Russian Chemical Reviews **66** (4) 363 – 373 (1997) © 1997 Russian Academy of Sciences and Turpion Ltd

UDC 577.154.158;577.152.277

Chemical approaches to the study of template biosynthesis: general problems and the study of transcription

D G Knorre, N V Kudryashova, O I Lavrik

Contents

I.	Introduction	363
II.	Certain general problems of the template biosynthesis of nucleic acids	364
III.	Specific chemical approaches used in the study of nucleic acid polymerases	365
IV.	Principal stages of transcription. The subunit structure of RNA polymerases	367
V.	Interaction of RNA polymerase and its subunits with DNA	368
VI.	Investigation of the segment of DNA unwound in the complex with RNA polymerase (the bulge)	369
VII.	The study of the RNA exit corridor	369
VIII.	Approaches to the study of the active centres of RNA polymerases by the affinity modification method	370
IX.	The study of the interaction of eukaryotic transcription factors with DNA	372
X.	Conclusion	372

Abstract. The chemical approaches to the study of systems for template biosynthesis are examined. The general problems of template biosynthesis solved with the aid of these approaches and the principal results obtained as regards transcription are described. Attention is concentrated on the results obtained in the study of the interaction of RNA polymerase and its subunits with DNA and the RNA product. The approaches to the study of the active centre in RNA polymerases by the affinity modification method as well as the interaction of eukaryotic transcription factors with DNA are described. The bibliography includes 64 references.

I. Introduction

The enormous role of chemical methods in the study of proteins, nucleic acids, and their complexes, which perform the most refined functions in living organisms, including template biosynthesis systems, is well known and unquestionable. Chemical methods constitute the basis of many most important stages in the determination of the primary structures of proteins and nucleic acids. It is sufficient to mention the Edman procedure for the stepwise cleavage of peptides, the basically chemical method of Maxam – Gilbert which revolutionised the determination of the DNA nucleotide sequences, and the Sanger sequencing method based initially on the employment of chemically synthesised dideoxynucleoside-5′-triphosphates and subsequently automated by the introduction of fluorescent dyes either into terminators or primers. The development of convenient and effective methods for the chemical synthesis of polypeptides and oligonucleotides has played an inestimable role in the study of biopolymers and their functions. We shall not consider these two groups of chemical approaches, but it is understood that the primary structure of any protein and of any nucleic acid can be established and polypeptides or oligonucleotides of sufficient length (tens or several hundreds of monomer units) may be synthesised chemically when this is needed or when it is of interest to do so.

In the present review, attention is concentrated on the use of the chemical reactions of proteins and nucleic acids to study systems of template biosynthesis. Three groups of chemical methods will be dealt with.

The first group includes methods for the chemical cleavage of polymer chains under mild conditions. Naturally, any biopolymer can be disintegrated into fragments on sufficiently severe treatment. For example, RNA may be hydrolysed to mononucleotides in an alkaline medium. However, under conditions necessary for alkaline hydrolysis the tertiary structure of RNA is disrupted. Furthermore, in the review attention is concentrated on the application of chemical approaches to the study of functionally active biopolymers, and for this purpose one has to deal with a native biopolymer whose tertiary structure is disrupted to a minimum extent. Methods of cleavage under mild conditions, which do not distort the steric organisation of biopolymers and their complexes, will therefore be described and considered. Up to now these approaches have found extensive application only in the study of nucleic acids.

The second and third groups include methods for the chemical modification of biopolymers. Chemical modification is understood as chemical reactions which do not lead to the degradation of the main polymer chain. Two principal types of chemical modification are distinguished: the group-specific and affinity modification. In the first case, use is made of a reagent capable of modifying particular types of biopolymer residues, usually amino-acid side groups or the heterocyclic bases of nucleic acids.

In the second case, a ligand capable of specific interaction with the biopolymer, for example the substrate of any kind of enzyme, is provided with a reactive group. As a result of the affinity for a particular site of the native biopolymer, the reagent is fixed on this site and the chemical reaction takes place in the vicinity of the latter.[1,2]

Much has been hoped for of the chemical modification methods as regards the study of the structures and reactivities of biopolymers.

Thus, by studying with the aid of a group-specific reagent which of the biopolymer residues potentially capable of modification should actually be modified, it may be inferred which of them are in a position accessible to the reagent, *i.e.* which are concealed within the protein globule and which participate in the hydrogen bonding forming the secondary or tertiary structure of the nucleic acid. Furthermore, by comparing the degree of

D G Knorre, N V Kudryashova, O I Lavrik Novosibirsk Institute of Bioorganic Chemistry, Siberian Branch of the Russian Academy of Sciences, prosp. Akad. Lavrent'eva 8, 630090 Novosibirsk, Russian Federation. Fax (7-383)235 16 65. Tel. (7-383) 235 64 41 (D G Knorre)

Received 17 January 1997
Uspekhi Khimii **66** (4) 401 – 413 (1997); translated by A K Grzybowski

biopolymer modification by such a reagent with the degree of loss of biological activity, the involvement of the modified residues in the functioning of the biopolymer may be inferred.

By employing affinity reagents, the modified residue may be identified. It was assumed that this is one of the residues forming the active centre of the biopolymer or is at least in the vicinity of the active centre.

In order to investigate systems consisting of several biopolymer molecules (subunits), bifunctional chemical reagents, capable of reacting with the two functional groups of the biopolymers, may be used. This led to the possibility of obtaining information concerning the subunit areas which are at a sufficiently close distance from one another, permitting one molecule of the reagent to react with both polymeric components (to cross-link them into a single structure).

However, during the last decade, many changes have occurred in physicochemical biology which have made it necessary to correct our view on the possibilities and principal tasks of the chemical modification methods. Thus the tertiary structures of many enzymes, including those participating in template biosynthesis, have been established by X-ray diffraction. In certain cases, the structures of complexes of proteins with specific ligands or their analogues were determined, *i.e.* data were obtained concerning the disposition of the ligand (substrate) on the protein. The importance of individual aminoacid residues in the functioning of the protein can be readily established with the aid of site-specific mutagenesis — the specific replacement by genetic engineering methods of a particular aminoacid residue by any other out of the 20 residues constituting proteins. Therefore, in applying chemical approaches to the study of systems of template biosynthesis, it is necessary to take into account the current possibilities of both direct physical methods and genetic engineering.

Together with the exposition of the main data obtained with the aid of chemical methods for the study of RNA and DNA polymerases, an attempt will therefore be made to examine which principal possibilities and advantages are provided by chemical approaches in the study and practical employment of systems of template biosynthesis. However, we shall deal initially with the general problems faced by investigators in the field of template biocatalysis.

II. Certain general problems of the template biosynthesis of nucleic acids †

Nucleic acid polymerases possess a set of features of many other enzymes. Like any other enzymes, they act on specific substrates, which are in this case the growing polynucleotide chain and nucleoside 5′-triphosphates (henceforth abbreviated in the general case and for ribonucleotides as NTP and for deoxyribonucleotides as dNTP).

Polymerases have an active centre which catalyses the transferase reaction, the transfer of the nucleotide residue of the NTP from the pyrophosphate to the 3′-OH group of the growing polynucleotide chain. In terms of the single-letter nucleotide nomenclature, this reaction can be formulated as follows:

$$-\mathrm{pX}_{i-2}\mathrm{pX}_{i-1}\mathrm{pX}_i + \mathrm{pppX}_{i+1} \longrightarrow -\mathrm{pX}_{i-2}\mathrm{pX}_{i-1}\mathrm{pX}_i\mathrm{pX}_{i+1} + \mathrm{pp}, \quad (1)$$

where X is the nucleoside, i the index designating the position of the nucleoside in the polynucleotide chain, and p the phosphoric acid residue.

However, there are many specific aspects characteristic either solely or predominantly of enzymes of template biosynthesis. These aspects are apparently characteristic also of the biosynthesis of proteins on ribosomes.

Firstly, the normal function of polymerases of nucleic acid requires the presence of a template with the aid of which one out of the four NTPs present in the system is selected at each step of chain lengthening. Four types of nucleic acid polymerases, differing in the chemical nature of the template and in the substrates, are known. DNA polymerases catalyse the template biosynthesis of polydeoxyribonucleotide chains, using either DNA (DNA-dependent DNA polymerases, usually called simply DNA polymerases) or RNA (RNA-dependent DNA polymerases or reverse transcriptases) as the template. RNA polymerases synthesise polyribonucleotides using either DNA (DNA-dependent RNA polymerases or simply RNA polymerases) or RNA (RNA-dependent RNA polymerases or RNA replicases) as the template. The processes catalysed by these enzymes are referred to simply as replication, reverse transcription, transcription, and RNA replication. DNA and RNA polymerases are characteristic of all living organisms without exception, and these enzymes have already been the subject of intensive investigations, by procedures including chemical methods, in the course of many years. Vigorous studies on the reverse transcriptases present as the obligatory component of retroviruses, which include the HIV-1 virus responsible for AIDS, have been initiated in recent years.

The selection of the required monomer at each stage of the lengthening of the polynucleotide chain via reaction (1) must serve as the signal for the switching on of the catalytic centre effecting the transfer of the α-phosphate of the nucleotide residue to the 3′-OH group of the growing chain. This may be associated with the more refined adjustment of the structures of the substrates to the groups of the catalytic centre of the enzyme in response to the recognition of the correct substrate.

