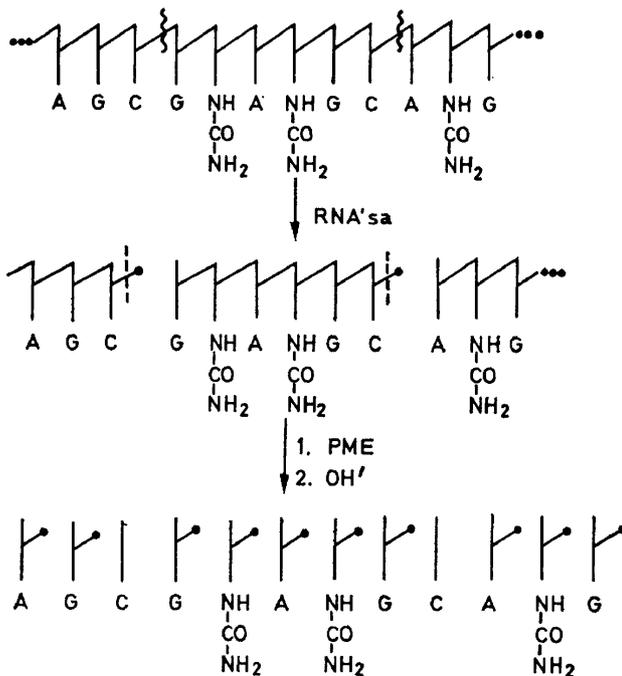
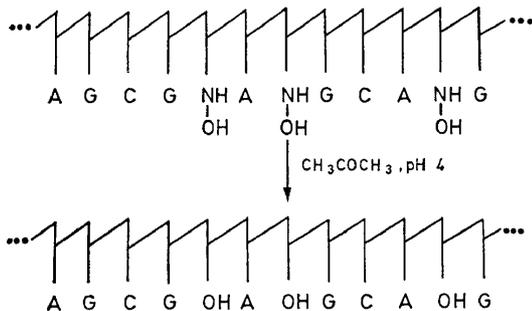


Figure 8

polynucleotide chain. It is clear that the presence of a carbonyl group or of some of its derivatives in β -position to the phosphodiester bond makes possible selective degradation here, involving β -elimination. Application of this principle to deuridilic RNA-II itself had no success, owing to the very slow degradation of the corresponding phosphodiester bondings. Therefore we elaborated the transformation of ribosyloxime residues in deuridilic RNA-II into ribosyl units by means of transoximation (Figure 9). This was carried out with nearly quantitative yield by treatment of deuridilic RNA-II with excess of acetone or cyclohexanone in an acetate buffer solution.



Deuridilic RNA - III

Figure 9

neighbouring group composition linked at 5'-position, while our approach gives grouping linked at 3'-position.

However the preparative-scale selective RNA degradation with isolation of oligonucleotides formed by this method seemed unlikely owing to rather drastic conditions of alkaline treatment.

In fact, such a reaction connected by side-processes, possibly by condensation of ribose or its transformation product makes impossible the elaboration of an analytical procedure in mild alkaline conditions.

Model experiments on ribose 3-phosphate destruction showed that aniline and especially *p*-anisidine are reagents of choice. The reaction proceeds under mild acid conditions (pH 4.5-5).

Under these conditions the degradation of deuridilic RNA-III resulted in formation of di-, tri- and tetranucleotide fraction comprising more than 60 per cent of total nucleotides (*Figure 11*). The degradation was completed after three hours.

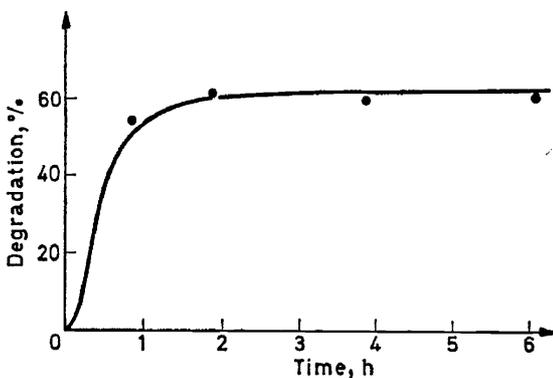


Figure 11

Special experiments have demonstrated that native RNA is unaffected under these conditions indicating the stability of usual phosphodiester bonds in this condition. The specificity of deuridilic RNA-III degradation followed from the fact that the spectrum of a fraction contained two-charged phosphate differed from that of all the oligonucleotide fractions and corresponded to the non-nucleotide substance, as shown in *Figure 12*.

