Russian Chemical Reviews **74** (1) 77 – 95 (2005) © 2005 Russian Academy of Sciences and Turpion Ltd

DOI 10.1070/RC2005v074n01ABEH000882

Covalent binding of modified nucleic acids to proteins as a method for investigation of specific protein – nucleic acid interactions

T S Zatsepin, N G Dolinnaya, E A Kubareva, M G Ivanovskaya, V G Metelev, T S Oretskaya

Contents

I. Introduction 77
II. Covalent binding of modified nucleic acids to proteins 78
III. Natural protein – nucleic acid conjugates formed in the enzymatic reactions 87
IV. Identification of protein sites covalently bound to nucleic acids 89

Abstract. Methods for specific covalent binding of modified nucleic acids to proteins and enzymes with conservation of their native structures are discussed. Possible applications of synthetic nucleic acids containing reactive (nucleophilic and electrophilic) groups and groups that can be photoactivated for testing protein – nucleic acid contacts and studying protein functioning are considered. The structures of protein – nucleic acid conjugates formed in the enzymatic reactions are described. Methods for the isolation and analysis of covalently bound protein – nucleic acid complexes are generalised. The main attention is given to methods for determination of protein sites attached to nucleic acids. The bibliography includes 226 references.

I. Introduction

Protein – nucleic acid interactions determine numerous processes occurring in cells, *viz*., replication, transcription and translation of DNA as well as RNA processing. Amino acid residues involved in the formation of protein – nucleic acid complexes are identified by various methods. These can be divided into two groups, *viz*., instrumental methods (*e.g*., X-ray diffraction analysis, NMR spectroscopy, *etc*.) and molecular-biological methods (site-directed mutagenesis of proteins, cross-linking, *etc*.).

Undoubtedly, the most reliable information about the structures of NA – protein conjugates can be obtained from X-ray diffraction data, however, isolation and crystallisation of the corresponding complexes are not always possible.† In addition, X-ray diffraction analysis often fails to establish the conformational reconstructions of the proteins in their interactions with NA and to determine the effects of the nucleotide sequences of NA flanking the protein-binding sites.[2] The dynamics of protein – nucleic acid complexes in solution is usually studied by NMR spectroscopy, but the latter is often inapplicable to the study of multisubunit complexes and processes associated with significant changes in the structures of biopolymers. Other spectroscopic methods, *e.g*., fluorescence spectroscopy including stop-flow technique, Raman and small-angle X-ray scattering and electron microscopy, provide little information about the structures of NA – protein complexes.

Among other molecular-biological methods used in the analysis of NA – protein interactions, site-directed mutagenesis and addressed chemical modification of proteins are most popular, but they are rather laborious and sometimes provide incorrect information due to changes in the biopolymer structure caused by the substitution of the functional amino acid residues. Moreover, this demands localisation of the protein domain involved in binding. This group of methods also includes covalent binding of proteins to NA, which is optimum for solution of many problems. In the first place, it allows identification of the protein domain and even particular amino acids interacting with a specific NA fragment within the NA – protein complex.[3 – 7] It is also used in the study of multiprotein complexes formed upon replication, translation, *etc*. In particular, this can also be employed for identification of a protein component interacting with a specific site in NA.[8] This approach can be used in the study of cell extracts (particularly, for detecting proteins selectively interacting with specific nucleotide sequences) and in the SELEX technology ‡ for selection of oligonucleotides with specific primary structures (*e.g*., NA aptamers) that bind to proteins irreversibly.[9, 10]

Covalent binding of proteins to NA has been used for the study of NA – protein interactions virtually from the moment of the origin of molecular biology, *i.e.,* since the mid-1960 s.[11] At present, photoaffinity modification of proteins is the most popular; this is based on covalent binding of non-modified nucleic acids to proteins upon UV irradiation.[12] However, it can hardly be regarded as a versatile procedure because of low yields of conjugates, radiation-induced damage of biopolymers and the necessity to identify the nucleotide unit to which the protein is attached. The use of NA analogues that can be photoactivated to generate radicals and biradicals under the influence of UV light seems to be a more promising approach. It excludes the necessity

T S Zatsepin, N G Dolinnaya, V G Metelev, T S Oretskaya Department of Chemistry, M V Lomonosov Moscow State University, Leninskie Gory, 119992 Moscow, Russian Federation. Fax (7-095) 939 31 81, tel. (7-095) 939 31 49, e-mail: tsz@yandex.ru (T S Zatsepin), tel. (7-095) 939 54 07, e-mail: dolinnaya@hotmail.com (N G Dolinnaya), metelev@belozersky.msu.ru (V G Metelev), tel. (7-095) 939 54 11, e-mail: oretskaya@belozersky.msu.ru (T S Oretskaya)
E A Kubareva, M G Ivanovskaya A N Belozersky Institute of Physico-Chemical Biology, Leninskie Gory, 119992 Moscow, Russian Federation. Fax (7-095) 939 31 81, tel. (7-095) 939 31 48, e-mail: kubareva@belozersky.msu.ru (E A Kubareva), tel. (7-095) 939 31 52, e-mail: img@genebee.msu.ru (M G Ivanovskaya)

Received 9 March 2003
Uspekhi Khimii **74** (1) 84 – 103 (2005); translated by R L Birnova

† Considerable progress in crystallisation of covalently bound NA – protein complexes and study of their structures by X-ray diffraction analysis achieved in the past decade is considered in detail in the review by Verdine and Norman.[1]

‡ SELEX is an abbreviation from 'Systematic Evolution of Ligands by EXponential enrichment'.

to identify the nucleotide unit involved in binding and diminishes the risk of destruction of biopolymers, since in this case irradiation is performed under less destructive conditions (*i.e.,* with longer wavelengths). Several reviews devoted to the synthesis,[13] properties [14] and applications [15–17] of NA analogues that can be photoactivated are available.

Chemical methods for the attachment of proteins to NA are based on either the use of non-specific binding reagents (*e.g.*, formaldehyde [18–20]), or selective introduction of reactive groups into the protein or NA. A novel procedure for the covalent binding of proteins to NA in the presence of metal complexes has been developed recently.[21] It was found that UV irradiation of DNA in the presence of complexes with platinum-group metals results in specific oxidation of guanosine residues within the oligonucleotide.[22–24] When such nucleotides come closer to the protein, cross-linking occurs. However, despite the obvious originality of this technique, it has not found wide application because of low yields of protein–NA conjugates and commercial inaccessibility of the most efficient metal complexes.

The present review is devoted to the use of modified NA for the study of specific protein–nucleic acid interactions (here, we do not consider the methods based on the use of chemically modified proteins).[25–27] The review by Sheflyan *et al.*[3] published in 1996 was devoted to methods for covalent attachment of nucleic acids and their derivatives to proteins; however, many new NA derivatives have been synthesised in the past 7 years and, most importantly, the methodology of determination of amino acid residues covalently bound to NA has been finally stated and brought to practice.

II. Covalent binding of modified nucleic acids to proteins

1. Methods for functionalisation of nucleic acids

Modified NA can be synthesised using chemical or enzymatic methods. The former approach includes preparation of modified nucleosides (as a rule, in the form of 3′-phosphoroamidite derivatives) which are further used in automated oligonucleotide synthesis. These nucleosides have to be stable under conditions of the synthesis and post-synthetic workup, which is not always possible. In addition, this method makes it possible to prepare oligomers having no more than 80–100 nucleotide units.

The enzymatic approach entails incorporation of modified 5′-nucleoside triphosphates into NA using various polymerases. As a rule, C(5)-substituted pyrimidine and C(8)-substituted purine nucleosides are used, the use of cytidine derivatives modified at N(4) has also been described. In this case, the length of the NA synthesised can reach thousands of nucleotides. However, the limitations of the enzymatic chain elongation method consist of impossibility of incorporation of certain modified nucleotides into the chain. The enzymatic method enables statistical functionalisation of NA and if 5′-nucleoside triphosphate terminates the growth of the oligonucleotide, it is the 3′-terminus of NA that is modified.

Substituents can be introduced into NA in the course of post-synthetic workup. Various reactive groups, *e.g.*, aliphatic amino groups, thiol groups, phosphorothionate internucleotide groups, *etc.*, can be used as substituents. An original procedure for the introduction of reactive internucleotide groups by chemical ligation of modified oligonucleotides has been developed in the laboratory headed by Z A Shabarova at MSU (see a review [28] and Section II.3).

2. Photoactivatable derivatives of nucleic acids

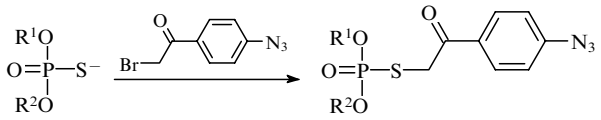
As mentioned above, the use of photoactivatable derivatives of NA in the study of specific protein–nucleic acid interactions is a promising approach. A comprehensive review by Meisenheimer and Koch [14] published in 1997 dealt with a detailed consideration of the mechanism and selectivity of covalent binding of side groups of amino acid residues of a polypeptide chain to the modified nucleotide units of nucleic acids containing photoactivatable groups.

Arylazide derivatives have become especially popular among other photoactivatable NA analogues due to their high reactivities.§ Nucleosides containing arylazido groups can be introduced into NA by both enzymatic and chemical methods.

The main disadvantages of arylazide derivatives of NA include the lack of selectivity in the binding to proteins and large size of the arylazido group, which can lead to disturbances in the local structure of an NA–protein complex and to cross-linking with remote protein fragments. Nevertheless, this type of modified NA is widely used in the study of structures of protein–nucleic acid complexes.

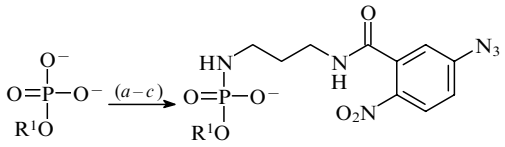
Arylazides can be introduced into NA

— at the internucleotide phosphorothioate bond of modified oligonucleotides: [29–31]

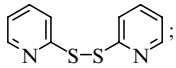


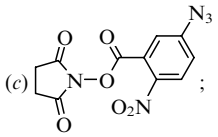
R^1,R^2 are fragments of the oligonucleotide chain;

— at the 5′-phosphate group of oligonucleotides:

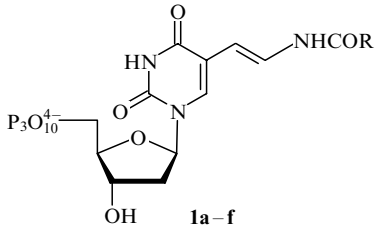


R^1 is a fragment of the oligonucleotide chain;

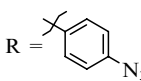
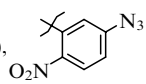
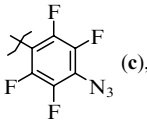
(*a*) Ph_3P, ; (*b*) $NH_2(CH_2)_3NH_2$;

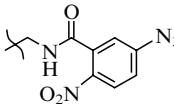
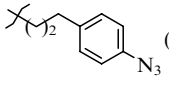
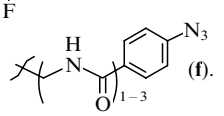
(*c*) ;

— in the course of a polymerase reaction using 5′-nucleoside triphosphates **1**–**3** modified at the heterocyclic base as one of the reactants (see Refs 32–43).

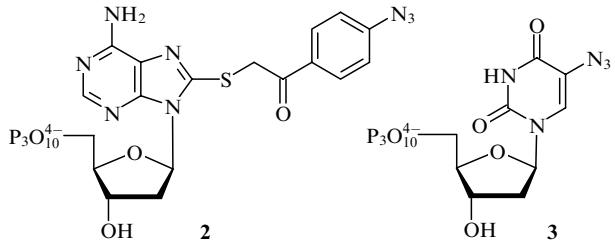


1a–f

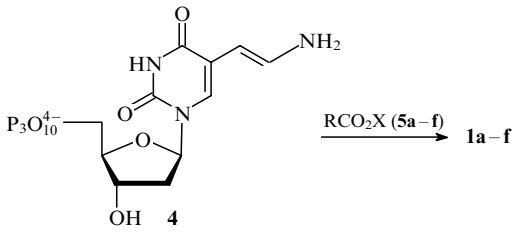
R =  (**a**),  (**b**),  (**c**),

 (**d**),  (**e**),  (**f**).

§ Highly reactive short-living arylnitrene biradicals, which react with the majority of amino acids, are formed upon irradiation of arylazido groups.



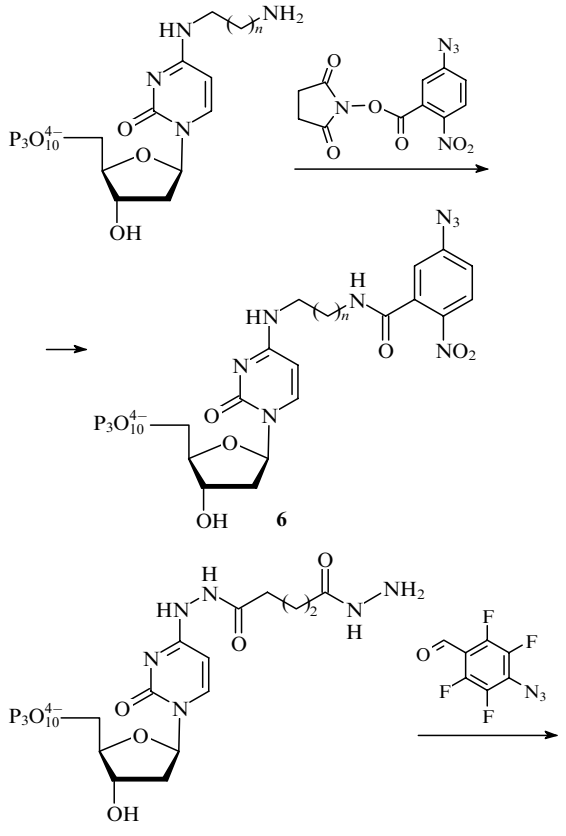
The photoactivatable triphosphates **1a – f** were prepared by the acylation of 5-(2-aminovinyl)-2′-deoxyuridine-5′-triphosphate (**4**) with an activated ester of the corresponding azidobenzoic acid (**5a – f**).