Secondly, at each chain lengthening stage one nucleotide base of the template, working as an information element, is involved in the selection process. The active centre of the enzyme consists of three main subsites (Fig. 1): the site accomodating the information unit of the template, ensuring the selection of the monomer (A) required at the given stage, the site in which the substrate is bound in its prereaction form (B), and the site accommodating the 3′-terminal unit of the growing chain to which the new nucleotide residue (C) is attached. In this scheme, the selection step takes place at the junction between sites A and B. If it is assumed that a certain time after the addition of the nucleotide unit the participants in the process remain in their places, the following situation is created: site C is occupied by the penultimate nucleotide residue of the lengthened chain; site A is occupied by the information element already employed; site B is occupied by the 3′-terminal

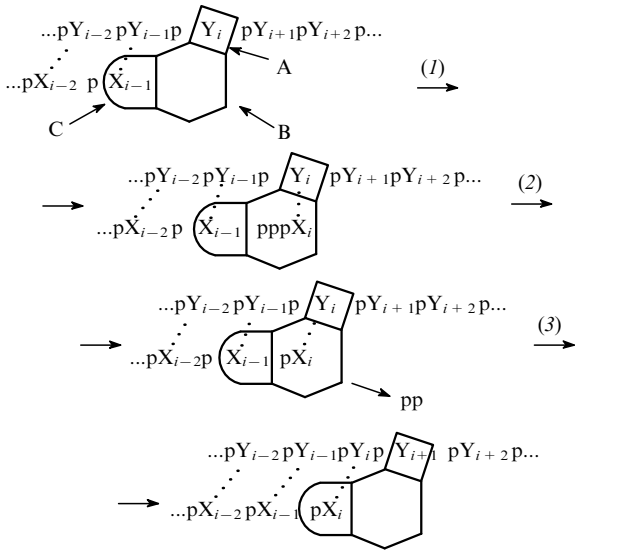


Figure 1. Processes occurring in the active centre of a template enzyme on elongation: Y is information elements of the template, X is monomer units of the growing chain; pppX$_i$ is free monomer; A is binding section for the information element of the template, B is monomer binding section, C is binding section for the end of the growing chain. (*1*) Selection of the monomer; (*2*) formation of a chemical bond; (*3*) translocation.

† The general problems of the template biosynthesis of nucleic acids have been examined by Knorre and Myzina.[3]

unit of the lengthened chain. It is reasonable to assume that all three domains of the enzyme are unique. In this case, the continuation of the synthesis requires the specific displacement of the template – growing chain system relative to the enzyme by one unit — translocation. The process can occur either synchronously with the chemical reaction or can follow it.

Thirdly, the initiation of the synthesis of new polynucleotide chains begins from a definite start point. The method of specifying this start point is different for polymerases of various types. In the case of transcription, the RNA polymerase must be located, before initiation, on a particular section of the DNA template, usually outside the region where transcription should occur. This section is referred to as the promoter if it is not too far (within the limits of tens of nucleotide pairs from the start point) or as the enhancer, as is the case for eukaryotes. There is, in addition to the promoter, a centre significant for the initiation of transcription located several hundreds or even thousands of nucleotide pairs from the start point. A random collision of the enzyme with the promoter appears to be a very improbable event. There is therefore the problem of the search for the promoter (or enhancer) by the enzyme or a special auxiliary protein — the transcription factor, which subsequently assists in the correct placing of the RNA polymerase near the start point.

It is noteworthy that in all the cases investigated initiation occurs only with the aid of RNA polymerases. Not one instance of initiation with the aid of DNA polymerase is known hitherto. These enzymes are capable only of lengthening an already existing chain. The synthesis of new DNA molecules in the cells therefore begins with the synthesis by a special enzyme of a short ribooligonucleotide, which is subsequently lengthened by the DNA polymerase, apparently this time with the aid of a second active centre. In the experiment, this can be achieved by introducing into the system a short deoxyribooligonucleotide complementary to the template. This oligomer can be subsequently lengthened with the aid of DNA polymerase. Such oligonucleotides have been called primers and the RNA polymerase generating the RNA primer for the subsequent synthesis of DNA is referred to as the primase. After the synthesis of the short RNA fragment by the primase, the growing end of the chain is switched to work with the DNA polymerase. This refers to cell enzymes. For enzymes the synthesis of which is programmed by viral nucleic acids, there are alternative paths. The initiation of the synthesis of DNA with the aid of the reverse transcriptase of retroviruses is being most intensively investigated. In this case, the genome viral RNA captures the primer from the cell, the role of the primer being assumed by one of the tRNA of the host cell, the 3′-end of which is complementary to a definite section of the genome RNA (the primer binding section), whereupon the reverse transcription begins with the attachment of the dNTP residue to the 3′-terminal ribonucleotide of the primer.

All the aspects of the functioning of polymerases enumerated — the adjustment of the selected substrate with the aid of the information element to the catalytic centre, translocation, the search for start points, and the directed switching of the template — product complex from one active centre to another — are dynamic events. Although dynamic elements are present also in the functioning of other types of enzymes, one can claim confidently that the role of dynamics in the functioning of template enzymes is especially important.

The second specific feature of template biosynthesis (especially characteristic of processes in eukaryotes) is that transcription and replication occur in a highly organised multicomponent system. In the case of eukaryotes, such a system is chromatin. In both transcription and replication, several components participate directly in the biosynthesis of nucleic acids. These must be combined in a single complex to effect the process. For example, transcription precedes the formation of a complex between the template and a set of transcription factors. The processivity of replication (the occurrence of a large number of elongation stages without the dissociation of the enzyme – substrate complex) is ensured by the presence of special clamping proteins which prevent dissociation. The knowledge of the structure of these complexes and of the dynamics of their changes during the functioning of the template system is extremely important for the understanding of the mechanism of the functioning of nucleic acid polymerases under the conditions in the living organism.

Yet another characteristic feature of template processes is that template systems must involve a template having an arbitrary primary structure. RNA polymerase must interact with an arbitrary transcribable sequence, which moves along the enzyme in the course of transcription. There are grounds for the assumption that the RNA synthesised remains for some time in contact with the enzyme. If this is so, then a certain 'exit corridor' for the transcription product should exist on the enzyme (Fig. 2, section *f*). DNA polymerase interacts with any template susceptible to replication, moving along it in such a way that only the information residue of the template and the 3′-end of the product are located in the active centre. It may be expected that a segment of the as yet nonreplicated template as well as a section of the duplex produced are in contact with the enzyme over a certain distance. The region of contact with the latter can also be referred to as the exit corridor. The problem of the determination of the region of the nonspecific contact between the enzyme, on the one hand, and the template and the polymeric product, on the other, therefore exists for template systems.

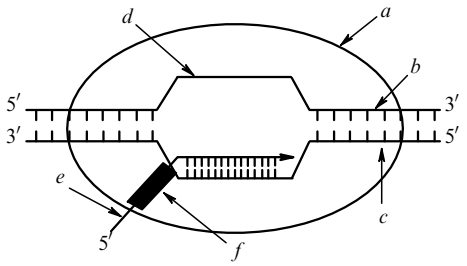


Figure 2. Schematic illustration of the interaction of the chains of the template and the product during transcription: (*a*) RNA polymerase; (*b*) nontranscribed DNA strand; (*c*) transcribed DNA strand; (*d*) temporarily unwound segment of DNA (bulge); (*e*) product RNA (transcript); (*f*) product RNA exit corridor.

By virtue of the length of nucleic acids and the presence as a rule of 'singular points' on the molecules of nucleic acids with the aid of which a definite catalytic step (initiation of transcription, onset of the reparation process) should take place, the problem of the search for these singular points arises — for example, the search for the promoter by RNA polymerase, the search for the damaged point by reparation enzymes, *etc*. In this connection, yet another problem of template biosynthesis is the determination of the mechanism of the search for such singular points.

III. Specific chemical approaches used in the study of nucleic acid polymerases

The majority of the problems arising in the study of template biosynthesis required the knowledge of the region of contact between nucleic acid and other components of the system, in the first place proteins. The region of contact is to a large extent protected against the influence of external factors, in particular enzymes and chemical reagents. When any kind of agent acts on the free biopolymer and on the polymer involved in a specific complex, it is in principle always possible to compare regions modified in the presence or absence of other components of the system and hence to determine which region is covered by these components. However, in the current state of experimental methods such a problem can be readily solved only for nucleic acids. The nucleic acid may be labelled at one of its ends. If it is treated by an endonuclease or a reagent cleaving the polynucleo-

tide chain at the modification point and the process is carried out to an extent such that on average there is one break for each chain, then the mixture obtained after separation by gel electrophoresis gives rise to a series of bands corresponding to cleavage products of different lengths (counting from the labelled end). When highly radioactive phosphorus or an intense fluorescent label is used, only fragments which have retained the labelled end are determined on the gel. If the nucleic acid has no tertiary structure, a series of bands corresponding to all the imaginable break points within the limits of the resolution of the electrophoresis are obtained. On the other hand, if during treatment the nucleic acid forms a complex with a protein or several proteins, then, within the limits of the region of contact, no cleavage occurs or at least it is not very pronounced. In this case, a 'white' spot, *i.e.* a region without radioactivity or a nonfluorescent region, appears on a series of bands. This spot is the imprint of the segment bound to the nucleic acid and is referred to as the footprint. Naturally, the effects of reagents used should not influence the stability of the complex investigated.