According to the degradation scheme such a substance can arise only from deuridilic RNA-III rupture at ribose units. Additionally alkaline hydrolysis of the dinucleotide fraction resulted in formation of a neutral substance with non-nucleotidic ultraviolet spectrum. It also indicated that the degradation proceeds specifically only at ribose units.

All these data presented demonstrate the transformation of RNA to deuridilic RNA-III via deuridilic RNA-II followed by treatment with aromatic amines, to be a new method of selective fragmentation of the polynucleotide chain at uridine units. This is the first purely chemical selective method of nucleic acids destruction. At the present time we are working on the details of this procedure. Particularly, we are studying the nature of

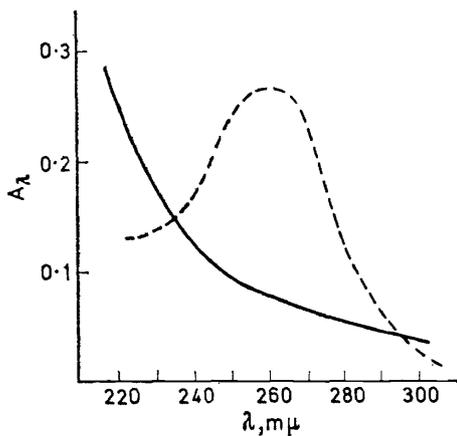


Figure 12

terminal residue in the degradation products since there is a possibility of formation of two ribose transformation products.

As indicated above, one of the most critical points in the elaboration of chemical modification methods is the possibility of application of reactions studied at monomer level to polymer. One should expect the reactivity of heterocyclic nuclei in the polymer to be decreased. Indeed, the rate constant of uracil nuclei modification in the high molecular weight RNA in the temperature interval 10–0°, where the secondary structure is practically absent, is half of that in uridine.

The detailed analysis of kinetic data, which is too lengthy for inclusion here, shows that at room temperature only non-cooperative stocking forces are involved in intramolecular interaction. But for lower temperature, say at 0° the base-pairing forces influence clearly appears and could be easily recognized by these cooperative characters. If so, this effect should be more important for the modification of RNA with stable double spiral regions. As can be seen from *Figure 13* the degree of uridine units modification in

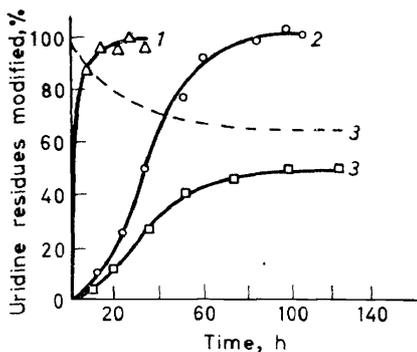


Figure 13. Modification of uridine residues in RNA by 10 M NH_2OH (pH 10.0) [1, at 37°; 2, at 10° in the presence of 7 M urea; 3, at 0°C. Broken line indicates change of hyperchromicity during modification at 0°C.

RNA sharply decreases with a decrease in temperature from 37° to 0°. At 0° only 40 per cent of the total uridine amount in tRNA was modified. At the same time the hyperchromic effect is lowered due to uracil nuclei cleavage, but not to secondary structure destruction. Obviously, in the case of a fixed secondary structure the modification of those uracil nuclei which are not paired with adenosine units by hydrogen bonds can take place. This