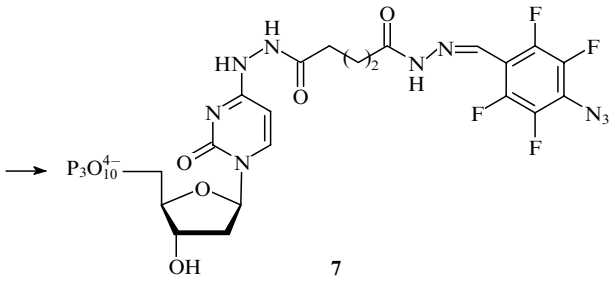


X = C_6F_5, succinimido.

The electron-withdrawing groups in the benzene ring of the arylazide stabilise the nitrene formed upon irradiation of the azide, as a result of which the yield of the protein – nucleic acid conjugate increases.[44]

The introduction of amino and hydrazide groups at C(4) of pyrimidine and their subsequent acylation or alkylidenylation is yet another approach to the synthesis of triphosphates containing an arylazide group. The synthesis of the corresponding compounds **6** and **7** has been described.[45 – 49]





The photoreagents **1 – 3**, **6** and **7** were used for the study of the human replication protein A,[32, 50 – 53] HIV-1 reverse transcriptase,[54] DNA polymerases,[55, 56] helicase Rep from *E. coli*[35] and excision repair proteins (BER, Base Excision Repair).[57, 58] The previously obtained data on the structures of DNA polymerases have been reviewed.[59] The use of arylazide derivatives of NA for structure – function studies has also been described.[15 – 18]

Lavrik *et al.*[60] have developed an original procedure for sensitised photomodification of DNA polymerase β which involves the use of an arylazido group. First, a nucleotide containing an arylazido group at the 3′-end was synthesised from a primer using DNA polymerase β. Then, pyrene-containing uridine-5′-triphosphate that binds to polymerase is added to the reaction mixture, and the resulting complex is irradiated with light with the wavelength of 365 – 450 nm. Irradiation induces the excitation of the pyrenyl residue and the fluorescence-resonance-energy transfer (FRET) to the arylazido group (Fig. 1). The activation of the latter brings about covalent binding of the 3′-end of the primer to the polymerase. Such a binary system allows the use of significantly longer irradiation wavelengths required for the formation of NA – protein conjugates in comparison with those commonly used for the photoactivation of arylazide derivatives (325 – 350 nm). The use of this approach has made it possible to reduce the photodamage of DNA and proteins and to increase (by an order of magnitude) the initial rate of photomodification. Later, this approach was repeatedly used for the study of functioning of DNA polymerases in both model systems and cellular extracts.[54, 61 – 65]

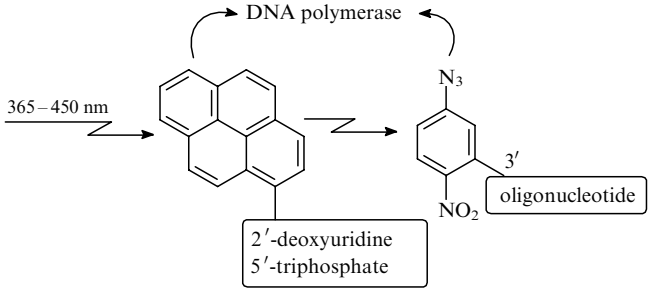
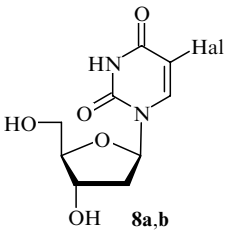


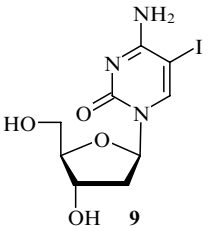
Figure 1. Sensitised modification of DNA polymerase.

Among other commonly used photoactivatable NA derivatives, mention can be made of 5-halogenopyrimidine and 8-halogenopurine nucleosides, *e.g.*, 5-bromo- (**8a**)[33, 66 – 72] and 5-iodo-2′-deoxyuridine (**8b**),[31, 40, 67, 73 – 95] 5-iodo-2′-deoxycytidine (**9**)[77, 96] and 8-bromo-2 -deoxyguanosine (**10**).[97]

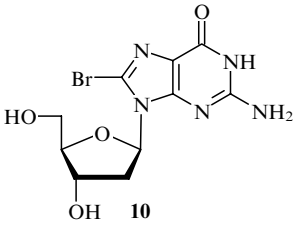


8a,**b**

Hal = Br (**a**), I (**b**).



9

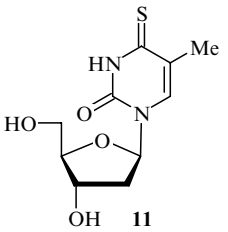


10

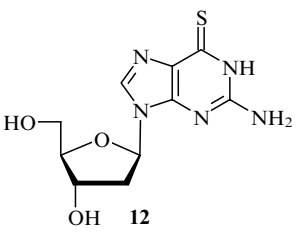
Photolysis of these photoactivatable NA produces stable radicals which react with electron-donor amino acid residues of proteins, particularly, with aromatic and sulfur-containing side groups. In the case of bromo derivatives, the optimum wavelength of the UV light necessary for the abstraction of the halogen atom is 308 nm, while that for iodo derivatives is 325 nm. The increase in the wavelength reduces photoinduced damage of proteins and NA, but decreases the yields of the photoaddition products. By virtue of their small size (the van der Waals radii of iodine and bromine atoms correspond approximately to that of the methyl group), halogen atoms do not induce any significant changes in the structure of NA and do not prevent protein–nucleic acid recognition. This makes halogen derivatives of NA highly attractive tools for probing protein–nucleic acid contacts in the binding sites of biopolymers.

Halogen derivatives of NA were used in the studies of a broad array of NA-recognising proteins, such as the transcription factor NF-κB;[33] the transcription factor binding the CCAAT motif of DNA (CBF);[69] the restriction–modification enzymes EcoP15I,[66] EcoRI,[78,87] EcoRII,[93,94,96] EcoRV,[84,95] SsoII,[88] CviBIII and TaqI;[81] homing endonuclease PI-SceI;[79,80] the eukaryotic ribosome L5 protein;[89] the herpes simplex virus helicase—primase subcomplex;[83,90] the replication terminator Tus,[67] HIV-1 integrase;[74,77] the repair protein MutS;[77] the protein HMG1;[91] the replication protein A;[82] the recombination proteins RAG1 and RAG2;[85] telomerase of *Euplotes aediculatus*;[75] the protein SSB;[92] the fibroblast growth factor ($bFGF_{155}$);[70] vimentine;[71] the phage capsid protein MS2[73] as well the multiprotein centromeric complex (CBF3).[68]

Nucleoside derivatives in which the oxo group in the heterocyclic base is substituted for the thioxo group, *e.g.,* 4-thiothymidine (**11**)[34,98–100] and 6-thio-2′-deoxyguanosine (**12**),[101–103] belong to yet another class of photoactivatable NA derivatives.



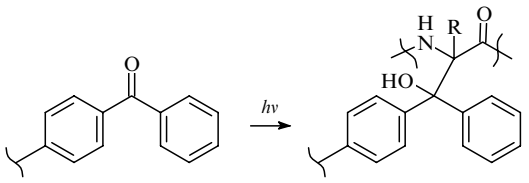
11



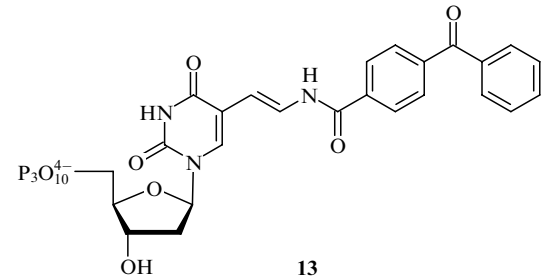
12

Replacement of the oxygen atom by sulfur induces minimum structural changes in the biopolymer, which, together with the high reactivity of thiones, makes them rather promising compounds. Upon irradiation with light with the wavelengths > 300 nm, these derivatives react with proteins by a free-radical mechanism. They were successfully used in the studies of the binding site of TAR RNA to the Tat-protein of HIV-1[98,103] (testing of NA–protein contacts in the large groove of the double helix of RNA) as well as of DNA methyl transferase EcoR124I,[102] the transcription factor SRF,[101] integrase of the bacteriophage lambda[100] and the protein Rev of HIV-1.[99]

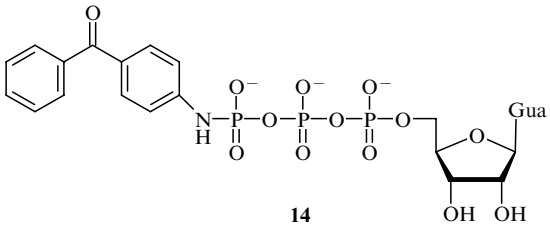
It is known that the benzophenone residue incorporated into NA[40,104] reacts with the α-carbon atom of the polypeptide chain upon UV irradiation.[105]



It seemed therefore interesting to consider possible applications of such NA for the study of NA–protein interactions. Thus triphosphates **13** and **14** containing a benzophenone residue were tested as photoreagents for the study of the transcription complex of yeast RNA polymerase III[40] at the RNA-binding site of RNA polymerase T7.[104]



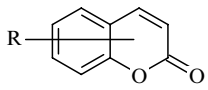
13



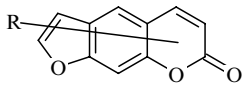
14

With rare exception, the yields of cross-linked products of such NA derivatives to proteins are extremely low; therefore, they have not found wide application.

Coumarins (**15**) and furocoumarins (**16**, psoralenes)¶



15



16

are widely used for the introduction of various non-radioactive labels[107,108] and the synthesis of covalently bound NA[109] as well as in the photochemotherapy[110] and diagnostics of genetic diseases.[111]

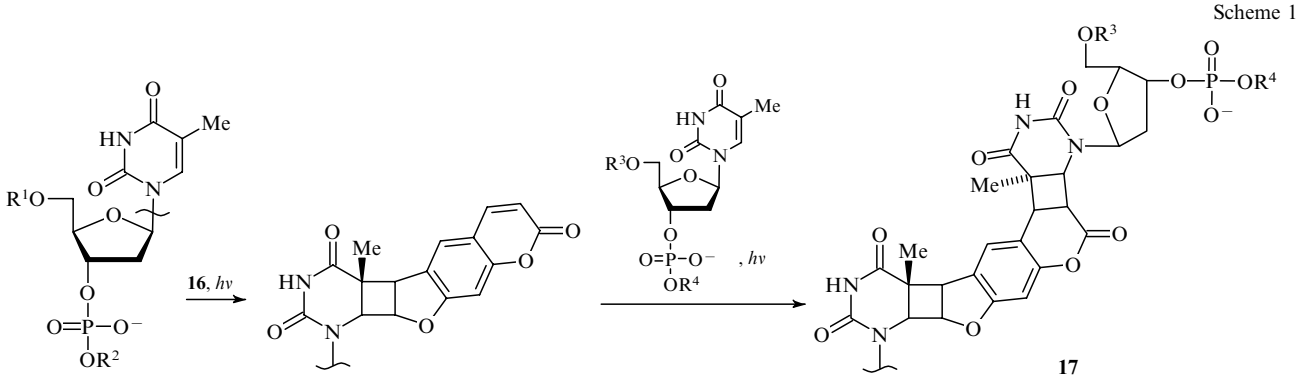
Psoralenes contain two photoactivatable double bonds by virtue of which they add to the double bond C(5)=C(6) of thymine to give compound **17** (Scheme 1). This reaction occurs upon UV irradiation (λ = 300–340 nm). It was suggested[112–115] to use DNA fragments containing the thymidine residue **18** modified by 4′-hydroxymethyl 4,5′,8-trimethylpsoralene for photoaffinity modification of proteins.

Coumarin can be introduced into NA by direct incorporation, into the oligonucleotide chain, of compound **19** with a carbohydrate-mimicking fragment.[111,116]

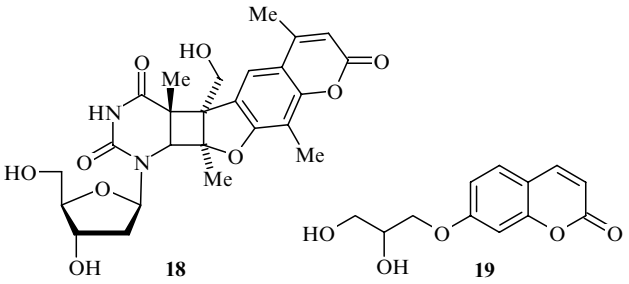
Psoralene- (or coumarin-)modified NA were used for the study of T7 RNA polymerase,[106,112–114] proteins RecA and Max,[114] as well as complex TFIIH proteins.[115]

¶ Psoralenes are also used for the modification of protein chains *in vivo*.[106]

Scheme 1

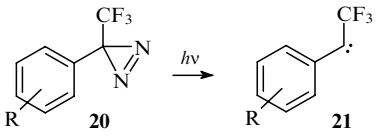


$R^1 - R^4$ are fragments of the oligonucleotide chain.