In footprinting, it is desirable that the specificity of the cleavage reaction in relation to heterocycles should be low. It is noteworthy that, together with chemical reagents, enzymes hydrolysing nucleic acids are frequently employed for footprinting. For this purpose, use is most often made of DNase I, which catalyses the cleavage of the inner phosphodiester bonds, including those in double-stranded DNA, leading ultimately to the formation of a set of relatively short oligonucleotides, as well as exonuclease III from *E. coli*, which catalyses the stepwise splitting off of the 5′-nucleotides from the 3′-ends of both strands of the double-stranded DNA. However, the resolution of the methods based on enzymic cleavage is restricted by the size of the enzyme (the larger the enzyme, the smaller the resolution), which in the case of DNA polymerase I is ~ 4 nm in diameter.[4]

A significantly better resolution is achieved when chemical reagents of small size are employed. The free radical $HO^{\bullet}$, generated in the $Fe(EDTA)^{2-}/H_2O_2$ system, has recently come to be widely used for the cleavage of DNA.[5] A specific feature of this method is that the reactions proceed via deoxyribose, which makes the method more specific in relation to the sequence of nucleotides. Footprinting with the aid of $KMnO_4$ is also used to investigate polymerases.[6]

In principle, it is possible to obtain a similar footprint of the protein molecule interacting with nucleic acid. Until recently, such studies were held back by the lack of effective methods for the determination of the points of modification of the protein, which made this problem extremely difficult. Very recently, a study has been published[7] in which a highly radioactive label was introduced into the transcription factor TFIIB by attachment to the C-end of the pentapeptide 'tail' — arginylarginylalanylseryl-valine, which is the target for phosphorylation with the aid of heart muscle protein kinase. This made it possible to introduce the [^{32}P] phosphate residue at the OH group of serine and then to carry out single-impact hydrolysis of the protein with the aid of weakly specific proteases with subsequent electrophoretic separation of the set of peptides obtained. As in the footprinting of nucleic acids, only peptides containing the C-terminal fragment and differing in their lengths appeared on the radioautograph. Comparison of the set of peptides obtained from the initial transcription factor and its complex with yet another factor TAF made it possible to identify the region protected by this factor against proteolytic cleavage by TFIIB.[7]

The existence of the elongation stage leads to interesting possibilities both in the methods whereby affinity modification can be carried out and in the design of reagents for affinity modification. Among the modification procedures, the method which has been called highly specific or competent labelling is especially interesting.[8] Nucleotides and their derivatives are in many cases capable of interacting not only with the active centre of the enzyme but also with a set of other regions on the surface of the protein. Cases are known where a large number of molecules of a reactive nucleotide derivative are bound to the enzyme molecule and such interaction is not always accompanied by the loss of enzyme activity. For example, the photoaffinity modification of phenylalanyl-tRNA synthetase from *E. coli* by ATP γ-*p*-azidoanilide can lead to the addition of 20 molecules of the ATP analogue without loss of enzyme activity.[9] In template biosynthesis, there is a possibility of the situation where the addition of a nucleotide to a protein does not prevent its participation in the elongation process. In this case by carrying out in the first stage of the experiment the modification by a nonlabelled reagent, designed as an affinity reagent, and then adding the next NTP carrying a radioactive label, for example α-^{32}P, it is possible to discriminate the modification point in the region of the active centre, because only the residue of the reagent attached covalently in this region undergoes elongation and contains the radioactive label. It must be emphasised that not all the residues of the affinity reagent attached to the enzyme in the region of the active centre are detected by this method because, if the modification involved the residue participating in the catalytic step, elongation cannot occur owing to the loss of catalytic activity. However, it is clear that no residues bound outside the active centre of the enzyme, even at functionally significant points, can be detected.

In the template biosynthesis of nucleic acids, the participants in the process are the template nucleic acid, the newly synthesised nucleic acid chain, and the corresponding NTP. The first two participants are polymeric molecules. The corresponding affinity reagents may be generated by synthesising the corresponding oligonucleotide fragments modelling these polymers, or by the modification of the polymers themselves. However, it is possible to introduce chemically reactive groups also during the elongation process. Derivatives of nucleoside 5′-triphosphates carrying reactive groups are used as substrates for such purposes. An example of a substrate of this kind for RNA polymerases is 5-(4-azidophenacylthio)-CTP.[10] In this case, the elongation product contains a reactive group template and is already an affinity reagent in relation to the template enzyme which was used in its synthesis. The substrate which participates in the elongation step but, being incorporated in the 3′-end of the chain, cannot combine with the next nucleotide unit is optimal. It is also possible to use reactive derivatives of terminators, for example dideoxynucleoside 5′-triphosphate in the biosynthesis of DNA. In this case, the reactive group is automatically fixed in the region of the active centre of the enzyme.

As already noted, it has been hoped that affinity modification is a method which will make it possible to determine the amino-acid residues in the active centre of the enzyme and in its vicinity. However, for this purpose it was necessary to know how to locate the modified residue on the polypeptide chain. The most direct approach consists in the cleavage of the modified protein at specific points, the identification in the set of peptides obtained that which contains the modified residue, and the analysis of this peptide by stepwise cleavage with the aid of the Edman method. The appearance in the course of this procedure of an unusual derivative or the arrest of cleavage at a particular unit indicates that this is the point where modification has occurred. The nature of the modification may be inferred provided that the product of the interaction of the modified residue with the reagent employed is known (from independent experiments).

In a number of instances, it is possible to apply indirect procedures whereby one can deduce the modification point with a particular degree of reliability. Such an approach has been proposed and widely developed in the studies of Grachev and coworkers[8] in relation to the location of the modified lysine and histidine residues in RNA polymerase by phosphorylating reagents. The properties of the phosphoramides formed on phosphorylation of the aliphatic ε-amino-group of lysine and of the imidazole ring of histidine differ quite sharply. Both phosphoramides are labile in an acid medium, but the kinetic characteristics of the hydrolysis are different. By measuring the rate of elimination of the attached label (demodification), it is possible to

compare the kinetic constants obtained with those measured for model compounds.‡ ATP γ-*N*-methylamide and ATP γ-imidazolide have been used as such compounds. The rate constants for their cleavage at pH 4.9 are 0.014 min^{-1} at 50 °C for the amide and 0.016 min^{-1} at 20 °C for the imidazolide, *i.e.* they differ quite sharply. After the determination of the peptide in which the modified residue is located, it is possible to determine from the primary structure of this peptide how many aminoacid residues are present in it and at what positions. If a residue of this kind, for example a histidine residue, is the only one in this peptide, then the location of the modification point is unambiguous. When the modified protein is cleaved into fragments and when particular procedures are used to separate the mixture of peptides, one must naturally bear in mind the stability of the modification product. For example, it is clear that in the case examined one cannot use proteases acting in an acid medium for the cleavage of the protein.

The use of photoaffinity modification by bromodeoxyuridine (brdU) residues may be quoted as yet another approach. Bromouracil is known to react photochemically with cysteine, tyrosine, phenylalanine, and histidine residues.[11] By employing this property, Skiba *et al.*[12] investigated the modification of the *E. coli*, RNA polymerase by the decanucleotide T_5brdUT_4 labelled with ^{32}P. Analysis of the peptides obtained by successive cleavage by bromocyanogen at the methionine residue and by trypsin hydrolysis at the lysine and arginine residues showed that the label is contained in the 721–731 peptide of the β-subunit and in the 984–990 peptide of the β′-subunit. Since each contains one aromatic aminoacid residue, the authors believe that the modification points are Tyr-726 and Phe-988 respectively.[12]

The above example demonstrates how important it is to have information about the chemical nature of the products of the modification of the aminoacid residues. Unfortunately, the modification products have been at present quite inadequately investigated for the most promising approach on the photoinduced modification, in particular for the widely employed method based on direct photocross-links between nucleic acids or nucleotides and the protein.

IV. Principal stages of transcription. The subunit structure of RNA polymerases §

So far, the main object of the detailed study of the mechanism of the functioning of RNA polymerases has remained the enzyme isolated from *E. coli*. From this example, it is easier to elucidate what can be obtained and what may be expected in the future from chemical approaches to the study of transcription. A few words must therefore be said about the enzyme itself and about the phenomenology of its functioning.

The main enzyme (core) consists of three types of subunits — α, β, and β′. The first of these is present in the form of two copies, so that the subunit structure of the enzyme is formulated as $\alpha_2\beta\beta'$. In the transcription initiation stage, the regulatory σ-subunit also enters into the composition of the enzyme, as a result of which the whole enzyme (the holoenzyme) $\alpha_2\beta\beta'\sigma$ functions. Only then is the correct complex between the enzyme and DNA formed in the region of the promoter, which in fact determines the point of the onset of transcription.

It is appropriate to mention that the numerals assigned to DNA nucleotide pairs in transcription are arranged in an increasing sequence on moving downwards, i.e. in the direction of motion of the enzyme during transcription, the numeral +1 being assigned to the pair coding for the first nucleotide of the transcript. In the nontranscribable section, located above the transcribable section, including the arrangement in the promoter, the numbers are negative and the numeral 0 is absent. According to the views of Yager and Hippel,[14] the transcription in prokaryotes takes place in six stages:

(1) The interaction of the enzyme with DNA is initially as a rule nonspecific but terminates with formation of an enzyme–promoter complex.

(2) Fusion of the template DNA in the region of contact between the promoter and the enzyme with formation of a locally unwound region (bulge) on the double-stranded DNA, which makes possible retranscription with the aid of the active centre of the enzyme located on the transcribable information strand (in the form of a ribonucleotide sequence).