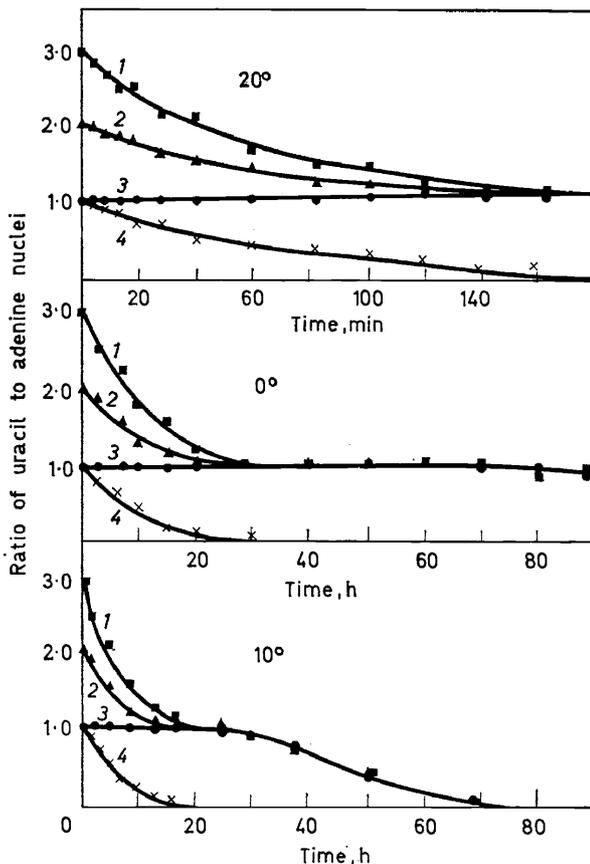


Figure 14. Modification of uracil nuclei in polyA-poly U mixtures. Ordinate ratio of uracil to adenine nuclei.

idea was confirmed by experiments on modification of tRNA under conditions when its secondary structure does not exist, namely by hydroxylaminolysis in 7 M urea.

The final evidence that the influence of the secondary structure on the reactivity of uridine units is connected with complementarity of these units with the corresponding adenosine units was obtained from the study of hydroxylaminolysis of poly-U: poly-A complexes. In Figure 14 one can see

CHEMICAL MODIFICATION OF NUCLEOTIDES

the course of hydroxylaminolysis of poly-U: poly-A mixtures with U to A ratio 1:1, 2:1 and 3:1. The excess of poly-U is subjected to rapid modification, while the 1:1 complex is modified only at a very slow rate. The reasons for the influence of formation of complementary pairs on the decreasing of reaction rate is not completely clear. Hydroxylaminolysis proceeds with initial nucleophilic attack of hydroxylamine anion of C₄ of uracil nucleus and hydrogen bonding with adenosine should increase the nucleophilicity at C₄ and hence increase the uracil nucleus reactivity. On the other hand, the secondary structure formation restricts the reagent penetration to the site of reaction. It seems likely that this steric effect predominates and the involving of uridine residues into spiral regions results in overall decrease of its reactivity.

The difference in reactivity of uridine residue in tRNA makes possible the use of such distinction for elucidation of some details of tRNA secondary structure. It should be noted especially that the degradation of uracil nuclei located in the non-spiral regions do not affect the secondary structure of tRNA and hence it is not altered during chemical modification. Thus the data on the reactivity of uridine units actually reflect the initially existing tRNA secondary structure.

It is known that about 25 per cent of the total nucleotides is located in the non-spiralized sections of tRNA. After estimating the amount of uridine units in these non-spiralized sections we were able to determine the nucleotide composition of spiralized and non-spiralized parts of tRNA. These data are summarized in *Table 2*. It can be seen that the spiral and non-spiral parts of tRNA differ substantially in nucleotide composition.

Table 2. Nucleotide composition of spiralized and non-spiralized parts tRNA at different temperatures

Nucleotides	Non-spiral regions		Spiral regions	
	moles/mole tRNA	%	moles/mole sRNA	%
U	5.4-8.0	37.0-55.0	9.2-6.6	63.0-45.0
A	2.9-5.5	18.8-35.8	12.5-9.9	81.2-64.2
G	3.9-1.3	19.0-6.3	16.6-19.2	81.0-93.7
C	4.7-2.1	22.5-9.9	16.6-19.2	77.5-90.1
A + G + C	11.5-8.9	20.2-15.5	45.7-48.3	79.8-84.5

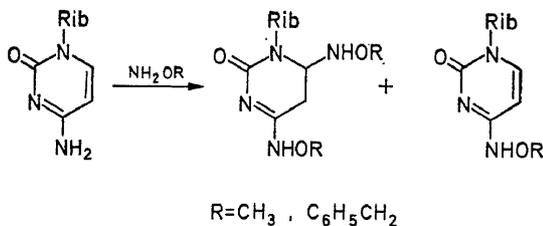
The influence of secondary structure on the reactivity of uridine residues presents new possibilities for preparative-scale selective tRNA modification. Hydroxylaminolysis of tRNA under mild conditions, say using 7 M hydroxylamine, at 0° during 96 hours, converts all the 'non-spiralized' uridines into ribosyloxime residues leading to tRNA modified at non-spiralized sections only.