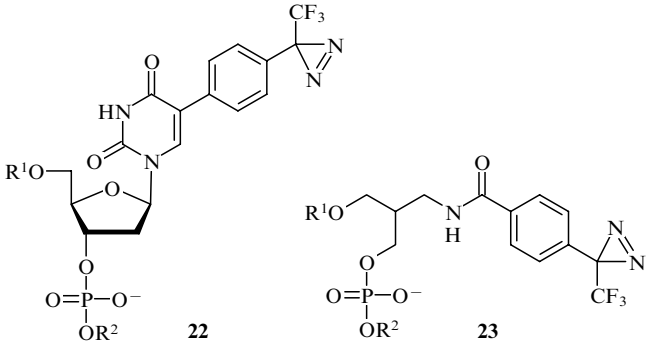


The optimum wavelength of UV light triggering the formation of protein – nucleic acid cross-links is 300 nm. The mechanism and selectivity of photoaddition of amino acid residues of proteins to psoralene-containing NA have not been established.

Aryl(trifluoromethyl)diazirine is yet another photosensitive reagent for the modification of NA. Upon UV irradiation, diazirines **20** are converted into reactive carbenes **21**, which react with virtually any molecule.



The review [117] published in 2000 is devoted to the use of aryl(trifluoromethyl)diazirines for the study of NA – protein interactions. Later, three other papers [94, 118, 119] devoted to the use of NA **22** and **23** containing diazirine residues were published.



R^1, R^2 are fragments of the oligonucleotide chain.

DNA methyl transferase EcoRII [94] and HIV-1B integrase [118] were used as subjects of studies.

Unfortunately, the yields of cross-linked binding products of diazirine derivatives of NA to proteins are low because of side reactions of carbenes with water.

3. Chemical modification of nucleic acids

The non-specificity of the above-described photoactivatable NA reagents is their main disadvantage. Besides, their activation requires UV irradiation which can damage biopolymers. In contrast, NA reagents containing reactive groups react selectively with spatially close amino acid residues of the protein without any external factors. The introduction of a reactive group into a preset position of the oligonucleotide chain allows the probing of point contacts between different regions of biomolecules, whereas variations of the chemical nature of the reactive group enable selective binding of a definite type of amino acid residues within the NA – protein complex.

Reactive groups can be introduced into any position of NA, *i.e.*, the heterocyclic base, the sugar – phosphate backbone or the ends of the oligonucleotide chain. The introduction of these groups into the sugar – phosphate backbone of NA is the most promising approach, while modification of heterocyclic bases can disturb complementary interactions of NA and specific interactions within the protein – nucleic acid complex.

The following properties should be characteristic of the NA reagents used for the cross-linking with proteins:

(1) the possibility to react specifically with spatially close functional amino acid groups of the protein;

(2) formation of stable conjugates with the protein without any additional activation at physiological values of pH (6.0 – 8.0);

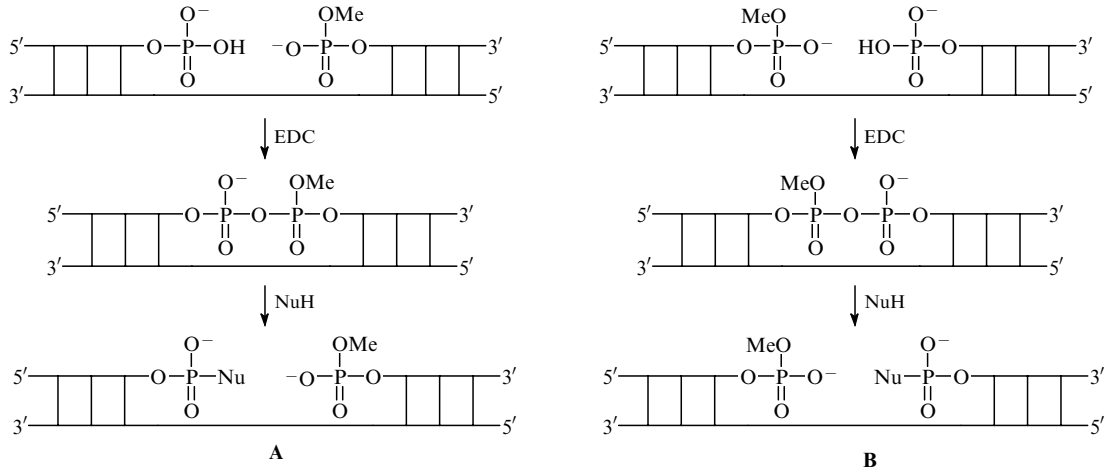
(3) retention of stability in neutral aqueous solutions over a long period of time without any loss of reactivity.

a. Modification of nucleic acids at the internucleotide phosphodiester bond

It is known that nonspecific interactions make a significant contribution to the structural organisation and stability of virtually all macromolecular complexes including NA – protein complexes. In this connection, modification of the phosphate groups of NA, which are involved in the majority of cases in these interactions, seems to be important for the study of NA – protein contacts. In the early 1990s, Shabarova *et al.*[120 – 123] proposed to substitute trisubstituted (TSP) groups for the natural 3′,5′-phosphodiester bonds in NA.

The trisubstituted pyrophosphate group possesses all the properties necessary for the interaction with proteins, *viz.*, it is stable in aqueous media and is readily cleaved by amines. It was found that oligonucleotides containing TSP groups react exclusively with lysine and histidine residues as well as with the terminal α-amino group of the protein;[124] the nucleophilic substitution occurs only with the disubstituted phosphate to yield a covalently linked protein – oligonucleotide complex. If the phosphoramide bond is formed with involvement of the ε-amino group of Lys or the terminal α-amino group of the protein, this bond is stable at pH > 4; if the amino group of His is involved, the bond is cleaved in weakly acidic media or under the action of *N*-methylimidazole, which allows discrimination of lysine and histidine residues.[125]

Scheme 2



NuH are Lys and His residues of the protein.

Template-directed condensation of oligonucleotides carrying mono- and disubstituted terminal phosphate groups using 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) (Scheme 2, chemical ligation) is employed for the introduction of the TSP group into the predetermined position of the oligonucleotide. The use of chemical ligation for the preparation of various single- and double-stranded TSP-modified oligonucleotides has been described in detail in the reviews. [3, 18, 126]

The formation of two isomeric TSP-containing oligonucleotides **A** and **B** (see Scheme 2) is possible depending on which of the terminal phosphate residues (3′- or 5′-) in the starting oligonucleotide is esterified. It was found that the synthesis of the isomer **B** is far more effective than that of the isomer **A** due to the difference in the reactivities and availabilities of the reacting mono- and disubstituted phosphate residues. At the same time, the isomer **B**, which is formed more easily, reacts more slowly with the amino group of lysine than the isomer **A**.

The human transcription factor NF-κB was used as a target protein.[127, 128] This belongs to the class of multifunctional eukaryotic proteins (the family Rel), which control gene expression and play an important role in such vital processes as immune response,[129] apoptosis,[130] cell growth [131] and proliferation.[130] It has been shown that synthetic DNA duplexes containing TSP groups in different positions of the recognition site of the transcription factor NF-κB differ substantially in the efficiency of their covalent binding to the protein. Thus in the case of isomers **A** the yield of the protein – nucleic acid conjugate was 35%, while for isomers **B** it did not exceed 7%. The observed regularity was taken into consideration in subsequent studies aimed at the design of DNA reagents containing the TSP groups.

A series of DNA duplexes containing *O*-methyl pyrophosphate in different positions of the binding site of the transcription factor HNF-1 was used for the probing of DNA contacts with this protein in nuclear extracts of rat liver [132] and the study of DNA formamido – DNA glycosylase contacts.[133]

The synthesis of TSP-containing oligoribonucleotides has been described by Shabarova *et al.*[134] These were used to confirm the existence of contacts between individual phosphate groups of RNA within the TAR RNA complex with the Tat-peptide [135] predicted on the basis of NMR spectroscopic data. The aim of this study was to specify the architecture of the natural TAR RNA – Tat-protein complex. The amino acids interacting with the RNA duplex were determined using chemical and enzymatic cleavage of the protein fragment in the covalently bound complex with subsequent analysis of the conjugate by mass spectrometry.[135] The natural TAR RNA – Tat-protein complex plays a key role in the activation of transcription of the human immunodeficiency virus. It was found that the Tat-protein after formation of the complex with TAR RNA initiates indirect activation of RNA polymerase II.[136] This reaction is the determining step in the initiation of transcription of HIV-1, since no elongation of transcription takes place in the absence of the Tat-protein.

The structural polymorphism of proteins including that induced by irradiation and crystallisation has been studied especially intensely in recent years. It remains to be hoped that the use of nucleic acids containing TSP groups will enable a more accurate determination of individual contacts between the biopolymers within the protein – nucleic acid complexes which exist in the native state in aqueous media and a comparison of the results obtained from X-ray diffraction analysis. The main advantages of this approach include selectivity of interaction (Lys or His) and the absence of extended spacers between the phosphate group responsible for the binding and the nucleophilic groups of proteins. Thus some of the contacts of the Lys and His residues in RNA polymerase of bacteriophage T7 with the phosphate groups of the promoter were established [137] with the use of synthetic fragments of the promoter (10 – 20 b.p.) containing a TSP group in positions that, according to data from X-ray diffraction analysis, are close to particular amino acids. The functional significance of such contacts has been established previously in biochemical studies.

This approach was used in the study of the complex of the transcription factor NF-κB with DNA duplexes with different primary structures.[128] Covalent binding made it possible to establish an important fact that in principle the structure of the NF-κB complex with DNA is determined not only by the nucleotide sequence in the protein-binding site, but also by flanking regions. Later, this fact was confirmed by X-ray diffraction analysis.[138 – 141]

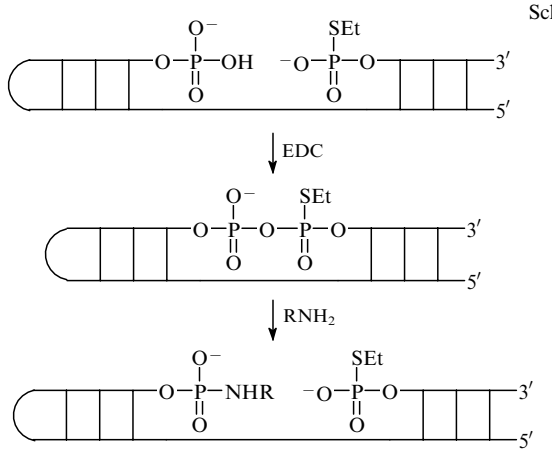
Studies by Shabarova *et al.*[142] have shown that DNA duplexes containing TSP groups are promising reagents for the probing of conformational changes in the enzyme during functioning of the enzyme – substrate complex. Restriction endonuclease MvaI (R.MvaI) was used as a subject of investigation, while a 14-membered DNA duplex containing a recognition site for the enzyme with a TSP group instead of a hydrolysable phosphodiester bond was used as a substrate.[142] This system was used for the study of conformational changes in the restriction endonuclease MvaI in the enzyme – substrate complex induced by different concentrations of magnesium ions (the latter is a cofactor of the cleavage of DNA). It was found that magnesium ions favour conformational rearrangements in the protein as a result of which one of the nucleophilic groups in the catalytic site of R.MvaI (presumably, histidine) comes closer to the modified bond in the DNA substrate.

In order to enhance the reactivities of internucleotide bonds, it was proposed to replace the TSP groups in DNA duplexes by their sulfur-containing analogues.[143] Oligonucleotides containing thio

analogues of the TSP groups appeared to be more reactive with respect to aliphatic amines, since dialkyl thiophosphoric acid is a stronger acid (in comparison with dialkyl phosphoric acid) and, as a consequence, the better leaving group.

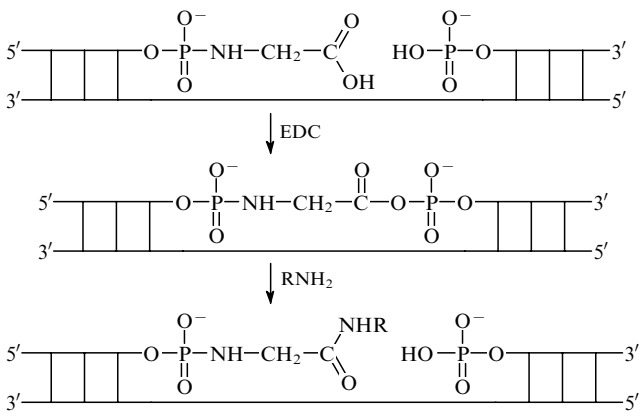
Hairpin DNA duplexes containing a recognition site of the human transcription factor NF-κB, the position of the modified fragment (*i.e.,* the sulfur-containing TSP group) in which corresponded to the position of the group the most close to the lysine residue, were bound covalenty to the protein (Scheme 3). In contrast to oxygen-containing TSP, both isomers differing in the position of the trisubstituted phosphorothioate residue reacted with the same efficiency. However, the lability of sulfur-containing TSP groups and the relative difficulty of their incorporation into oligonucleotides prevent large-scale application of this type of DNA reagents.