(3) Abortive initiation, which begins with the reaction of two ribonucleoside 5′-triphosphates (NTP) selected by the DNA template,

$$pppN_1 + pppN_2 \longrightarrow pppN_1pN_2 + pp, \qquad (2)$$

and continues in accordance with Eqn (1) until ∼10 elongation steps have been completed, after which the σ-subunit is split off from the complex and the latter is markedly stabilised [up to this moment, the possibility of the dissociation of the complex with ejection of a short (abortive) product, an oligonucleotide, remains].

(4) Elongation occurring in the stable complex with migration of the enzyme and the bulge along the DNA; within the limits of the bulge the 3′-terminal fragment of the newly synthesised DNA forms a duplex with the transcribable DNA strand (see Fig. 2).

(5) Encounter between the enzyme and the nucleotide sequence in the transcribable DNA, which is potentially capable of promoting termination.

(6) Decomposition of the transcription complex with liberation of the transcription product.

Naturally the active centre moves along the strand being transcribed and a hybrid duplex comprising the fragment of the strand being transcribed and the RNA product is formed on the region between the start of the bulge and the active centre. At a certain instant, the RNA product begins to separate from the template, whereupon the latter is rejoined to the nontranscribed strand. On the other hand, the DNA strands continue to move apart ahead of the active centre migrating along the strand being transcribed.

In recent years, studies by chemical methods of eukaryotic RNA polymerases have begun to be published. These enzymes have a significantly more complex structure than the enzyme from *E. coli*. They consist of approximately ten different subunits. Furthermore, three different RNA polymerases, usually designated as RNA polymerases I, II, and III, function within the cores. For yeast the corresponding enzymes are designated by the Latin letters A, B, and C. These enzymes are respectively responsible for the transcription of the genes of the ribosomal RNA, the genes of information RNA, and the genes coding for various small RNA, including transport and ribosomal 5S RNA. The subunit composition of certain, most intensively investigated enzymes is presented in Table 1. An important distinctive feature of eukaryotic enzymes is the need for the involvement of special transcription factors in the transcription initiation process. According to the available data, the functionally significant interaction of all three RNA polymerases with DNA begins with the formation of a complex between DNA and the set of transcription factors specific to each of the three polymerases.

Thus the following problems are solved (or can be solved) with the aid of chemical methods. The first is the establishment of the nature of the subunits interacting with DNA and the study of the dynamics of the primary interaction of the enzyme with DNA. These problems can be solved by the direct photocross-linking method. The second is the establishment of the region of DNA interacting with the enzyme at different transcription stages, which may be achieved with the aid of footprinting. The third is the determination of the size and position relative to the enzyme of the 'bulge' migrating along DNA, which can be achieved by identifying the DNA residues not involved in the Watson–Crick

‡ If the phosphorylation occurred at serine, threonine, or tyrosine residues, the modification product would be stable in the pH range used for demodification. Thus the knowledge of the chemical nature of the modified residue makes it significantly easier to locate it.

§ For the principal concepts, see Alberts *et al.*[13]

Table 1. The subunit composition of eukaryotic RNA polymerases (in kDa).

Yeast			Wheat seedlings	Calf thymus		Human HeLa cells		Drosophila	
A	B	C	I	I	II	I	II	I	II
190	220	160	180		210	197	240	195	175
	185				180		214		
135	150	128	140–150		145	140	180	125	135
		82				126			
49	53			51		52		48	
43	44.5			44	44			41	
40		40			36			38	
34.5	32	31	30						
27	27	27		25	25			25	
23	23	23			20				
19		19			18			18.5	
	16.5			16.5	16				
					15, 12, 11.5				

interactions. The fourth is the identification of the subunits participating in the interaction with the DNA strands in the region of the bulge, with the newly synthesised RNA before and after its separation from the template DNA (establishment of the exit corridor), and with the substrates (NTP). The last range of problems may be solved with the aid of the affinity labelling of the components by the corresponding reactive analogues, including the use of competent labelling, or by direct photocross-linking on UV irradiation. For adequately stable products, it becomes possible to establish the positions of the interacting aminoacid residues by identifying the modified aminoacids. In this case, one obtains information about the individual elements of the active centre of the enzyme. For eukaryotic enzymes, the study of the region of DNA interacting with the transcription factors and also the interaction of the latter with the enzyme itself are added to these tasks. We shall examine consecutively the principal results obtained by applying chemical approaches to the solution of these problems.

V. Interaction of RNA polymerase and its subunits with DNA

The first experiments designed to investigate the interaction of RNA polymerase with DNA were carried out by means of the chemical method easiest to use — the method of direct photocross-links. A mixture of DNA and the enzyme irradiated with UV light (usually at $\lambda = 254$ nm) is treated with the corresponding nucleases, after which only small fragments of the attached DNA are retained in the form bound to the enzyme. The enzyme subunits are separated by gel electrophoresis and the subunits bound covalently to DNA on irradiation are found from the positions of the radioactively labelled bands. It is also possible to treat the irradiated mixture with restriction enzymes, which cleave DNA at strictly defined sections, to precipitate with the aid of antibodies to RNA polymerase the fragments which contain the covalently bound enzyme, to remove the enzyme from the precipitated fragments by treatment with the corresponding proteinase, and, having separated them by gel electrophoresis, to find from the positions of the bands which of the fragments and in what amount were cross-linked with RNA polymerase on irradiation.

The results of the first study of this kind are presented in the paper of Hillel and Wu.[15] It was noted that, when an excess of the enzyme is mixed with ^{3}H-labelled DNA of T7 phage and on subsequent UV irradiation of the mixture, the main part of the enzyme is bound outside the promoters, *i.e.* nonspecifically. If an excess of unlabelled DNA is added to the mixture before irradiation, then the complexes between labelled DNA and the promoters are retained for a fairly long time and the labelled DNA in the nonspecific complexes is replaced by the unlabelled DNA, so that on UV irradiation the radioactivity is retained only in the specific complexes. It was established in this way that three subunits (β, β′, and σ) participate in the nonspecific interaction, while only β and σ are involved in the formation of specific complexes. In both cases, there is no interaction with the subunit α.

It has been shown using DNA in which thymine had been substituted by bromouracil, affording more stable cross-linking products,[16] that the third heterocycle in the upward direction from the start point (the −3 pair) participates in one of the cross-links in the σ-subunit, while the +3 pair participates in one of the cross-links in the β-subunit.

Similar results, which agree in principle with those presented on the basis of photocross-links, have been obtained on cross-linking the enzyme with DNA by means of formaldehyde.[17] The procedure developed by Mirzabekov[17] for the study of the interaction of DNA with histones in chromatin was used for these purposes. It is based on the elimination of purines from DNA to a slight extent (the purine heterocycles are split off from certain nucleotide residues with formation of the open-chain form of deoxyribose). The aldehyde groups of such DNA can interact with the ε-amino-groups of the lysine residues in close proximity with formation of Schiff bases, which are readily converted into stable alkylamines by treatment with sodium tetrahydroborate. The most detailed picture of the cross-linking of RNA polymerase with the promoter section of lac UV5 has been obtained[18] by this method.

The results of the modification of *E. coli* RNA polymerase by the alkylating derivative of the oligonucleotide carrying at the 5′-terminal phosphate the 4-[*N*-methyl-*N*-(2-chloroethyl)amino] benzylamine residue and having a sequence corresponding to the fragment from −13 to −2 of the nontranscribed DNA strand containing the *spc* promoter have also been described. It has been established that the oligonucleotide itself has a high affinity for the enzyme, which may be evidence in support of the mechanism involving the unwinding of the double-stranded DNA in the region of the promoter as a result of the preferential binding of the nontranscribed strand and hence the liberation of the strand being transcribed. The alkylating derivative of this oligonucleotide modifies the σ-subunit. The phenomenon involving the effective binding of the octaribonucleotide, complementary to the section from −13 to −6 of the nontranscribed strand, to the enzyme, which has not so far found a satisfactory explanation, is also described in the above study; the alkylating derivative of the octaribonucleotide then modifies the β,β′-subunits.[19]

The dynamics of the cross-linking of the enzyme with DNA has been investigated by means of a special apparatus for rapid mixing and subsequent irradiation after a specific time interval by a powerful short (10 μs) UV light pulse.[20] The distribution of the attached enzyme between eight different sections of DNA, obtained as a result of the highly specific fragmentation of DNA with the aid of the Kpn I and Mbo I restriction enzymes, was analysed for this purpose. It was found that in such cleavage the three main promoters of the T7 phage DNA are located in one fragment. During the first tens of milliseconds, the distribution was close to statistical, *i.e.* the amount of the RNA polymerase bound to the fragment was correlated satisfactorily with the size of the fragment. The redistribution of the enzyme between the fragments with its gradual accumulation in the fragment containing the promoter then took place gradually. This made it possible to determine the rate constant for the formation of primary nonspecific complexes and the rate constant for the migration of DNA in the direction of the promoter. The nature of this migration agrees with the mechanism of migration by means of one-dimensional diffusion along DNA.