Using such a partially modified tRNA prepared in this way, we can

perform the degradation at the sites of cytidine or uridine in loop section of this polymer by enzymatic or chemical methods mentioned before. This approach seems very promising for investigation of loop parts of tRNA.

As was mentioned before, we have investigated hydroxylaminolysis of not only uridine but also cytidine units in polynucleotide chain. This reaction also proceeds selectively under appropriate conditions. At the start of the study of this reaction we hoped to clarify some rather doubtful points in the matter of mutagenic effect of hydroxylamine, since this effect is connected with reaction of this reagent with cytidine moieties.

As the investigation of Schuster, Freese, Brown *et al.* shows, the reaction of hydroxylamine with cytidine gives rise to uridine oxime and 5,6-dihydro-6-hydroxylamino uridine oxime.

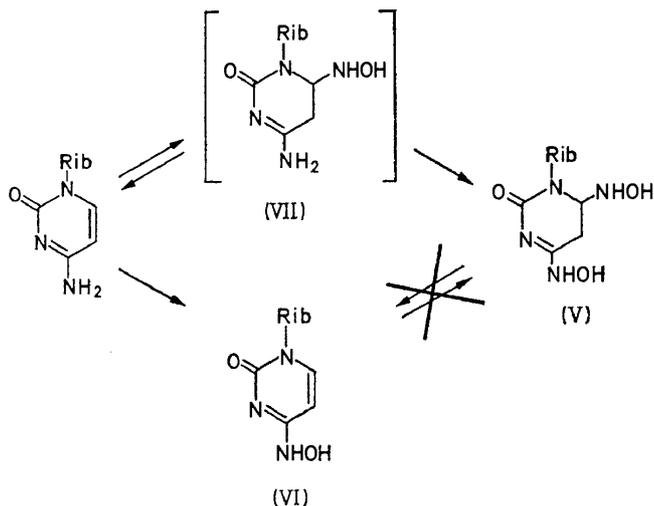


Consequently after action of hydroxylamine on nucleic acids, for example, DNA, cytidine units could be converted into one of these modified units. One could expect the uridine oxime residue, with aromatic character, to be responsible for mutagenesis due to possibility of complementary pair formation. At the same time dihydro-derivative lacking of aromatic character and of complementary pair formation possibility is responsible for inactivation effect of hydroxylamine. This suggestion was confirmed by direct experiments carried out by Budowsky, Sverdlov and Osterman, who used the modified cytidine triphosphates as precursors for RNA biosynthesis by DNA-dependent RNA-polymerase. It was shown that uridineoxime triphosphate incorporated like to cytidinetriphosphate while dihydrouridineoxime triphosphate was not incorporated in place of cytidine or uridine.

Until recently it remained obscure why the interrelation of mutagenic and inactivating effects of hydroxylamine depend on its concentration, in particular, why the inactivating effect decreased at hydroxylamine concentration more than 0.1 M.

These effects are clearer after explanation of the mechanism of reaction of cytidine with hydroxylamine, which is apparently a typical one for other cases of nucleophilic reagents action on cytosine nucleus. The detailed study of the reaction which is represented below showed the rate of reversible transformation V \rightleftharpoons VI to be very small and may not be taken into account and that the intermediate product VII is very unstable. Therefore, factually, the reaction proceeds in two directions and the true mechanism should be expressed as follows.