Scheme 3



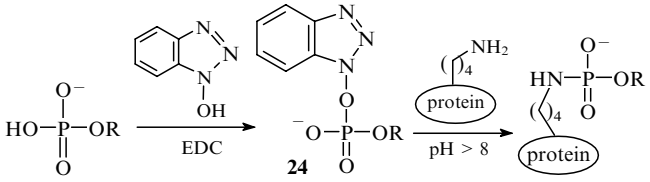
It was proposed [124, 144] to use oligonucleotides containing a reactive acylphosphate group instead of one phosphodiester bond for the study of NA – protein interactions. This group was introduced into the preset position of the sugar – phosphate backbone by template-directed chemical ligation of two oligonucleotides one of which served as a phosphate donor, while the other one was used as a carboxyl donor (Scheme 4).[124] The mixed anhydride formed undergoes cleavage in aqueous solution (pH $\geqslant$ 8) under the action of amines to give a carboxamide (rather than phosphoroamide) (see Scheme 4).[124] In reactions with amines, such acyl phosphate-containing oligonucleotides play the role of mild acylating reagents. A comparison of the properties of DNA duplexes containing an acyl phosphate or a TSP group in the same position of the recognition site of the transcription factor NF-κB revealed that the yields of the protein – nucleic acid conjugates in the case of acyl phosphate-containing DNA were three times lower than those from TSP-containing DNA.[124] The lower reactivities and the formation of side products in the synthesis of acyl-phosphate oligonucleotide derivatives are the reasons for their low utility for the covalent binding to proteins.†

Scheme 4



b. Modifications of the terminal phosphate groups of nucleic acids

Oligonucleotide derivatives with activated terminal phosphoester groups are more accessible than other reactive phosphorylating agents. *N*-Hydroxybenzotriazole esters (HOBT-derivatives) of oligonucleotides **24** are the most popular reagents of this type. They are prepared in one step from oligonucleotide-3′- or 5′-phosphates and *N*-hydroxybenzotriazole under the action of water-soluble carbodiimide in water-organic media.[147, 148]



R is the oligonucleotide residue.

HOBT-Derivatives of the oligonucleotides **24** manifest phosphorylating properties in reactions with amino acids, peptides and proteins. In this case, the reaction with the protein occurs at the N-terminal α-amino group, at the ε-amino group of Lys or at the imidazole residue of histidine.[149]

Reactive *N*-hydroxybenzotriazole esters of oligonucleotides were successfully used in the study of DNA-[125] and RNA-binding [149] proteins. The structure of the complex of RNA polymerase of *E. coli* with a DNA promoter has been studied [125] with the help of a large series of hydroxybenzotriazole derivatives of single- and double-stranded oligonucleotides corresponding to various fragments of the promoter responsible for different transcription initiation steps (−35 and −10 relative to the transcription initiation point). The RNA polymerase subunit of the holoenzyme built up of five subunits and the amino acid residues involved in the binding to each oligonucleotide were determined based on the results of covalent binding.[125] The results obtained with the use of hydroxybenzotriazole derivatives were confirmed by Naryshkin *et al.*[150] who have used photoactivatable arylazide analogues of DNA for the solution of the same task.

The use of hydroxybenzotriazole derivatives of RNA has also been described in the study [149] on the binding of Tat-proteins to the fragment of TAR RNA. The high efficiency of the covalent binding of RNA to the peptide suggests that this type of derivatives can be used for the construction of 'decoys' that slow down the transcription of the immunodeficiency virus in infected cells.

It should be noted that HOBT and TSP derivatives of nucleic acids containing modified phosphate groups both inside the NA chain and at its termini are the only currently available reagents for testing the contacts of Lys and His residues of proteins. The high efficiency of formation of protein – nucleic acid conjugates (yields up to 50%), the high selectivity of binding to proteins and the availability of chemical means of analysis of covalent binding products open up broad opportunities for further applications of these reagents for determination of the structures and analysis of conformational rearrangements in NA – protein complexes during their functioning or upon changes in their environments.

c. Modification of sugar moieties of nucleic acids

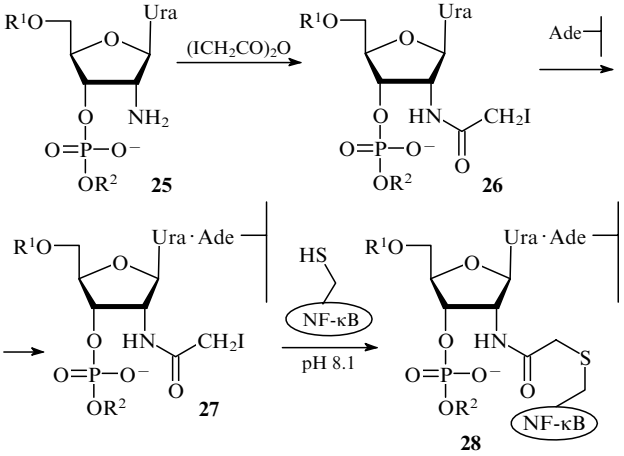
Only a small number of NA reagents containing reactive groups in the sugar moiety have been obtained so far. At the same time, it is known that NA-recognising proteins react with NA exclusively by generating hydrophobic and van der Waals contacts with the

† The methods for the introduction of carboxy groups into DNA and the chemical properties of such derivatives have been studied fairly extensively.[145, 146]

ribose residue.[151] Of special interest are NA reagents modified in the 2′-position of ribose which plays a crucial role in the discrimination of DNA and RNA molecules by proteins.

It was suggested [152] that the 2′-iodoacetamide group within DNA duplexes can be used for selective covalent binding of proteins containing a cysteine residue in the active centre. The iodoacetamido group was chosen due to its higher (in comparison with the bromoacetamido group) activity in the alkylation of the thiol group of cysteine.[153] Synthetic oligonucleotides containing 2′-amino-2′-deoxyuridine **25** in the preset position instead of thymidine were used as starting compounds for the directed introduction of the iodoacetamido group (this approach enables easy changes in the position of the reaction centre in the target oligonucleotide sequence). The 2′-deoxy-2′-iodoacetamidouridine derivatives **26** were prepared by acylation of the amino group with an excess of iodoacetic anhydride (Scheme 5).

Scheme 5



R^1, R^2 are fragments of the oligonucleotide chain.

Studies of reactions of DNA reagents containing an iodoacetamide group in position 2′ of the uridine residue with various low-molecular-mass nucleophiles (including amino acids and peptides) revealed that they react only with thiol-containing compounds, particularly, with cysteine-containing proteins. It is of note that the iodoacetamide groups at the 2′-position of the sugar – phosphate backbone of DNA react with the cysteine residues of the protein only within specific protein – nucleic acid complexes (*e.g.*, within the duplex **27** , see Scheme 5).

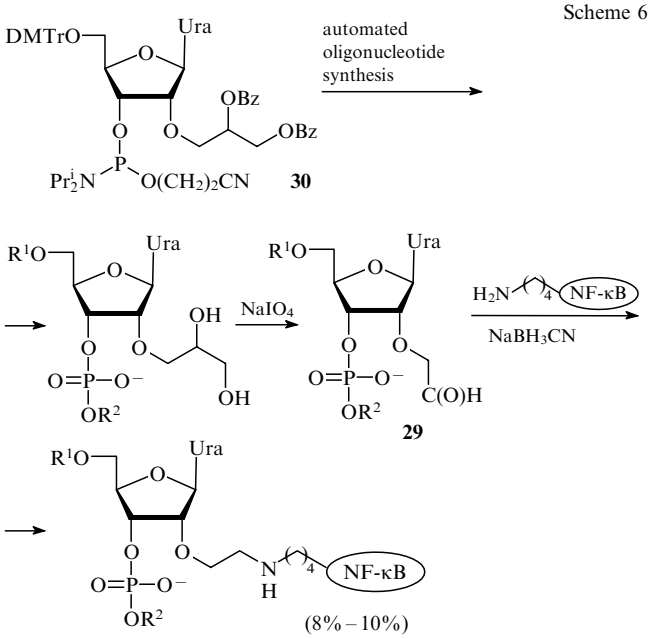
Such DNA duplexes were used for the affinity modification of a series of DNA-recognising cysteine-containing proteins, *e.g.*, the transcription factor NF-kB, the restriction – modification enzyme SsoII and uracil DNA glycosylase.[152] Thus the reaction of the modified DNA duplexes **27** with the transcription factor NF-kB is complete within 15 min at pH 8.1 and yields the covalently bound adducts **28**. These data altogether suggest that this reaction occurs with good efficiency (the yield of the conjugates reaches 25%) and high specificity.

The amino groups of lysine residues of DNA-recognising proteins can also be involved in the interaction with the sugar – phosphate backbone of DNA. To this end, aldehyde groups are to be introduced into the sugar – phosphate backbone. The synthesis and properties of oligonucleotides containing dialdehyde groups are considered in the reviews.[154 – 156]

Synthesis of oligodeoxyribonucleotides **29** containing a 2-oxoethyl group at the O(2′) atom has been described.[157, 158] The protected phosphoramidite 2′-*O*-(2,3-dihydroxypropyl)uridine derivative **30** was used as a starting compound in the oligonucleotide synthesis of such DNA (Scheme 6).

Oxidation of the diol group into the aldehyde group was carried out under mild conditions by treatment with a twofold

Scheme 6

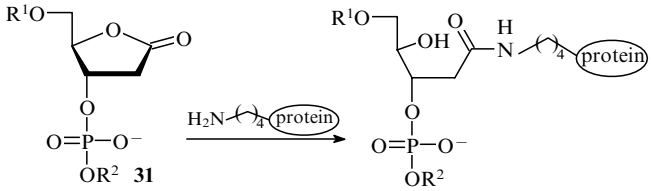


DMTr is dimethoxytrityl;
R^1,R^2 are fragments of the oligonucleotide chain.

excess of sodium periodate. The oligonucleotides **29** obtained react efficiently with different amino compounds.[159] The possibility of covalent binding of DNA containing a 2′-*O*-(2-oxoethyl)-uridine fragment to the transcription factor NF-kB (see Scheme 6) has been demonstrated.[160] The oligonucleotide **29** appeared to be more efficient in the reaction with NF-kB than the dialdehyde derivatives, since in this case β-elimination is excluded and the small size of the 2′-*O*-(2-oxoethyl) residue allows the ε-amino groups of lysine to approach the most close to DNA. The advantage of this methodology is the possibility to vary the position of the modified unit and the number of modifications in the oligonucleotide. The yields of the addition products of the transcription factor NF-kB to DNA ligands were 8% – 10%.

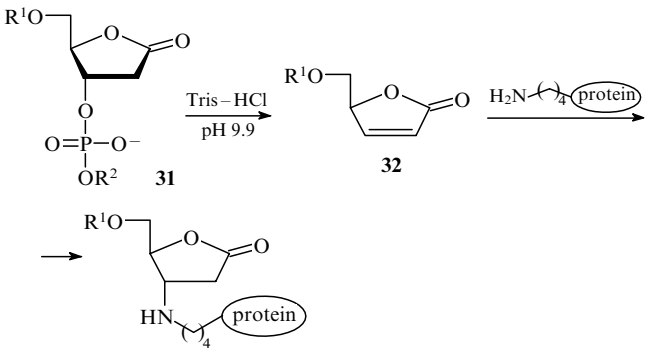
It was proposed [161] that the DNA ligands **29** were used for the study of the mechanism of functioning of DNA methyltransferase SsoII. With a set of duplexes containing one or two modifications in the recognition site, two types of DNA – protein conjugates have been obtained in 70% yields. From determination of molecular masses of such conjugates it was suggested that DNA methyltransferase SsoII interacts with DNA as a dimer.

Oligonucleotides **31** containing a 2-deoxyribonolactone fragment were used for the covalent binding of some repair enzymes.[162, 163]



R^1,R^2 are fragments of the oligonucleotide chain, the protein is endonuclease III or human DNA polymerase.

Such oligonucleotides appeared to be versatile for the covalent addition of various excision repair proteins. It was suggested [164] that the oligonucleotides **31** were used for the synthesis of oligonucleotides **32** containing furan-2(5*H*)-one at the 3′-end of the chain.



R^1,R^2 are fragments of the oligonucleotide chain.

The latter represent substrates for only those BER enzymes that catalyse β,δ-elimination of the apurine site, which allows selective binding of these proteins, particularly, in cellular extracts.

d. Modification of nucleic acids at the heterocyclic bases

It has been noted above that the introduction of reactive groups into heterocyclic bases is seldom used for modification of NA. Nevertheless, such examples are known. In particular, substituents containing −SH or −SS groups were introduced into the heterocyclic base in order to achieve covalent binding of cysteine-containing proteins to DNA through formation of disulfide bridges. Introduction of these groups into the oligonucleotide allows its covalent binding to the protein either by thiol – disulfide exchange or as a result of oxidation of the thiol groups under the action of atmospheric oxygen.