The region of contact between the *E. coli* RNA polymerase and DNA has been repeatedly investigated with the aid of footprinting. Complexes formed with the promoter in the absence of transcription and in the course of the latter were in-vestigated. In the first case, it was possible to detect and characterise three states

of the complex — closed, intermediate, and open. The best results were obtained when OH radicals were used.[21] In the closed complex, formed at 4 °C, the region protected by the enzyme against the action of radicals is that between −4 and −53 and the transcribed and nontranscribed strands are alternately protected with the periodicity corresponding to the helical struc-ture of the DNA. This fact shows that only one side of DNA comes into contact with the enzyme in this region. On formation of the intermediate complex (20 °C), the turns of the DNA helix are fully protected and full protection of the nontranscribed strand along the section from −10 to +21 and of the transcribed strand along the section from −13 to +18 is observed. It had been proposed earlier[22] that this section be referred to as the fusion domain, whereas the region of the incomplete contact located above it should be called the recognition domain. It must be emphasised that the fusion domain does not coincide with the fused region of DNA, namely with the size of the bulge which enters into the composition of the domain, *i.e.* is significantly smaller. The part of DNA in the fusion domain, adjoining the bulge, is still (to the right) or already (to the left) in the double-stranded form. Finally, at 37 °C, whereupon the open transcription-active complex is formed, the region of contact with the nontranscribed strand is retained while a section from −7 to +2 accessible to modification arises on the strand being transcribed.

VI. Investigation of the segment of DNA unwound in the complex with RNA polymerase (the bulge)

In the study of transcription with the aid of footprinting, it is possible to detect the migration of the edges of the bulge. Thus, in the transcription the promoter lac UV5 (the first transcribed nucleotides of which form the sequence TTAACA...) in the complex with the enzyme or with the enzyme and the primer ApA simulating the first two transcription steps, the residue C in the +5 position is not alkylated by dimethyl sulfate, whereas after the addition of dTTP, which elongates the substrate by a further two units, the same residue C becomes accessible to alkylation.[23]

Further investigations required the formation of a series of transcription complexes with different lengths of the transcripts. This was achieved by employing nucleotide sequences in the transcribed part of DNA favourable for this purpose. For example, the transcription has been carried out[22] from the fragment of T7 DNA with the sequence (3′)T(+1)AGCTCTCC. .CTGTGCCGCTTA...(5′) on the strand being transcribed. When the transcription is performed using the trinucleotide AUC as the primer and ATP and GTP as the substrates, the synthesis proceeds as far as the eleventh unit, since, owing to the absence of CTP from the mixture, the next step cannot be carried out. Similarly, by carrying out the transcription in the presence of CTP, ATP, and GTP, it is possible to obtain a transcription complex in which the length of the transcript is 20, because in the absence of UTP the 21st step of the synthesis cannot occur. By altering the nucleotide sequences in DNA as a result of the introduction of mutations, it is possible to obtain transcription complexes with most diverse lengths of the newly synthesised RNA. It has been established[22] with the aid of footprinting by OH radicals that, when the transcript reaches a length of 11 units, the disruption of the contact between the enzyme and the recognition domain without change in the size of the fusion domain is observed. However, a further increase in the length of the transcript leads to a decrease in the fusion domain from ~30 units to 23 units. The disruption of the contacts between the enzyme and the recognition domain after the attainment of a length of 11 units by the transcript has also been recorded by means of footprinting with the aid of the Cu^{2+} – 1,10-phenanthroline system.[24]

A more universal procedure is based on the immobilisation of RNA polymerase on a carrier.[25] By employing stepwise elongation (by omitting at each stage one of the four substrates), it was possible to obtain transcription complexes with different lengths of the transcript. The length of the transcript for each such complex yields the position of the catalytic centre of the enzyme and footprinting (in the above study, exonuclease III was employed) yields the lower point of the contact between DNA and the enzyme. Systematic measurements of the distance between these two points showed that the transcription proceeds mainly continuously, *i.e.*, for each step of the migration of the enzyme active centre, the boundary of the region of the DNA–enzyme interaction undergoes unit migration. However, in certain cases it was found that the migration of the active centre of the enzyme itself does not lead to the migration of the enzyme relative to the template and after several such steps the enzyme, as it were, executes a jump over several pairs of template nucleotides.

The migration of the enzyme in jumps has been detected also in a study[26] involving the application of a more refined chemical technique. In this investigation, a series of transcription complexes were obtained with the length of the transcript from 11 to 20 nucleotides. These complexes were obtained by transcription on a DNA fragment containing the promoter and the sequence of the first 20 transcribed nucleotides (3′)d(TAGCTCTCCCTGTG. .CCGCTT...)(5′). In order to obtain the set of complexes, the transcription was initiated with the aid of the ribonucleotide pAUC using GTP and ATP as the substrates. This system yielded a complex with a transcript length of 11 units (complex 11). After its isolation by means of gel filtration and by adding CTP, complex 12 was obtained, while the addition of CTP and ATP afforded complex 14. Complex 14 was converted into complex 16 in the presence of GTP, into complex 18 in the presence of CTP + GTP, and into complex 20 in the presence of CTP + GTP + ATP. The position of the uncoiled section of DNA (the bulge) was determined by means of three reagents, each of which was specific to one type of heterocycle: thymine, adenine, and cytosine were determined with the aid of OsO_4, diethyl pyrocarbonate, and dimethyl sulfate followed by treatment with hydrazine respectively. In combination with footprinting by OH radicals, these investigations showed that, up to and including complex 18, the double helix is uncoiled ahead of the RNA polymerase and the position of the upper contact between DNA and the enzyme does not change. After the attainment of a length of 20 units by the transcript, a jump occurs and the helix is closed, passing to the double-stranded state.

The change in the size of the region of contact between the template and RNA has been recorded for transcription complexes with RNA polymerase II from calf thymus with transcripts having lengths corresponding to 135, 138, and 143 nucleotides. Footprinting with the aid of DNase I showed that the dimensions of the sections protected against hydrolysis correspond respectively to 48, 55, and 51 pairs of nucleotides.[27]

VII. The study of the RNA exit corridor

In order to trace the migration path of the 5′-end of the transcript formed, two principal approaches were employed. One of them is based on the introduction of a reactive group into the initiating substrate or primer. The first experiments of this kind were described in studies[28, 29] where ATP γ-*p*-azidoanilide was used as one of the substrates. The latter may serve for both initiation and elongation. Naturally the photoreactive section of γ-*p*-azidoaniline is retained only at the 5′-end of the transcript. Photomodification was detected only in the β- and σ-subunits. A more detailed study of the transcript by means of this procedure has been carried out in a study[30] where transcription by the *E. coli* RNA polymerase was carried out on a poly[d(A–T)] template using labelled [α-^{32}P] ATP and a primer containing the photoreactive label *p*-$N_3C_6H_4C(O)CH_2SP(O)(O^-)$AdoPUrd. After the transcription, the transcription complex was irradiated, the labelled subunits were separated from one another, were separated by dialysis from the substrates, and were treated with PheHgOAc in order to split off the label from the transcript. The transcripts were separated by electrophoresis and were radioautographed. This made it possible to determine the amount of transcript of a

particular length bound to each subunit. It was found that transcripts with lengths of 6–8 units modify DNA preferentially, those with a length of 9–12 units modify the σ-subunit, and those with a length of 13–19 units modify the β/β′-subunit.[31]

The second procedure is based on the hypothesis that the newly synthesised template RNA within the limits of the template DNA bulge is in the form of a duplex with the DNA strand being transcribed. In this case, it can be modelled by an oligonucleotide complementary to several residues of the open part of the strand being transcribed and carrying a reactive group at the 5′-end. This was done on the A2 section of the T7 phage DNA promotor which, forming part of a complex with *E. coli* RNA polymerase, forms the open section (3′)d(ATGTTAGCGA) on the strand being transcribed, the first seven residues referring to the non-transcribed region (i.e. corresponding to numbers from −7 to −1), while the dC residue is the first transcribed nucleotide.

A series of derivatives of oligonucleotides complementary to the open section and having the general structure R(pdN)$_n$pTprC with a 3′-terminal dinucleotide complementary to the fragment A(−2)G(−1) (the first group of reagents) and R(pdN)$_n$pdGprC with the terminal dinucleotide complementary to the fragment C(+1)G(+2) (the second group of reagent), where R is the reactive group, where prepared. The group R contained either a benzaldehyde residue (see below) or a fragment of an aromatic 2-chloroethylamine derivative capable of alkylating the nucleophilic centres of proteins and nucleic acids, or a phosphate activating group, for example *N*-methylimidazole. After the affinity modification with the aid of the first group of reagents, [α-^{32}P]GTP was added to the reaction mixture and competent labelling was carried out, *i.e.* the label was introduced into the oligonucleotide bound covalently to the enzyme. In the case of the second group of reagents, [α-^{32}P]UTP was added in order to carry out the competent labelling. Next a study was made of the distribution of the radioactivity among the subunits and the presence of the histidine and lysine residue modification was detected (with the aid of the criteria described above, based on the kinetics of the demodification of the phosphorylation products). In the case of alkylating reagents, the modification of DNA was also observed. The distribution and nature of the labelled aminoacid residues differed significantly for reagents with different lengths of the oligonucleotide component, *i.e.* with different distances between the chemically reactive group and the G·C(+1) pair. Fig. 3 presents the results of the determination of the 'exit corridor' obtained with the aid of the first group of reagents.[32] The data obtained using the second group of reagents [33] agree qualitatively with those presented in Fig. 4 for reagents of the first group. However, there are also distinct differences. For example, the σ-subunit is modified by reagents of the first group at both lysine and histidine residues and the group R cannot be attached closer than to the residue forming the −7 pair. On the other hand, when reagents of the second group were employed, the modification of the histidine residues in the σ-subunit was already observed in the modification by the hexanucleotide (n = 4).

The determination of the RNA exit corridor for human RNA polymerase II (HeLa cells), based on the photoaffinity modification of the enzyme with the aid of the transcript into which 4-thiouridine monophosphate (s^{4}UMP) residues were introduced during transcription, has also been described. The length of the transcripts was regulated by introducing 3′-CH_3O-GTP as the terminator. Mainly large subunits with masses of 240 and 140 kDa were modified.[34] It was found that the length of the transcript must correspond to at least 16 nucleotide residues for sufficiently effective modification. This makes it possible to infer the distances between the active centre and the start of the exit of the transcript from the duplex formed by the transcribed DNA strand and the transcription product.