CHEMICAL MODIFICATION OF NUCLEOTIDES



It means that uridineoxime is formed directly from cytidine. Since the first step of the reaction in formation of dihydro-derivative is reversible, that ratio of the reaction products has to change depending on conditions: hydroxylamine concentration, temperature, pH values, etc. As can be seen in *Figure 15* the ratio of two reaction products VI:V increases with the

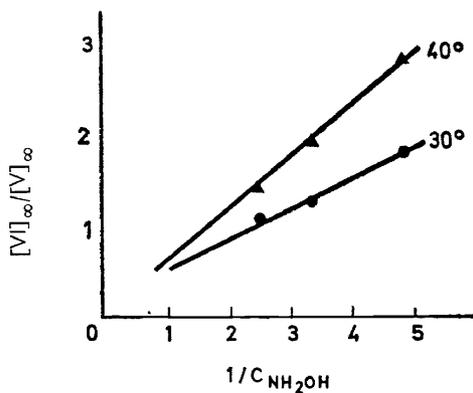


Figure 15. Dependence of the ratio VI/V from the reaction conditions

decrease of hydroxylamine concentration, leading to increase of inactivating effect of hydroxylamine with decrease of its concentration.

Comparing the mechanism of cytidine and uridine hydroxylaminolysis one can make another important conclusion. The main distinction is that the uracil nucleus is attacked by hydroxylamine anion, while cytosine is

attacked by nitrogen atom of hydroxylamine: in particular, the use of such a hydroxylamine derivative for which anion formation is difficult would result in selective attack at the cytosine nucleus. Therefore the reactions of nucleosides and nucleic acids with *O*-substituted hydroxylamines, namely *O*-methyl and *O*-benzylhydroxylamines, were investigated. These hydroxylamine derivatives react only with cytidine and no alteration of uridine and purine nucleotides were observed (cf. page 270).

This reaction is very close to that of cytidine with hydroxylamine, as follows from conditions, from the structure of products and from mechanism, which have been estimated on the kinetic measurements. The only essential difference is that *O*-alkylhydroxylamines react only with cytidine. These data led to elaboration of a highly specific reagent which is now widely used, particularly in genetic experiments. The reaction with *O*-alkylhydroxylamine was used also for nucleic acids modification. Using the conditions similar to those of hydroxylaminolysis the modified RNA could be easily obtained. At pH 6.0 the reaction is completed during 24–48 hours. Unfortunately all attempts to use this modified RNA for selective polynucleotide chain cleavage failed. It was found that the introduction of *O*-methyloxime grouping and even of more bulky *O*-benzyloxime grouping does not block the pyrimidyl-ribonuclease action at these units. The incubation of modified RNA with the enzyme leads to the formation of almost the same set of oligonucleotides as in the case of parent RNA. These and some other negative results confirm the suggestion on the dependence of pyrimidyl-ribonuclease specificity on the presence of some characteristic grouping in heterocyclic nuclei rather than on its geometry.

The story of finding specific action of *O*-alkylhydroxylamine indicates the great possibilities for search of other agents for specific modification of nucleic acids. One of the most promising approaches seems to be the detailed study of mechanism of action of known reagents. This allows choice of new derivatives with a greater selectivity. Of course the investigation of new reactions with pyrimidines and purines also is of great importance but unfortunately the chemistry of these heterocyclic systems is developing rather slowly. It seems that research into new specific reactions of heterocyclic bases, and the detailed study of its mechanism and kinetics is one of the most urgent aspects of nucleotides and nucleic acids chemistry. In the case of transfer RNA, the simplest member of this family, there arises a new very important problem—the study of minor components chemistry and the selection of corresponding specific reactions.

Finally mention should be made to put into perspective the method of chemical modification itself as an approach to the investigation of nucleic acids. It is already well-known, and also follows from data presented above, that this method seems to be of interest for study of RNA primary and secondary structures. I am convinced that its value will grow when we progress from simple transfer RNA to high molecular DNA and RNA. The successful utilization in these cases of the principles known from pioneering work of Holly seems rather doubtful. The elaboration of new approaches to the high molecular nucleic acids structure analysis is connected, in particular, with the hope of advanced use of electron microscopy as follows from Beer's works. In this case the chemical modification of bases should be of very

important application and factually should determine the accuracy and scope of this method. The specific mark of bases due to the action of selective reagent permits recognition by the electron microphotography with accuracy inherent in the chemical modification itself. Of course in this case new aspects and new very serious difficulties arise which cannot be discussed here but must be kept in mind during the search for new specific reactions, elaboration of methods of modification and investigation of mechanisms.