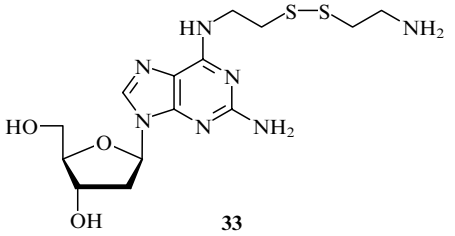
The thiol – disulfide exchange

$$R^1SH + R^2SSR^3 \longrightarrow R^1SSR^2 + R^3SH$$

proceeds efficiently at room temperature in water at physiological values of pH despite the fact that this reaction includes the cleavage and formation of a stable covalent S−S bond with the energy of 60 kcal mol^{-1} (see Ref. 165). The formation of the disulfide bond through the oxidation of thiols by oxygen occurs in two steps.[166]

$$R^1SH + O_2 \xrightarrow[-H_2O_2]{H_2O} R^1SOH \xrightarrow[-H_2O]{R^2SH} R^1SSR^2.$$

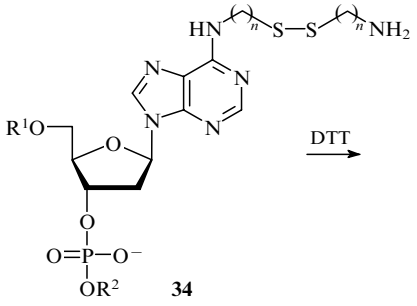
For the first time, the formation of a disulfide bond in the DNA – protein complex has been used in order to confirm the specificity of binding of a modified oligonucleotide ligand to the active site of (methylguanine) – DNA methyltransferase (MGMTase) of *E. coli* containing a cysteine residue.[167] The action of this key repair enzyme is directed at the removal of the methyl and other alkyl groups from position O(6) of guanine or from position O(4) of thymine by their transfer to the cysteine residue in the active site of the protein. In the 'dormant' state, the cysteine residue is localised inside the protein globule.[168] It was assumed that the binding of MGMTase to the substrate or its repair are accompanied by conformational rearrangements; however, the information on the nature of these changes was absent. In order to corroborate this hypothesis, 2-amino-6-cystamino-2′-deoxy-β-D-ribofuranosyl purine (**33**) was introduced into the 21-membered oligodeoxyribonucleotide instead of deoxyguanosine by automated synthesis.

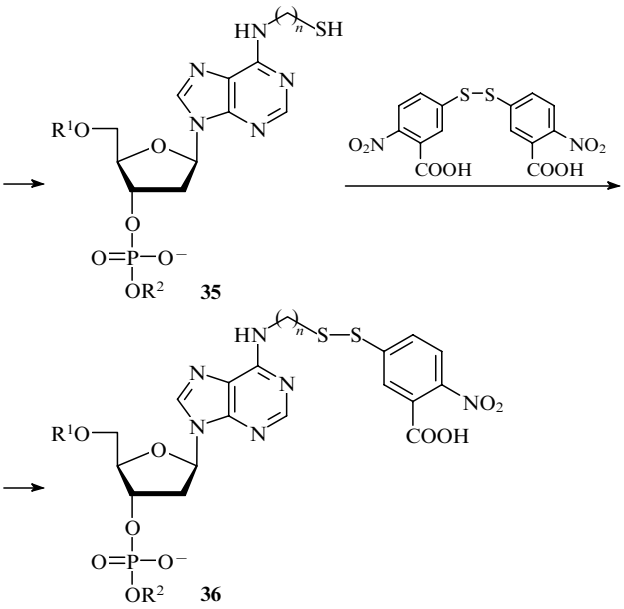


The modified oligonucleotide containing the fragment **33** was incorporated into NA and the disulfide bond was reduced by dithiothreitol (DTT). A covalent bond between the double-stranded oligonucleotide and MGMTase was formed under the action of atmospheric oxygen (the yield of the NA – protein conjugate was low). The formation of the disulfide bond between DNA and the protein was confirmed by its cleavage with DTT. Analysis of the primary structure of the polypeptide fragment covalently bound to the oligonucleotide confirmed the fact that it was cysteine that was involved in the formation of the disulfide bridge. Supporting evidence for the involvement of this amino acid residue in the formation of the covalent complex MGMTase – DNA was obtained in the following experiment. The cysteine residue in the active site of the enzyme was methylated by its incubation with an oligomer containing an *O*(6)-methylguanosine residue. The protein obtained did not form conjugates with reduced DNA substrates containing the nucleoside analogue **33** with a free thiol group. These studies have shown that MGMTase undergoes conformational rearrangements during its functioning which favour the convergence of the cysteine residue to the alkylated guanosine derivative.[167]

A similar approach based on the site-specific attachment of proteins to reactive analogues of DNA aimed at the formation of disulfide bonds was used [169] in order to establish the contacts in the active complex of HIV integrase with viral cDNA. At the beginning, additional cysteine residues were introduced into the predetermined positions of integrase by site-directed mutagenesis and modified fragments of 2′-deoxyadenosine containing a disulfide group at the N(6) atom were introduced into the predetermined positions of DNA by automated synthesis. It was expected that the cysteine residue in the DNA – protein complex would be localised in the vicinity of the modification site of DNA, since thiol – disulfide exchange demands a more 'close' arrangement of two interacting groups than photo-induced covalent addition.

The reactivity of DNA **34** containing a disulfude group could be enhanced by converting it into the thiol form by treatment with DTT; the reaction of DNA **35** with the cysteine residue of a protein occurs upon oxidation with atmospheric oxygen. Treatment of DNA **35** with 5,5′-dithiobis(2-nitrobenzoic acid) yields an even more reactive DNA derivative **36**.





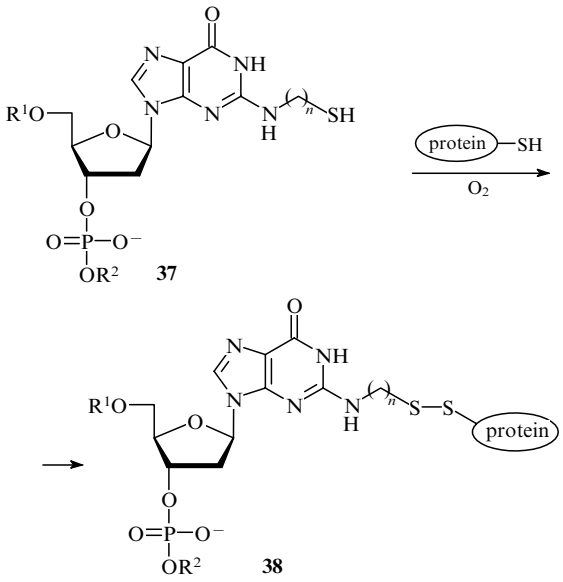
R^1,R^2 are fragments of the oligonucleotide chain; $n = 2, 3$.

It is of note that the introduction of reactive groups at N(6) of adenine does not influence its ability to form Watson–Crick hydrogen bonds.

It was found that the yields of covalent binding products of the highly reactive substrates **36** to all mutant forms of integrase containing single additional cysteine residues are nearly identical. The use of less reactive substrates **35** or **34** is recommended to increase the specificity of the covalent binding. In fact, the required specificity of the covalent binding was achieved with the least reactive oligonucleotides **34**: a noticeable yield of the conjugate was observed only for one mutant form of integrase. It was shown also that the use of a propylene linker instead of an ethylene linker increases significantly the yields of the protein–nucleic acid conjugates. The effect of 2-mercaptoethanol as an additive on the yields of the protein–nucleic acid conjugates has also been examined.[167] It was shown that high concentrations of 2-mercaptoethanol (> 5 mmol litre^{-1}) block completely the formation of the reaction product; however, lower concentrations (2 mmol litre^{-1}) enhance the specificity of the covalent binding of HS- and SS-containing DNA. Under these conditions, the formation of the protein–nucleic acid compex is reversible and spatially close arrangement of the reactive groups is the only prerequisite for cross-linking.

These studies have demonstrated the *trans*-interaction between viral cDNA and integrase, *i.e.*, the enzyme–substrate complex represents a DNA bound to the C-terminal domain of one integrase monomer and the catalytic domain of the other integrase monomer. Moreover, site-specific covalent binding revealed conformational changes in the integrase–DNA complex concomitant with the terminal processing reaction.

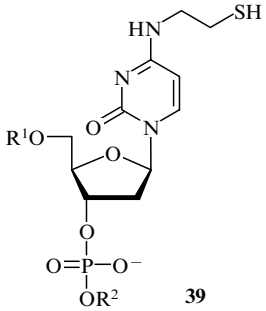
Oxidation of thiol groups by atmospheric oxygen was successfully used in the synthesis of a covalently bound complex of reverse transriptase of HIV-1 with a DNA duplex consisting of a matrix and a primer. This reaction was carried out in the presence of deoxynucleoside triphosphates, which provided the formation of a functionally active supramolecular associate.[170, 171] The reactive mercaptoalkyl group was introduced at the amino group in position 2 of guanine (compound **37**). This group was localised in the small groove of DNA and was spatially close to the cysteine residue‡ of the H helix of reverse transcriptase. One of the conjugates (complex **38**) was crystallised and examined by X-ray diffraction, which made it possible to determine the structure of the reverse transciptase complex with the primer, the DNA template and 2′-deoxynucleoside triphosphate.



R^1, R^2 are fragments of the oligonucleotide chain; $n = 2, 3$; the protein is HIV-1 reverse transcriptase.

A similar strategy was used for the study of conformational changes in reverse transcriptase during its binding to 2′-deoxynucleoside triphosphate.[172, 173]

He and Verdine[174] used the oligonucleotide **39** containing a cytidine residue modified at the amino group for the study of the N-terminal domain of the mutant Ada-protein from *E. coli*. Studies of specific covalent addition of modified DNA **39** to mutant proteins has made it possible to establish some DNA–protein contacts.

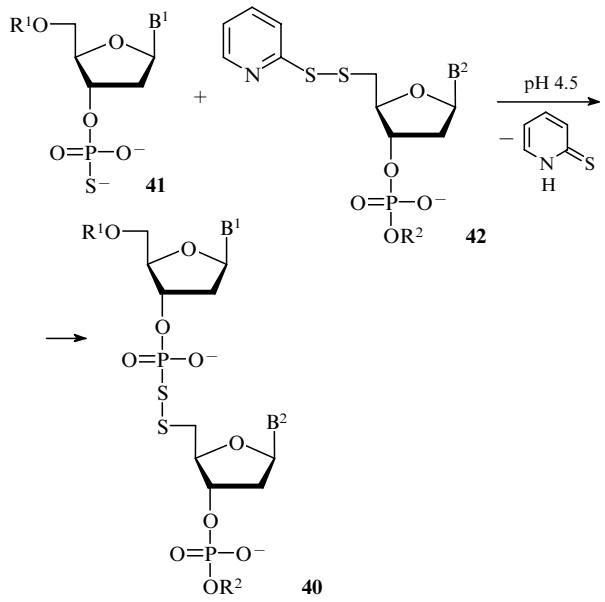


R^1, R^2 are fragments of the oligonucleotide chain.

e. Miscellaneous examples of modification of nucleic acids

It was suggested[175–178] that oligodeoxyribonucleotides containing phosphoryl disulfide and diphosphoryl disulfide groups in the sugar–phosphate backbone were used for selective binding of cysteine in proteins and peptides. Interaction of DNA methyltransferase SsoII with DNA containing a methylated site was studied using a series of DNA duplexes where one of the phosphodiester internucleotide bonds was replaced by the phosphoryl disulfide (PDS) group (one of the duplex strands represented a modified oligonucleotide **40**).[179, 180] The modified bond was introduced into the positions directly flanking the methylated cytidine residue using thiol–disulfide exchange between the 3′-thiophosphorylated oligonucleotide **41** and the 2-pyridyldithio-substituted oligonucleotide **42**.

‡ The cysteine residue was introduced instead of glutamine into the native enzyme using site-directed mutagenesis.

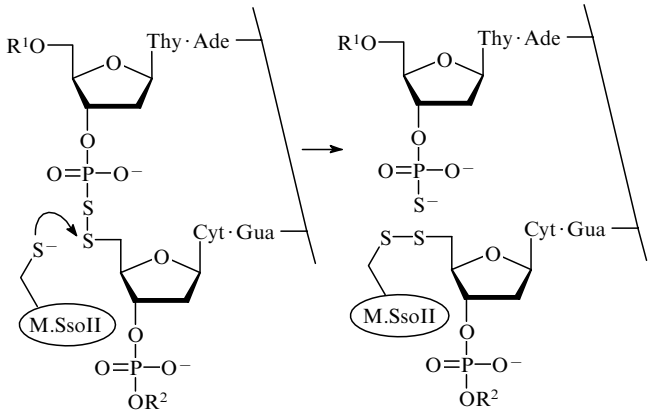


R^1, R^2 are fragments of the oligonucleotide chain; B^1, B^2 are heterocyclic bases.

This reaction proceeded with equal efficiency both in the presence and in the absence of the template oligonucleotide.[177, 178, 180] It was shown that PDS-containing DNA duplexes form conjugates with M.SsoII with different degrees of efficiency, which depended on the position of modification.[179]

It is known that the cleavage of the mixed disulfide R^1SSR^2 by a nucleophilic thiolate ion R^3S^- results in the formation of a more stable anion R^1S^- and a new mixed disulfide R^3SSR^2 (see Refs 165 and 181). With regard to PDS-containing DNA duplexes, this suggests that the thiolate ion attacks the sulfur bound to the carbon atom (but not the sulfur bound to phosphorus), since thiophosphoric acid is a much stronger acid than thiol. This was confirmed experimentally by an analysis (PAAG electrophoresis and mass spectrometry) of conjugates of PDS-containing DNA duplexes with M.SsoII (Scheme 7) and glutathione, a simple tripeptide model of cysteine-containing proteins.[179]

Scheme 7



R^1, R^2 are fragments of the oligonucleotide chain.