This procedure has also been used to identify the subunits in contact with the transcript, formed in the isolated nucleoli, in which ribosomal RNA genes are transcribed with the aid of RNA polymerase. For this purpose, s^{4}UTP and labelled CTP were introduced into the nucleoli of the HeLa cells and UV-irradiation was carried out for a time. Analysis of the radioactivity associated with the RNA polymerase I isolated from the irradiated nucleoli showed that subunits with masses of 200 and 120 kDa are modified in the first place. With increase in the size of the transcript, the fraction of the transcript bound covalently to the subunit with a mass of 52 kDa also increases.[35]

The same system has been used to locate RNA polymerase II in the nucleus. It was established that, after the irradiation of the nuclei in which transcription with participation of s^{4}UTP and labelled CTP took place, the enzyme molecules are not bound to the nuclear template but move along loops formed by chromosomal DNA.[36]

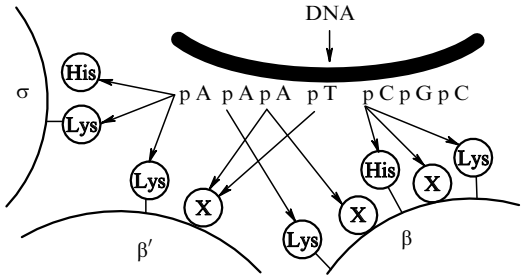


Figure 3. Topography of the product RNA exit corridor obtained by labelling the transcription complex of the *E. coli* RNA polymerase by reagents containing aldehyde and alkylating groups, and by phosphorylating derivatives of oligoribonucleotides. The arrows indicate the interactions of the reactive groups of the reagents with the aminoacid residues of the protein — Lys, His, and X (X is an aminoacid residue other than lysine and histidine).

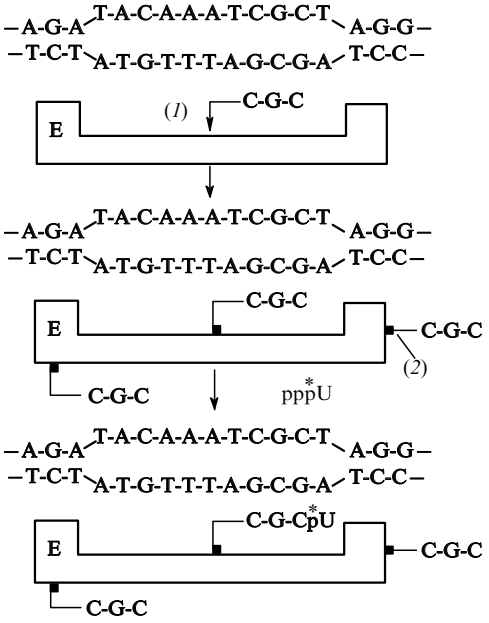


Figure 4. Schematic illustration of the highly selective affinity labelling of RNA polymerase. (*1*) Phosphorylating residue; (*2*) site of addition of the phosphorylating residue to the protein.

VIII. Approaches to the study of the active centres of RNA polymerases by the affinity modification method

Most of the information about the aminoacid residues of the enzyme participating in the formation of the active centre or adjoining it has been obtained by the highly selective labelling method described above. The initiating region of RNA polymerases from various sources such as the enzyme produced by the T7

phage,[37] three yeast RNA polymerases,[38] and the RNA polymerase II from human placenta [39] and from wheat seedlings [40] has been investigated by this method. Derivatives of the initiating purine nucleotide containing benzaldehyde at the 5′-end and forming a Schiff base with the ε-aminogroup of lysine were used as the affinity reagent in the above investigations. The Schiff base was reduced with tetrahydroborate to an alkylamine, after which elongation was carried out with the aid of the labelled [α-^{32}P]NTP. Adenosine 5′-trimetaphosphate was used to modify the region of the initiation of transcription. In the case of eukaryotic enzymes, the second subunit in terms of size was labelled. In the case of the T7 RNA polymerase, the 631 lysine residue was modified.[41] In the *E. coli* RNA polymerase, the modification points were the 1065 lysine residue and the 1237 histidine residue of the β-subunit.[42] In the yeast RNA polymerase II, the 979 and 987 lysine residues were modified.[43]

By virtue of its high selectivity, the competent labelling method may be applied also in cell extracts. Although this has been done on RNA replicase (which we do not consider because of the scantiness of the available data),[44] we thought it useful to present an example of the determination of the functions of two nonstructural proteins of the tick encephalitis virus, because it demonstrates the possibilities of the use of the above approach in complex supermolecular systems. The UUC(3′) sequence is present at the 3′-end of the plus-chain of the viral RNA, from which replication should begin in order to form the double-stranded replicated RNA. The reactive analogue from which replication may begin should be a derivative of GMP. In competent labelling, elongation should be achieved with the aid of [α-^{32}P]ATP. When these components were introduced into the cell extract from pig embryo kidneys infected with the virus, the radioactive label was detected in a protein with a molecular mass of 69 kDa 8 h after infection, which, judging from the spectra of the viral RNA, corresponds to the nonstructural (*i.e.* not entering into the composition of the mature virus) NS3 protein. In the extract obtained from the cells isolated 45 h after the start of infection and subjected to the same treatment, the label was present in another nonstructural protein (NS5) with a molecular mass of ~100 kDa. In conformity with the logic of the events characteristic of RNA-containing viruses, one may assume that the NS3 protein participates in the replication of double-stranded RNA, required for the conversion of the viral RNA into the double-stranded form while NS5 participates in the replication of the double-stranded RNA needed for the accumulation of new plus-chains entering into the composition of new viral particles.

5-Azido-UTP [45] and 8-azido-ATP have also been proposed for the direct modification of NTP-binding centres.[46] The first proved to be the substrate of the enzyme from *E. coli* with $K_m \sim 80$ μM. Photoirradiation of the enzyme in the presence of this reagent leads to approximately the same incorporation of the label into the β-, β′-, and σ-subunits, the presence of DNA not influencing the distribution of the label between the subunits. 8-Azido-ATP is a poor substrate for this enzyme, but it inhibits the enzyme competitively (in relation to ATP). Photoaffinity modification leads mainly to the modification of the β′- and σ-subunits. When the analogue is incorporated in the 3′-terminal position in the transcript and on subsequent irradiation, the β′-subunit is modified within the limits of the 932–1020 peptide.[47] 8-Azido-ATP has also been used for the photoaffinity modification of the RNA polymerase coded for by the T7 phage [48] and of the RNA polymerase II from calf thymus.[48] Peptides containing the attached photolabel were isolated in these two investigations.

Affinity modification makes it possible to detect the migration of the substrate within the limits of one enzyme molecule or between subunits. As an example, one may quote a study [50] in which the transformation of the transcription-active complex formed by the enzyme from *E. coli* into an inactive complex was observed when transcription ceased as a result of the lack of substrate for the next elongation step. This transformation occurs when the system is incubated for several tens of minutes at 37 °C and is irreversible, i.e. the addition of substrate no longer leads to continuation of transcription. The complex was obtained by stepwise elongation on the immobilised enzyme [25] and it contained the photoreactive 4-thio-UMP residue at the 3′-end. The complex was irradiated at 360 nm. Analysis of the modification of the subunits showed that in the transcription active complex the β-subunit is modified within the limits of the 515–660 and 1091–1107 peptides and the β′-subunit is modified within the limits of the 400–465 peptide; in the inactive complex,only the β′-subunit is modified within the limits of the 1091–1107 peptide.

Other less traditional approaches to the study of the active centres of RNA polymerase have been developed recently. Thus it has been proposed [51] that the chemical and genetic approaches be combined in order to establish the region where the Mg^{2+} ion is bound in the active centre of the enzyme with replacement of the Mg^{2+} ion by Fe^{2+}. It was found that oxidative degradation of both the template and the enzyme takes place in this instance. The degradation proved to involve mainly the β′-subunit, cleavage occurring within the limits of the peptide located between the 466 and 400 methionine residues. Since the conservative sequence (*i.e.* a sequence characteristic of a wide variety of nucleic acid polymerases) NADFDGD (in terms of single-letter designations, N = asparagine, A = alanine, D = aspartate, F = phenylalanine, and G = glycine) is located on this section, it was concluded that the above fragment participates in the binding of the magnesium ion. A mutant variant of RNA polymerase, in which the β′-subunit contained in the same place the NAAFAGA sequence (the aspartate residues were replaced by alanine residues) was obtained by the protein engineering method for additional confirmation of the cleavage site. After the replacement of Mg^{2+} by Fe^{2+}, the enzyme was not cleaved in the presence of this fragment. The cleavage of template DNA at rigorously defined sections in the −1 and −2 positions relative to the transcription start point was observed in the same experiment.

The use of a chimeric reagent, consisting of the conjugate of the initiating substrate (ATP) with an antibiotic (rifampicin), for affinity modification has been described.[52] This antibiotic is a powerful inhibitor of prokaryotic RNA polymerases and it is therefore an effective antitubercular preparation. A series of conjugates with different lengths of the $-NH-(CH_2)_n-NH-$ spacer, connecting the carboxy-group of the antibiotic with the ATP γ-phosphate, were prepared. It was found that, when the length of the linker is $n > 3$, the conjugate serves as the initiating transcription substrate, whereupon the optimum length of the spacer corresponds to $n = 5$. Rifampicin is known to be bound strongly to the enzyme. This means that the distance between the centres binding the antibiotic and initiating the substrate is ~1.5 nm.