The internucleotide PDS bond within DNA is selective with respect to the cysteine residues and does not react with the ε-amino group of lysine.[179]

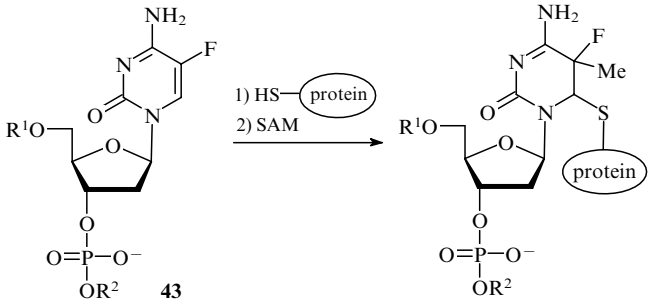
It is of note that the formation of a new disulfide bond between the cysteine residue and the PDS group is only possible where these are spatially close in the supramolecular DNA – protein complex. It has been shown [182] that the formation of the conjugate of the restriction endoculease SsoII (R.SsoII) and the DNA duplex containing a PDS group in the recognition site proceeds with an extremely low efficiency. This phenomenon was attributed to the remoteness of the Cys residue from the binding and catalytic sites of R.SsoII. Evidence for this hypothesis can be derived from the fact that the R.SsoII variants in which the Arg-(116, 117, 119) or Lys-122 residues interacting with the sugar – phosphate backbone of DNA were replaced by the Cys residue form conjugates with PDS-containing DNA substrates in high yields.[182]

III. Natural protein – nucleic acid conjugates formed in the enzymatic reactions

Reactions of certain enzymes with NA yield unstable intermediates which represent covalently bound biopolymer complexes. In some cases, dissociation of these intermediates can be prevented by modification of NA and they can be investigated.

The transfer of the methyl group onto 5-fluoro-2′-deoxyuridine monophosphate by thymidylate synthase is a classical example of an enzymatic reaction which results in the formation of a stable covalently bound protein – nucleic acid complex.[183] Substitution of the fluorine atom for the hydrogen atom in 2′-deoxyuridine monophosphate prevents the cleavage of the nucleotide – enzyme bond.[184] Cytosine-5-DNA methyltransferases and methyltransferases that react with RNA also belong to enzymes that form conjugates with oligonucleotides. Oligonucleotides containing a 5-fluorocytidine or 5-fluoro-2′-deoxycytidine residues (compound **43**) are widely used for the stabilisation of such conjugates.

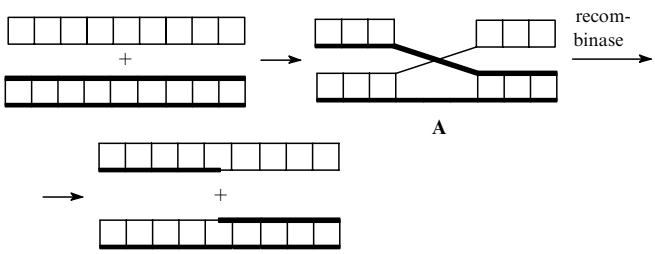
A more detailed description of these processes is given in the reviews;[10, 185] see also Ref. 186.



R^1, R^2 are fragments of the oligonucleotide chain, protein is cytosine-5-DNA methyltransferase, SAM is *S*-adenosyl-L-methionine.

Landy *et al.*[187–189] studied the covalent binding of DNA to recombinases catalysing the homologous recombination of DNA in the cross-like structures **A** (the Holiday structures) (Scheme 8) resulting in transfer of fragments of one region of a genome into the other region.

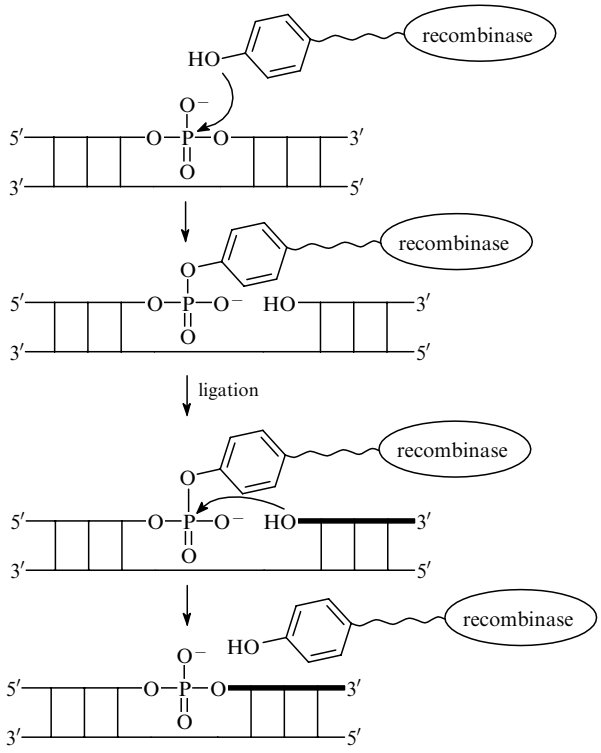
Scheme 8



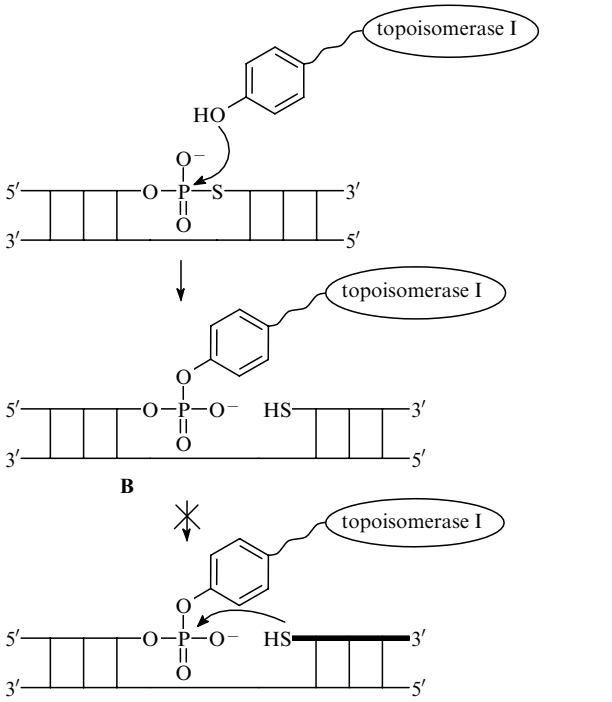
Hereinafter, ═══ are the strands of recombining fragments of DNA.

The mechanism of action of recombinases includes an attack of the hydroxy group of tyrosine on the internucleotide phosphate of DNA (Scheme 9) resulting in the cleavage of the phosphodiester bond and the formation of a conjugate between the 5′-terminal site of DNA and recombinase with subsequent ligation

Scheme 9



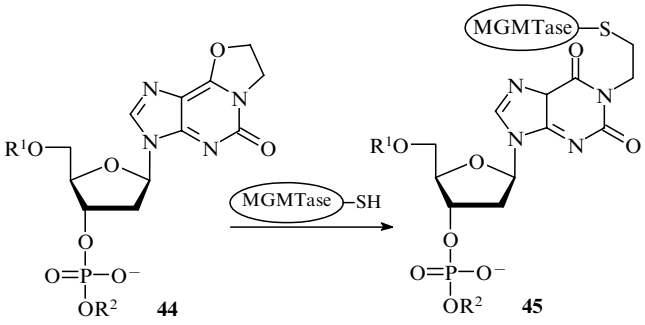
Scheme 10



of recombining DNA fragments and release of recombinase. It was suggested that shortened oligonucleotides were used so that cleavage in a single strand resulted in dissociation of the short 3′-terminal fragment of DNA. This prevented recombination in the ligation step and, as a consequence, stabilised the recombinase–DNA conjugate.

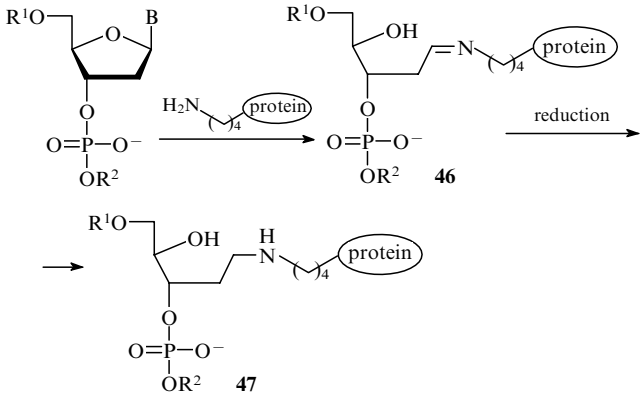
It was suggested [190, 191] that modified DNA containing phosphorothioate internucleotide bonds (Scheme 10) were used for the study of covalent binding of topoisomerases.§ No ligation took place after the formation of the protein–nucleic acid conjugate **B**, since the replacement of the P–O bond by the P–S bond is energetically unfavourable. This enabled selective inhibition of topoisomerase I and isolation of a conjugate in high yield.

The modified oligonucleotide **44** was used [192] for the study of the mechanism of action of MGMTase. The reaction of the oligonucleotide **44** with MGMTase containing a cysteine residue in the active site gave the conjugate **45**. Even in the presence of a 10-fold excess of the protein this reaction was rather slow (48 h).



R^1,R^2 are fragments of the oligonucleotide chain.

In a review [193] on bifunctional DNA glycosylases, which possess N-glycosylase and β-lyase activities, the formation of the Shiff base **46** upon reaction of the lysine side amino group of DNA glycosylase with an apurine site of the oligonucleotide was considered as a key reaction; subsequent reduction was aimed at the access to the stable conjugate **47**.

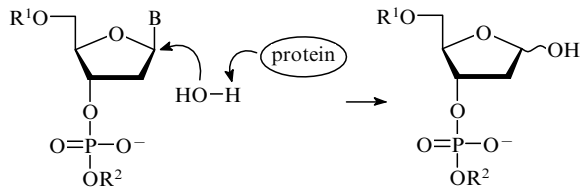


R^1,R^2 are fragments of the oligonucleotide chain; the protein is DNA glycosylase.

Strengthening of the bond between the enzyme and the substrate by the reduction of the Shiff base has also been used elsewhere.[194–197] Detailed studies of the mechanism of this reaction were carried out using model peptides.[198] Virtually all studies on glycosylases [193] made use of MutY from *E. coli* [194–196] and human endonuclease III.[197] Study of 8-oxoguanine-DNA glycosylase made it possible, for the first time, to establish the involvement of one of the reaction products, *viz*., 8-oxoguanine, in the cascade of reactions triggered by the β-lyase activity of this enzyme.[199]

It is of note that this approach cannot be applied to monofunctional DNA glycosylases, since they function according to a different mechanism which includes clevage of the N-glycosidic bond with the participation of the water molecule.

§ Topoisomerases initiate single-chain breaks in DNA.



R^1,R^2 are fragments of the oligonucleotide chain; the protein is DNA glycosylase

IV. Identification of protein sites covalently bound to nucleic acids

Amino acid residues in the proteins involved in the complex formation with NA can be identified by direct and indirect methods. The former entail isolation of the NA – protein conjugate, partial cleavage of the protein, isolation of an oligonucleotide – peptide fragment and elucidation of the primary structure of the peptide and identification of the amino acid involved in the formation of the covalent bond. The latter make use of mutant proteins and X-ray diffraction analysis. The amino acid residue of the protein, which is thought to react with DNA, is replaced by a non-reactive residue with subsequent analysis of whether the covalent bond is formed between the protein and the oligonucleotide.[162, 163] Obviously, this method is inapplicable to photoactivatable NA, which react with any amino acids. Data from X-ray diffraction on the arrangement of side groups of amino acid residues are used as evidence of proximity of definite amino acids of the protein to the specific fragment of NA.

Let us consider the direct method used for the analysis of protein – nucleic acid conjugates in more detail. The conjugates are usually isolated by polyacrylamide gel (PAAG) electrophoresis under denaturing conditions (Table 1).

The analysis of the data presented in Table 1 suggests that the conditions of isolation of protein – nucleic acid conjugates are analogous or are much the same as those used for the separation of proteins by PAAG electrophoresis according to Laemmli (12.5% PAAG, 250 mM Tris – HCl, 0.1% sodium dodecyl sulfate, pH 8.8).[202] The conjugate is eluted from the gel using a buffer in which subsequent partial cleavage of the protein is carried out. In some cases, the isolated product is desalted, dialysis being the most efficient and versatile procedure.

Medium-pressure ion-exchange chromatography (FPLC MonoQ HR5/5)[203] and anion-exchange HPLC (Porous HQ, PerSeptive Biosystems),[204] which permit separation of the DNA – protein conjugate from the nonbound protein, are also used for the isolation of conjugates.

The protein and the covalently bound biopolymer complex were precipitated with 10% trichloroacetic acid or antibodies, which made it possible to separate the bulk of the nonbound NA.[69, 74, 205] Dialysis of the reaction mixture can also be used for this purpose.[100] Nucleic acid – protein conjugates could be separated from excess protein if oligonucleotides carrying biotin at the 3′-end were used with subsequent immobilisation of the oligonucleotide (both free and within the conjugate) on avidin-coated magnetic particles.[74] Isolation of NA and the NA – protein conjugate by affinity chromatography was described.[76] To this end, an oligoadenylate sequence (oligo-dA) was attached to the 3′-end of a modified oligonucleotide followed by chromatography on a microcolumn with oligothymidylate-cellulose (oligo-dT). In this case, the separation of the complex from NA was achieved due to the formation of the oligo-dA · oligo-dT duplex.

It should be noted that the purification of the covalently bound protein – NA complex is not compulsory and in some cases[12, 79, 88, 206] the covalent binding is immediately followed by the next stage, *viz*., partial chemical or enzymatic cleavage of the protein.

The methods for the cleavage of the protein within the protein – nucleic acid complex are summarised in Table 2.

The data listed in Table 2 suggest that such proteases as trypsin and chymotrypsin are most commonly used for partial cleavage of the protein component of the conjugate. Enzymatic treatment is carried out in weakly alkaline solutions under

Table 1. Isolation of a NA-protein conjugate by gel electrophoresis.

Protein	Composition of gel [a]	Conditions of isolation from gel	Desalting conditions	Ref.
Protein Tax of leukemia virus	10% PAAG, 250 mM Tris – HCl, 0.1% SDS, pH 8.8	10 mM Tris – HCl, 1 mM EDTA, 0.1% SDS, pH 7.0, 4 °C, 18 h	no desalting	37
Replication terminator Tus	10% PAAG, 250 mM Tris – HCl, 0.1% SDS, pH 8.8	H_2O, 37 °C, 24 h (3-fold)	sedimentation with 0.3 M AcONa, EtOH	67
DNA methyltransferase EcoRI	12% PAAG, 250 mM Tris – HCl, 0.1% SDS, pH 8.8	–	dialysis against buffer 50 mM NH_4HCO_3, pH 7.8, 4 °C, 8 h	87
DNA methyltransferase EcoRII	12% PAAG, 90 mM Tris – HCl, 90 mM H_3BO_3, 2 mM EDTA, 7 M urea, pH 8.0	see [b]	no desalting	96
	12% PAAG, 100 mM Tris – HCl, 0.1% SDS, pH 8.8	1% SDS, 37 °C, 5 – 7 h	the same	94, 200
HIV-1 integrase	NuPage 4% – 12% MES gel (Novex, Inc.), pH 6.4	20 mM Tris – HCl, 0.5% SDS, pH 8.5, 37 °C, 18 h	precipitation with 10% CCl_3CO_2H, acetone	77
HIV-1Tat-protein	20% PAAG, 100 mM Tris – HCl, 0.1% SDS, pH 8.3	45 mM Tris – H_3BO_3, 1 mM EDTA, 3 M AcONa, pH 5.5	reversed-phase HPLC	98
Ttk-protein of *Drosophila melanogaster*	16% PAAG, 250 mM Tris – HCl, 0.1% SDS, 7 M urea	10 mM Tris – $HSCH_2CO_2H$, 0.1 mM EDTA, 0.1% Triton X-100, 0.1 mM PMSF,[c] pH 8.0, 45 °C, 4 h	no desalting	201

[a] Abbreviations: PAAG is polyacrylamide gel; Tris is tris(hydroxymethyl)aminomethane; SDS is sodium dodecyl sulfate; MES is morpholinoethanesulfonic acid. [b] The conjugate was not eluted from the gel. [c] PMSF is phenylmethylsulfonyl fluoride.

Table 2. Cleavage of the protein component within the NA – protein conjugate.

Reagent or enzyme (A)	Protein (B)	c_A or A : B (see [a])	Reaction conditions	Ref.
Chymotrypsin	homing endonuclease PI-SceI	40 μg ml^{-1}	50 mM Tris – HCl, 1 mM $CaCl_2$, pH 8.0, 37 °C, 2 h	79
	DNA methyltransferases CviBIII and TaqI	1 : 1	10 mM Tris – HCl, 20 mM $CaCl_2$, pH 8.0, 37 °C, 16 h	81
	restriction endonuclease SsoII	500 μg ml^{-1}	100 mM NH_4HCO_3, 37 °C, 2 h	88
Chymotrypsin (see [b])	protein U1 of the small nuclear ribonucleoprotein	1 : 20	50 mM Tris – HCl, 1 M urea, 1 mM DTT, 1 mM $CaCl_2$, pH 7.5, 37 °C, 16 h	205
BrCN	DNA methyltransferase EcoRII	1 M	HCOOH, 25 °C, 48 h	94, 200
	HIV-1 integrase	100 mM	70% CF_3CO_2H, 70 °C, 45 min	77
Protease Glu-C	the same	1 : 5	50 mM Tris – HCl, 1 M NaCl, 5 mM EDTA, 5 mM DTT, 0.1% SDS, pH 8.5, 37 °C, 12 h	77
	"	1 : 5	50 mM Tris – HCl, 1 M NaCl, 5 mM EDTA, 5 mM DTT, 0.01% SDS, pH 8.0, 37 °C, 12 h	74
BNPS-scatol (see [c])	DNA methyltransferase EcoRII	70 mg ml^{-1}	1% SDS, 70% AcOH, 37 °C, 2 h	94
Trypsin (see [d])	the same	1 : 25	400 mM NH_4HCO_3, 2 M urea, 0.1 mM $CaCl_2$, 37 °C, 24 h	12
Trypsin (see [e])	"	1 : 1 to 1 : 10	50 mM Tris – HCl, 1 mM $CaCl_2$, 10% MeCN, pH 8.5, 37 °C, 16 h	96
Trypsin	Ttk protein of *Drosophila melanogaster*	5 μg ml^{-1}	100 mM NH_4HCO_3, 0.2% Triton X-100, 1 mM DTT, 37 °C, 6 h	201
	DNA methyltransferase EcoRI	1 : 20	50 mM NH_4HCO_3, pH 7.8, 37 °C, 12 h	87
	HIV-1 Tat-protein	0.05 μg ml^{-1}	100 mM Tris – HCl, 10 mM $CaCl_2$, pH 7.8, 25 °C, 4 h	98
	the same		100 mM Tris – HCl, pH 8.5, 37 °C, 24 h	67
	"		100 mM NH_4HCO_3, 2 M urea, pH 7.8, 37 °C, 24 h	100
	DNA polymerase β	1 : 10	50 mM NH_4HCO_3, 10% MeCN, pH 8, 37 °C, 3 h	203
	DNA endonuclease VIII	1 : 6	20 mM Tris – HCl, 3 .4 M urea, pH 7.5, 37 °C, 18 h	204
Trypsin (see [b])	HIV-1 Tat-protein	1 : 20	50 mM Tris – HCl, 1 M urea, 1 mM DTT, 1 mM $CaCl_2$, pH 7.5, 37 °C, 16 h	205
	the same	5 μg ml^{-1}	2.5 mM Tris – HCl, 5 mM $CaCl_2$, 0.25 mM EDTA, 0.025% SDS, pH 8.0, 37 °C, 2 h	37
	"	160 μg ml^{-1}	50 mM Tris – HCl, 2 mM $CaCl_2$, pH 8.0, 37 °C, 18 h	76
Clostripaine	repair protein MutS	80 μg ml^{-1}	30 mM Tris – HCl, 2 mM $CaCl_2$, 2 mM DTT, 0.5% SDS, pH 7.5, 37 °C, 18 h	76
Proteinase K	gene conversion endonuclease PI-SceI	2 mg ml^{-1}	50 mM Tris – HCl, 1 mM $CaCl_2$, pH 8.0, 37 °C, 16 h	79

[a] The concentration of the reagent (c_A) or the enzyme : substrate ratio (A : B). [b] Preliminary denaturation of the protein – nucleic acid conjugate. [c] Functionalisation with 3-bromo-3-methyl-2-(2-nitrophenylsulfenyl)-3*H*-indole. [d] Preliminary denaturation of the protein – nucleic acid conjugate and treatment with iodoacetic acid. [e] Preliminary treatment of the protein – nucleic acid conjugate with iodoacetamide.

denaturing conditions. In some cases, preliminary denaturation of the protein, reduction of the disulfide bonds and alkylation of the thiol groups of cysteine with iodoacetic acid or iodoacetamide are performed in order to reduce the proteolysis time and to obtain more homogeneous protein – nucleic acid conjugates. Proteinase K is often used for the treatment of analytical amounts of protein – nucleic acid conjugates; however, this enzyme is not used in a large scale due to its low specificity, which results in the formation of a mixture of peptide – nucleic acid conjugates and, as a consequence, prevents the identification of the covalently bound site of the protein. Nevertheless, the use of proteinase K permits a comparison of stabilities of DNA – protein conjugates and of the original protein, which facilitates the selection of conditions for the trypsinolysis of a conjugate in subsequent studies.

Partial cleavage of the protein is followed by purification of the peptide – oligonucleotide conjugate. This is achieved by reversed-phase HPLC, gel filtration, anion-exchange HPLC and PAAG electrophoresis (Table 3). Elucidation of the primary structure according to Edman or using mass spectrometry demands preliminary isolation and desalting of the conjugates. The most efficient chromatographic procedures make use of buffer solutions based on volatile salts, *e.g.*, ammonium formate or triethylammonium acetate. In this case, multiple evaporation with water or aqueous ethanol is sufficient to desalt the conjugate isolated by HPLC. However, in some cases protein admixtures cannot be removed by reversed-phase HPLC and gel filtration, and a two-step purification procedure of a conjugate seems to be more reliable, first, by anion-exchange HPLC or electrophoresis in PAAG with subsequent desalting by microcolumn gel filtration or reversed-phase chromatography or ultracentrifugation. The transfer of a conjugate to poly(difluorovinylidene) (PVDF) membrane (blotting) with subsequent determination of the primary

Table 3. Isolation of peptide – oligonucleotide conjugates.

Isolation procedure	Separation conditions	Detection procedure [a]	Desalting procedure	Ref.
Anion-exchange HPLC				
Vydac 301VHP552	25 mM $Et_3NH^+AcO^-$, pH 8.0; gradient (0.5 M NaCl): 0% to 60% (40 min)	SP (215 + 260)	reversed-phase chromatography on a cartridge C_{18} (Sep-Pak Plus, Waters), precipitation with ethanol	87
Mono Q (Amersham-Pharmacia-Biotech)	50 mM Tris – HCl, 1 mM EDTA, pH 7.5; gradient (1 M NaCl): 0% to 80% (40 min)	RA	precipitation with ethanol or gel filtration on a cartridge NAP-10, 10 mM NH_4HCO_3	79, 88
Sychrom AX 300	20 mM NaH_2PO_4, pH 6.8, 20% MeCN; gradient (NaCl): 0 (30 min) to 1 M (60 min), 1 M (20 min)	SP (254)	–	100
Poros HQ10 (PerSeptive Biosystems)	10 mM Tris – HCl, pH 7.4; gradient (1 M KCl): 0% (5 min), 0% to 100% (15 min)	SP (260)	ultrafiltration (Microsep 3K, Pall-Gelmann Sciences)	81
Gel filtration				
Superose 12 (Amersham-Pharmacia-Biotech)	0.1 M NH_4HCO_3, 2 M urea, 1 M NaCl, pH 8.0	–	gel filtration on Sephadex G25 column, 10 mM HEPES,[b] 10 mM NaCl	12
	20 mM Tris – HCl, 50 mM NaCl, 5 mM EDTA, 8 M urea, pH 7.8	SP (254)	precipitation with ethanol	205
Reversed-phase HPLC				
C_{18}, Vydac	0.1% CF_3COOH, 5% MeCN; gradient (MeCN): 5% (30 min), 5% to 45% (120 min), 45% to 90% (20 min)	SP (220, 260)	no desalting	205
	10 mM $Et_3NH^+AcO^-$; gradient (MeCN): 0% to 30%	SP (254)	the same	100
C_{18}, Phenomenex	0.1% HCOOH; gradient (MeCN): 0% to 40%, 20 min	MS	–	70
μK3C C_2/C_{18} SC 2.1/10 (Amersham-Pharmacia-Biotech)	5 mM K_2CO_3, pH 7.0; gradient (MeCN): 0% to 70%, 50 min	SP(214, 254, 280)	–	206
Electrophoresis				
10% PAAG	–	RA	no desalting	76
12% PAAG	45 mM Tris – H_3BO_3, 1 mM EDTA, 6 M urea, 1 mM $HSCH_2CO_2H$, pH 7.0	–	gel filtration on a column with Sephadex G25, 10 mM NH_4HCO_3	79, 88
	50 mM Tris – H_3BO_3, 1 mM EDTA, 8 M urea	RA	elution from the gel with a buffer 20 mM NH_4HCO_3, 10% MeCN, pH 8.5, evaporation	37
20% PAAG	50 mM Tris – H_3BO_3, 1 mM EDTA, 8 M urea	SP	elution from the gel with a buffer 10 mM Tris – HCl, 1 mM EDTA, pH 7.5	204
	50 mM Tris – H_3BO_3, 1 mM EDTA, 8 M urea	RA	elution from the gel with a buffer 45 mM Tris – H_3BO_3, 1 mM EDTA, 3 M AcONa, pH 5.5; reversed-phase chromatography on a cartridge C_{18}	98
NuPage 4% – 12% MES gel (Novex, Inc.)	–	RA	transfer to PVDF membrane (Trans-Blot, Bio-Rad Laboratories)	77
15% Tris-Tricin (Novex, Inc.)	–	RA	transfer to PVDF membrane	74

[a] Abbreviations: SP, spectrophotometric detection (the wavelengths, nm, are indicated in parenthesis); RA, autoradiographic detection; MS, mass-spectrometric detection. [b] HEPES is 2-[4-(2-hydroxyethyl)-1-piperazinyl]ethanesulfonic acid.

structure of the peptide in the immobilised NA – peptide complex has been described. The transfer of conjugates onto anion-exchange filters is also used.[207]

The final step includes determination of the primary structure of the covalently bound peptide using chemical or enzymatic sequencing. In some studies, [37, 67, 74, 76, 79, 81, 87, 88, 96, 100, 201, 205] the primary structures of the proteins were determined by sequencing which included chemical degradation according to Edman.[208] This procedure is usually performed on automatic sequencers, which demands significant amounts of the conjugate (50 – 150 pmol). The substituted amino acid is determined by the absence of a standard *N*-phenylthiocarbamoyl derivative (the so-called 'blank' sequencing cycle). In the case of incomplete purification of the conjugate, *i.e.*, when a mixture of oligonucleotide –

peptide conjugates differing in one or two amino acids is present in the solution, the use of this approach is significantly complicated.