A conjugate the spacer of which included the alkylating group $-NH(CH_2)_2N[C(O)CH_2Br](CH_2)_6$ has also been synthesised. In this case, the modification of the transcribed DNA strand in the C(−3) and C(−2) positions was detected.

The authors also employed a procedure which they called a pseudoenzymic reaction. For this purpose, the alkylating fragment $-NH(CH_2)_nNHC(O)CH_2Br$ was attached to the carboxy-group of rifampicin and an ATP analogue, containing a sulfur atom instead of one of the oxygen atoms in the γ-phosphate, was introduced into the reaction mixture containing the enzyme and the rifampicin derivative. When these compounds were disposed favourably in the corresponding active centres, the alkylation of the thio-analogue of ATP was observed. The most effective alkylation was noted for $n = 4$, which corresponds to the size of the spacer with six carbon atoms.

Nucleoside triphosphates can function in the transcription process in ways other than as substrates of the polymerisation process. In the study of the photoaffinity modification of RNA polymerase II from calf thymus with the aid of 8-azido(naphthalene-1′-sulfonyl) [α-^{32}P]ATP, it was established that, when one residue of the reagent is attached covalently, modification accompanied by the complete inactivation of the enzyme takes place. The inactivation occurs competitively in relation to ATP. However, other substrates — CTP and GTP — are not competitive

inhibitors in relation to the reagent and hence the modified section is not a substrate binding site. Since for the initiation of the transcription by RNA polymerase II the hydrolysis of one ATP group is required, it has been suggested that the modification takes place at a specific ATP binding site in which the hydrolysis required for the initiation occurs.[53] Using the 2′-*O*-dansyl-ATP fluorophor, which according to the authors forms a complex with the modified enzyme at the substrate binding site, it was possible to determine the distance between the ATP and the other substrate binding site, by measuring the resonance energy transfer. It was found to be ~2.5 nm.

All the methods described above lead to the labelling of the enzyme in the NTP binding region. There have been individual studies in which an attempt was made to bind the enzyme to the 3′-end of the growing RNA chain. The syntheses of several short transcripts into the 3′-ends of which the s^4UMP residue had been introduced have been described.[54] Short single-stranded templates with the general structure (3′)d(CCCCCCGCTTGTGGTG – X), where X = A, TA, TTA, or TTTA, were used for this purpose. They were transcribed with the aid of yeast RNA polymerase I in the presence of the GpCpG primer and three substrates: ATP, [α-^{32}P]CTP, and s^4UTP (the transcribed part of the template is underlined). It is not difficult to note that the transcription began with the site where the 3′-end of the primer is located and, since there is a single dA residue on the transcribed strand, located at the 5′-end of the template, a single s^4UMP residue, coded for by this dA, was incorporated in the 3′-end of the transcript. After the completion of the process, the transcripts obtained, with a length equivalent to 12 – 15 nucleotide residues, were found to be part of the composition of the transcription complexes containing the enzyme, the template, and the transcript with a photoreactive residue located in the region of the active centre. Irradiation led to the attachment of the radioactive label to two subunits with molecular masses of 190 and 135 kDA. The ratio of the levels of radioactivity of the two subunits changed as a function of the length of the transcript, although it would seem that in all the systems investigated the photoreactive residue has always been in the same position on the enzyme, corresponding to the 3′-end of the transcript. This is unexpected and has not found a convincing explanation.

The same approach with the same set of template and transcripts labelled at the 3′-end has been used for the affinity modification of RNA polymerase II from yeast and wheat seedlings. In both cases, the affinity modification led to the incorporation of radioactivity solely in the largest subunit (220 kDa for the yeast enzyme and 180 kDa for the enzyme from wheat seedlings).[55]

IX. The study of the interaction of eukaryotic transcription factors with DNA

For each of the three eukaryotic RNA polymerases, there is an intrinsic set of transcription factors participating in the transcription initiation stage. In accordance with the number of the polymerase, they are designated as TFI, TFII, and TFIII. Within the limits of each set, they are additionally designated by Latin letters The majority of the factors consist of several subunits. In the transcription of the genes of information RNA, the set of principal factors necessary for the transcription of any genes participates in the initiation, while additional highly specialised factors are needed to trigger a whole series of genes. The principal factors include TFIIA, TFIIC, and TFIID, which form a primary complex in the immediate vicinity of the section of DNA containing the d(TATA) sequence at a short distance upward from the start point of the transcription, as well as the factors TFIIE, TFIIF, TFIIB, and TFIIH. One of the subunits of the last factor is a protein kinase and catalyses the phosphorylation of the large subunit of the RNA polymerase, which is necessary for the manifestation of enzyme activity.

Figure 5 presents a schematic illustration of the disposition of the subunits of the factors TFIIIB and TFIIIC on genes.

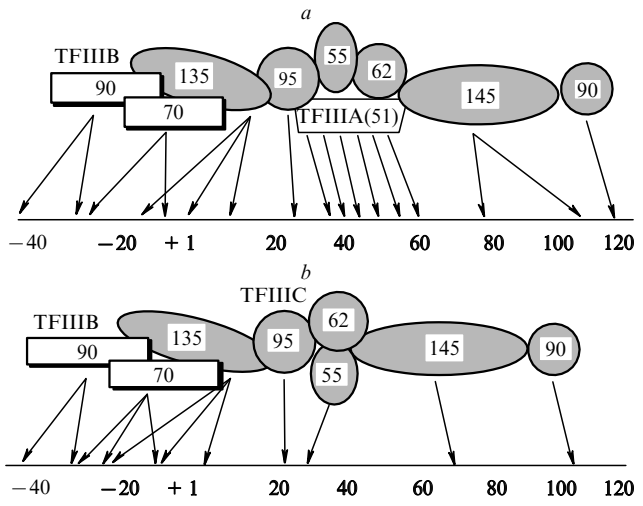


Figure 5. Schematic illustration of the interaction of the transcription factors III with the genes of 5S rRNA (*a*) and the 4th suppressor $tRNA^{Tyr}$ (*b*). The masses of TFIIIA, TFIIIB, and TFIIIC are indicated in kDa. The arrows indicate the position in the DNA sequence which have been photocross-linked with a particular subunit of the transcription factors.

A detailed arrangement of the subunits on two DNA specimens coding respectively one of the tRNAs and ribosomal 5S RNA has been established for the factors TFIII with the aid of photoaffinity modification. The disposition of part of the promoter region within the limits of the coding section is characteristic of RNA polymerase III. A large series of double-stranded DNA have been obtained by genetic engineering techniques.[56, 57] One of the strands was elongated by several units with the aid of DNA polymerase with incorporation into the mixture of the elongating substrates of a photoreactive dNTP — 5-(3-*p*-azidobenzoylaminoallyl)dUTP. The photoreactive label was introduced in this way into various positions in DNA, after which a complex with the corresponding factors was formed, the photoaffinity modification of the factors was carried out, and it was established to which genes the covalent binding had taken place.

X. Conclusion

The methods based on chemical modification began to be applied to template biosynthesis systems primarily on ribosomes. The first studies of the modification of the tRNA binding region with the aid of an alkylating derivative of tRNA and of the mRNA binding region with the aid of an alkylating derivative of an oligonucleotide were published at the beginning of the 1970s.[58, 59] These were followed by systematic investigations of the structural-functional topography of the *E. coli* ribosomes with the aid of photoreactive derivatives of tRNA and reactive derivatives of ribooligonucleotides (the main results of these investigations are presented in a review [60]). In recent years, similar studies have been initiated on ribosomes from human placenta. Information has been obtained about the protein environment and the sections of ribosomal RNA in the region where genetic information is decoded.[61, 62] Bifunctional cross-linking methods have found extensive application in the determination of the steric organisation of ribosomes. They make it possible to determine the mutual disposition of the ribosomal proteins and their disposition relative to particular sections of ribosomal RNA. In combination with data on the mutual disposition of proteins found by the neutron scattering method, this made it possible to construct a detailed three-dimensional model of the disposition of the 16S ribosomal RNA in the small subunit of the *E. coli* ribosome.[63]

From the data presented in the review, it is seen that the systematic application of chemical methods to the study of transcription has only just begun.

With the exception of the investigation of ribosomes, the bifunctional cross-linking method has not so far found extensive application in the study of transcription. As an individual example, mention may be made of the study [64] whose authors used a bifunctional reagent (*N*-hydroxysuccinimidyl-4-azidobenzoate) — to elucidate the interaction of the subunits of the *E. coli* RNA polymerase with the protein ω, which is present in a highly purified enzyme preparation in a stoichiometric amount but is not an essential component for transcription. Bifunctional cross-linking data have shown that the protein is bound to the β′-subunit of the whole enzyme. The bifunctional cross-linking method is particularly effective in the study of systems with a large number of proteins and nucleic acids and one may therefore expect that its importance will increase as one proceeds to the study of transcription in more complex systems, for example in complexes of chromosomal DNA with eukaryotic RNA polymerases and with a set of transcription factors, including those forming part of the composition of chromatin.

It is seen from the examples quoted that chemical approaches lead to new unique possibilities for the study of template biosynthesis. The possibility of investigating the mechanism of the search by the enzyme for points at which the functionally significant process begins,[20] the determination of the positions and dimensions of the unwound section of DNA on RNA polymerase and its migration in the course of transcription,[26] and the determination of the RNA exit corridor, *i.e.* of the actual 'track' of the migration of the product on the enzyme as it moves away from the catalytic centre where biosynthesis takes place, are the most impressive features at the present time.[29–34] It is at present difficult to even imagine experiments not based on chemical approaches which would make it possible to look quite so deeply into the dynamic aspects of template biosynthesis. Some of the procedures described may be and are already employed for the study of the transcription process in cell organelles.[35, 36]

References

1. D G Knorre, V V Vlassov *Affinity Modification of Biopolymers* (Boca Raton, FL: CRC Press, 1989)
2. D G Knorre, S D Myzina, in *Biologicheskaya Khimiya* (Biological Chemistry) (Moscow: Vysshaya Shkola, 1992) p. 199
3. D G Knorre, S D Myzina, in *Biologicheskaya Khimiya* (Biological Chemistry) (Moscow: Vysshaya Shkola, 1992) p. 232
4. B Leblank, T Moss, in *DNA-Protein Interactions, Principles and Protocols* (Ed. G G Kneale) (New York: Humana Press Totowa, 1994) p. 1
5. M A Price, T D Tulius *Eur. J. Biochem.* **212** 194 (1992)
6. W C Suh, W Ross, M T Record *Science* **259** 358 (1993)
7. R Hori, S Pyo, M Carey *Proc. Natl. Acad. Sci. USA* **92** 6047 (1995)
8. M A Grachev, T I Kolocheva, E A Lukhtanov, A A Mustaev *Eur. J. Biochem.* **163** 113 (1987)
9. V N Ankilova, D G Knorre, V V Kravchenko, O I Lavrik, G A Nevinski *FEBS Lett.* **60** 172 (1975)
10. M M Hanna, Y Zhang, J C Reidling, M J Thomas, J Jou *Nucleic. Acids Res.* **21** 2073 (1993)
11. K C Smith *Biochem. Biophys. Res. Commun.* **34** 354 (1969)
12. N P Skiba, V A Efimov, V M Lipkin, O G Chakhmakhcheva, Yu A Ovchinnikov *Bioorg. Khim.* **12** 1023 (1986) [a]
13. B Alberts, D Bray, J Lewis, M Raff, K Roberts, J D Watson, in *Molecular Biology of the Cell* (New York: Garland Publ. Inc., 1994) p. 365
14. D T Yager, P H Hippel *Biochemistry* **30** 1097 (1991)
15. Z Hillel, C-W Wu *Biochemistry* **17** 2954 (1978)
16. R B Simpson *Cell* **18** 277 (1979)
17. K L Brodolin, V M Studitsky, A D Mirzabekov *Nucleic Acids Res.* **21** 5748 (1993)
18. A Chenchick, R Beabealashvilli, A Mirzabekov *FEBS Lett.* **128** 46 (1981)
19. L K Savinkova, A A Sokolenko, I T Tulokhonov, V L Knorre, R I Salganik, A G Ven'yaminova, M N Repkova, N I Komarova *Mol. Biol.* **27** 64 (1993) [b]
20. C S Park, F Y-H Wu, C-W Wu *J. Biol. Chem.* **257** 6950 (1982)
21. P Schickor, W Metzger, W Werel, H Lederer, H Heumann *EMBO J.* **9** 2215 (1990)
22. W Metzger, P Schickor, H Heumann *EMBO J.* **8** 2745 (1989)
23. A Spassky *J. Mol. Biol.* **188** 99 (1986)
24. A Spassky *Biochemistry* **31** 10502 (1992)
25. E Nudler, A Goldfarb,M Kashlev *Science* **265** 793 (1994)
26. E Zaychikov, L Denissova, H Heumann *Proc. Natl. Acad. Sci. USA* **92** 1739 (1995)
27. G A Rice, M J Chamberlin, C M Kane *Nucleic Acids Res.* **21** 113 (1993)
28. E D Sverdlov, S A Tsarev, N N Modyanov, V M Lipkin, M A Grachev, E F Zaichikov, A G Pletnev *Bioorg. Khim.* **4** 1278 (1978) [a]
29. M A Grachev, E F Zaychikov *FEBS Lett.* **130** 23 (1981)
30. J M Zengel, L Lindahl *Nucleic Acids Res.* **21** 2429 (1993)
31. T M Stackhouse, C F Meares *Biochemistry* **27** 3038 (1988)
32. I G Tsarev, A A Mustaev, E F Zaichikov, T Yu Alikina, A G Ven'yaminova, M N Repkova *Bioorg. Khim.* **16** 765 (1990) [a]
33. T S Godovikova, M A Grachev, I V Kutyavin, I G Tsarev, V F Zarytova, E F Zaychikov *Eur. J. Biochem.* **166** 611 (1987)
34. B Bartholomew, M E Dahmus, C F Meares *J. Biol. Chem.* **261** 14226 (1986)
35. M Roberge, E M Bradbury *J. Biol. Chem.* **263** 18553 (1988)
36. M Roberge, M E Dahmus, E M Bradbury *J. Mol. Biol.* **201** 545 (1988)
37. A R Schaffner, E D Jorgensen, W T McAllister, G R Hartmann *Nucleic Acids Res.* **15** 8773 (1987)
38. M Riva, A R Schaffner, A Sentenac, G R Hartmann, A A Mustaev, E F Zaychikov, M A Grachev *J. Biol. Chem.* **262** 14377 (1987)
39. T G Maksimova, A A Mustaev, E F Zaichikov, A V Polukhin, V K Rait *Bioorg. Khim.* **16** 1145 (1990) [a]
40. M A Grachev, G R Hartmann, T G Maximova, A A Mustaev, A R Schaffner, H Sieber, E F Zaychikov *FEBS Lett.* **200** 287 (1986)
41. T G Maksimova, A A Mustayev, E F Zaychikov, D L Lyakhov, V L Tunitskaya, A Kh Akbarov, S V Luchin, V O Rechinsky, B K Chernov, S N Kochetkov *Eur. J. Biochem.* **195** 841 (1991)
42. A Mustaev, M Kashlev, J Lee, A Polyakov, A Lebedev, K Zalenskaya, M Grachev, A Goldfarb, V Nikiforov *J. Biol. Chem.* **266** 23927 (1991)
43. I Treich, C Carles, A Sentenac, M Riva *Nuceic Acids Res.* **20** 4721 (1992)
44. O V Morozova, N A Belyavskaya, E F Zaychikov, E A Kvetkova, A A Mustaev, A G Pletnev *Biomed. Sci.* **2** 183 (1991)
45. A-Y M Woody, R K Evans, R W Woody *Biochem. Biophys. Res. Commun.* **150** 917 (1988)
46. A-Y M Woody, C R Vader, R W Woody, B E Haley *Biochemistry* **23** 2843 (1984)
47. S Borukhov, J Lee, A Goldfarb *J. Biol. Chem.* **266** 23932 (1991)
48. M M Hanna, Y Zhang, J C Reidling, M J Thomas, J Jou *Nucleic Acids Res.* **21** 2073 (1993)
49. E Freund, P M McGuire *Biochemistry* **25** 276 (1986)
50. V Markovtsov, A Mustaev, A Goldfarb *Proc. Natl. Acad. Sci. USA* **93** 3221 (1996)
51. E Zaychikov, E Martin, L Denissova, M Kozlov, V Markovtsov, M Kashlev, H Heumann, V Nikiforov, A Goldfarb, A Mustaev *Science* **273** 107 (1996)
52. A Mustaev, E Zaychikov, K Severinov, M Kashlev, A Polyakov, V Nikiforov, A Goldfarb *Proc. Natl. Acad. Sci. USA* **91** 12036 (1994)
53. I Kang, J Lin, J H Wang *Biochemistry* **33** 2696 (1994)
54. S Kelly, N Sheng, D Dennis *J. Biol. Chem.* **265** 7787 (1990)
55. N Sheng, E B Mougey, S Kelly, D Dennis *Biochemistry* **32** 2248 (1993)
56. B Bartholomew, G A Kassavetis, B R Braun, E P Geiduschek *EMBO J.* **9** 2197 (1990)
57. B R Braun, B Bartholomew, G A Kassavetis, E P Geiduschek *J. Mol. Biol.* **228** 1063 (1992)
58. E S Bochkareva, V G Budker, A S Girshovich, D G Knorre, N M Teplova *FEBS Lett.* **19** 121 (1971)
59. V G Budker, A S Girshovich, N I Grineva, G G Karpova, D G Knorre, N D Kobets *Dokl. Akad. Nauk SSSR* **211** 725 (1973) [c]
60. G G Karpova, D G Knorre *Usp. Biol. Khim.* **32** 3 (1991)
61. K N Bulygin, N B Matasova, D M Graifer, A G Veniyaminova, V I Yamkovoy, J Stahl, G G Karpova *Biochim. Biophys. Acta* (1997)
62. A A Malygin, D M Graifer, K N Bulygin, M A Zenkova, V I Yamkovoy, J Stahl, G G Karpova *Eur. J. Biochem.* **226** 715 (1994)
63. R Brimacombe, J Atmadja, W Stiege, D Schuler *J. Mol. Biol.* **199** 115 (1988)
64. D R Gentry, R R Burgess *Biochemistry* **32** 11224 (1993)

[a] — *Russ. J. Bioorg. Chem. (Engl. Transl.)*
[b] — *Mol. Biol. (Engl. Transl.)*
[c] — *Dokl. Chem. (Engl. Transl.)*