An interesting approach to determination of the primary structure of a cross-linked peptide was proposed by Mucke *et al.*[96] These authors carried out a comparative analysis of trypsinolysis products of free restriction endonuclease EcoRII and its conjugate with an oligonucleotide using reversed-phase HPLC. A short protein–nucleic acid conjugate was not absorbed on the column and the 'extra' fragment present in the tryptic hydrolysate of the free protein could unambiguously be determined. Sequencing of this peptide allowed identification of the amino acid that reacted with the 5-iodopyrimidine residue of NA. Unfortunately, this method is not versatile, since the presence of several reactive residues in the peptide under study casts doubts on the structure of the conjugate.

Sequencing of peptides can also be performed using mass spectrometry.[209–215] A comparison of several approaches to the determination of substituted amino acids using mass spectrometry has been carried out.[92] Currently, determination of molecular masses and sequencing of biopolymers are usually performed using matrix-assisted laser desorption ionisation (MALDI)[209] and electrospray ionisation (ESI) mass spectrometry.[210] In MALDI MS, the biomolecule is separated from the neighbouring molecules using matrices, *i.e.*, various low-molecular-mass aromatic compounds excitable under the action of laser irradiation. Irradiation of samples is accompanied by ionisation of a biomolecule with the predominant formation of a single-charged ion. In ESI MS, the solution containing the biopolymer is fed into the ionisation chamber of the mass spectrometer through a capillary. Application of a strong electrostatic field results in the formation of ionised aerosol at the capillary tip. The ionisation of the biopolymer occurs after evaporation of the solvent.

In the analysis of proteins, it is positively charged ions that are registered, while negatively charged ions are recorded in the analysis of oligonucleotides, which is the main problem in the use of MS for the characterisation of protein–nucleic acid conjugates. Nevertheless, some authors believe[216–218] that MALDI MS and ESI MS can be used for detecting and sequencing of these conjugates. The choice of an effective matrix in the sequencing of protein–nucleic acid conjugates by MALDI MS is a central task, since no versatile matrices for the detection of both peptides and oligonucleotides exist. The mass-spectrometric analysis of protein–nucleic acid conjugates by ESI MS is facilitated by removal of the nucleic acid component using enzymatic degradation of NA.[70, 83, 87, 219, 220] The advantage of the ESI MS is the possibility of its application for the analysis of fractions after high-performance liquid chromatography or electrophoresis.

In the study,[92] the structure of a DNA–protein conjugate was determined by mass spectrometry. After partial trypsinolysis of the isolated conjugate, the reaction mixtures were analysed by MALDI TOF (time-of-flight) MS and the data obtained were compared with the analogous data for the free protein. This enabled determination of the peptide fragment covalently bound to DNA. The use of proteinase K instead of trypsin has made it possible to diminish the peptide fragment to four amino acid residues. However, the amino acid covalently bound to the oligonucleotide could only be determined after complete degradation of DNA by phosphodiesterase and alkaline phosphatase and subsequent sequencing of the resulting DNA–protein conjugate by tandem nanoESI MS.¶

Tandem MS entails the presence of two consecutively arranged detectors with an ionisation chamber in-between. The first detector permits the passage of only ions with a definite mass/charge ratio into the ionisation cell. These undergo fragmentation in the ionisation chamber. The ionisation products (the so-called daughter ions) pass into the second detector for determination of the primary structure of the peptide. As a rule, this method allows determination of amino acid sequences in proteins containing up to 30 amino acids.

Tandem ESI MS is the most efficient method for the determination of the structures of peptide–oligonucleotide conjugates and, hence, of the target amino acid. The use of tandem nanoESI MS[92] has made it possible to determine the structure of a protein at its content in the solution of about 10^{-10} moles. However, low yields of the conjugates and inevitable losses in numerous intermediate steps require femtomolar amounts of the protein. The sensitivity of mass-spectrometric methods was found to exceed significantly those of traditional sequencing methods based on the Edman procedure, since they permit direct determination of the attached amino acid from its molecular mass (tandem MS) and direct sequencing of both the oligonucleotide and the peptide and do not require thorough purification of the samples.

The use of mass-spectrometric techniques for the determination of the structures of the attached peptides has also been considered elsewhere.[206, 207, 219, 221, 222] However, despite the obvious merits of mass spectrometry in the analysis and determination of structures of peptide–oligonucleotide conjugates, this method has not found wide application due to low accessibility of tandem mass spectrometres.†

Hybrid instruments, *e.g.*, quadrupole time-of-flight (Q-TOF) and Fourier-transform ion-cyclotron resonance mass spectrometres,[92] (FTICR) and quadrupole ion traps[92] (QIT), are used along with traditional time-of-flight mass spectrometres with a linear detector designed for the study of biomolecules. The use of FTICR MS ensures a much higher accuracy in the determination of molecular masses (> 10 kDa) in comparison with TOF MS; however, its high cost and complexity of calibration for the analysis of peptide–oligonucleotide conjugates make them unclaimed, at least at present. Mass spectrometres supplied with quadrupole ion traps do not ensure the resolution comparable to that of TOF MS. A more detailed description of various mass-spectrometric techniques used in the study of structures of protein–nucleic acid complexes and conjugates can be found in the reviews.[224–226]

* * *

This review has been written with the financial support of the Russian Foundation for Basic Research (Project Nos 03-04-48957, 04-04-48714, 02-03-32950) and the programme Universities of Russia (Grant 07.03.066).

References

1. G L Verdine, D P G Norman *Annu. Rev. Biochem.* **72** 337 (2003)
2. M Deibert, S Grazulis, G Sasnauskas, V Siksnys, R Huber *Nature Struct. Biol.* **7** 792 (2000)
3. G Ya Sheflyan, E A Kubareva, E S Gromova *Usp. Khim.* **65** 765 (1996) [*Russ. Chem. Rev.* **65** 709 (1996)]
4. D G Knorre, N V Kudryashova, O I Lavrik *Usp. Khim.* **67** 486 (1998) [*Russ. Chem. Rev.* **67** 435 (1998)]
5. D G Knorre, T S Godovikova *FEBS Lett.* **433** 9 (1998)
6. D G Knorre, T S Godovikova *Izv. Akad. Nauk, Ser. Khim.* 1225 (1999) [a]
7. D G Knorre, N D Kobets *Usp. Khim.* **68** 365 (1999) [*Russ. Chem. Rev.* **68** 333 (1999)]
8. S R Kassabov, N M Henry, M Zofall, T Tsukiyama, B Bartholomew *Mol. Cell. Biol.* **22** 7524 (2002)
9. M C Golden, B D Collins, M C Willis, T H Koch *J. Biotechnol.* **81** 167 (2000)

¶ The prefix 'nano' designates that the solution in the capillary is fed at the rate of nl s^{-1} (this enables to reduce the amount of the protein and to increase the sensitivity of the detector).

† In the past decade, significant investments were made to investigations in the field of proteomics,[223] which made it possible to increase significantly the number of tandem mass spectrometres.

10. D Smith, B D Collins, J Heil, T H Koch *Mol. Cell. Proteomics* **2** 11 (2003)
11. K C Smith *Photochem. Photobiol.* **3** 415 (1964)
12. W M Rehrauer, S C Kowalczykowski *J. Biol. Chem.* **271** 11 996 (1996)
13. L A Sylvers, J Wower *Bioconjugate Chem.* **4** 411 (1993)
14. K M Meisenheimer, T H Koch *Crit. Rev. Biochem. Mol. Biol.* **32** 101 (1997)
15. N Naryshkin, Y Kim, Q Dong, R H Ebright *Methods Mol. Biol.* **148** 337 (2001)
16. F Robert, B Coulombe *Methods Mol. Biol.* **148** 383 (2001)
17. J Persinger, B Bartholomew *Methods Mol. Biol.* **148** 363 (2001)
18. F Gohring, F O Fackelmayer *Biochemistry* **36** 8276 (1997)
19. M-H Kuo, C D Allis *Methods* **19** 425 (1999)
20. V Orlando *Trends Biochem. Sci.* **25** 99 (2000)
21. I M Gavin, S M Melnik, N P Yurina, M I Khabarova, S G Bavykin *Anal. Biochem.* **263** 26 (1998)
22. R P Hikerson, C L Chepanoske, S D Williams, S S David *J. Am. Chem. Soc.* **121** 9901 (1999)
23. K L Nguyen, M Steryo, K Kurbanyan, K M Nowitzki, S M Butterfield, S R Ward, E D A Stemp *J. Am. Chem. Soc.* **122** 3585 (2000)
24. E M Boon, M A Pope, S D Williams, S S David, J K Barton *Biochemistry* **41** 8464 (2002)
25. D Bastia *Structure* **4** 661 (1996)
26. Y-B Zhou, S E Gerchman, V Ramakrishnan, A Travers, S Muyldermans *Nature (London)* **395** 402 (1998)
27. A F Ghetu, D C Arthur, T K Keppola, J N M Glover *RNA* **8** 816 (2002)
28. I E Kanevskii, S A Kuznetsova *Usp. Khim.* **67** 688 (1998) [*Russ. Chem. Rev.* **67** 617 (1998)]
29. L M Hales, R I Gumport, J F Gardner *Nucleic Acids Res.* **24** 1780 (1996)
30. T S Heuer, P O Brown *Biochemistry* **37** 6667 (1998)
31. A M Kimzey, W S Dynan *J. Biol. Chem.* **273** 13 768 (1998)
32. D L Kolpashchnikov, S N Khodyreva, D Yu Klimankov, M S Wold, A Favre, O I Lavrik *Nucleic Acids Res.* **29** 373 (2001)
33. I A Udalova, J C Knight, V Vidal, S A Nedospasov, D Kwiatkowski *J. Biol. Chem.* **273** 21 178 (1998)
34. O I Lavrik, D M Kolpashchnikov, H P Nasheuer, K Wiesshart, A Favre *FEBS Lett.* **441** 186 (1998)
35. I Wong, T M Lohman *Proc. Natl. Acad. Sci. U.S.A.* **93** 10 051 (1996)
36. M S Bartlett, M Thomm, E P Geiduschek *Nature Struct. Biol.* **7** 782 (2000)
37. A L Kimzey, W S Dynan *J. Biol. Chem.* **274** 34 226 (1999)
38. J Persinger, B Bartholomew *J. Biol. Chem.* **271** 33 039 (1996)
39. T Lagrange, T K Kim, G Orphanides, Y W Ebright, R H Ebright, D Reinberg *Proc. Natl. Acad. Sci. U.S.A.* **93** 10 620 (1996)
40. J J Tate, J Persinger, B Bartholomew *Nucleic Acids Res.* **26** 1421 (1998)
41. S M Sengupta, J Persinger, B Bartholomew, C L Peterson *Methods* **19** 434 (1999)
42. C Costas, E Yuriev, K L Meyer, T S Guion, M M Hanna *Nucleic Acids Res.* **28** 1849 (2000)
43. M M Hanna, K Liu *J. Mol. Biol.* **282** 227 (1998)
44. M M Hanna *Methods Enzymol.* **274** 403 (1996)
45. W A Wlassoff, M I Dobrikov, I V Safronov, R Y Dudko, V S Bogachev, V V Kandaurova, G V Shishkin, G M Dymshits, O I Lavrik *Bioconjugate Chem.* **6** 352 (1995)
46. I V Safronov, N V Shcherbik, S N Khodyreva, W A Wlassov, M I Dobrikov, G V Shishkin, O I Lavrik *Bioorg. Khim.* **23** 576 (1997) [b]
47. D M Kolpashchikov, A L Zakharenko, S V Dezhurov, N I Rechkunova, S N Khodyreva, S Kh Degtyarev, V V Litvak, O I Lavrik *Bioorg. Khim.* **25** 129 (1999) [b]
48. D M Kolpashchikov, L A Aleksandrova, N F Zakirova, S N Khodyreva, O I Lavrik *Bioorg. Khim.* **26** 151 (2000) [b]
49. D M Kolpashchikov, T M Ivanova, V S Boghachev, H P Nasheuer, K Weisshart, A Favre, P E Pestryakov, O I Lavrik *Bioconjugate Chem.* **11** 445 (2000)
50. O I Lavrik, D M Kolpashchikov, K Weisshart, H P Nasheuer, S N Khodyreva, A Favre *Nucleic Acids Res.* **27** 4235 (1999)
51. D M Kolpashchikov, P E Pestryakov, W A Wlassov, S N Khodyreva, O I Lavrik *Biokhimiya* **65** 160 (2000) [c]
52. D M Kolpashchikov, P Hughes, A Favre, G Baldacci, O I Lavrik *J. Mol. Recognit.* **14** 239 (2001)
53. I A Drachkova, I O Petruseva, I V Safronov, A I Zakharenko, G V Shishkin, O I Lavrik, S N Khodyreva *Bioorg. Khim.* **27** 197 (2001) [b]
54. A L Zakharenko, D M Kolpashchikov, S N Khodyreva, O I Lavrik, L Menendez-Arias *Biokhimiya* **66** 999 (2001) [c]
55. D Yu Khlimankov, N I Rechkunova, S N Khodyreva, I O Petruseva, Zh K Nazarkina, E A Belousova, O I Lavrik *Mol. Biol.* **36** 1044 (2002) [d]
56. S Doerhoefer, S Khodyreva, I V Safronov, W A Wlassov, R Anarbaev, O I Lavrik, E Holler *Microbiology* **144** 3181 (1998)
57. D Yu Khlimankov, N I Rechkunova, D M Kolpashchikov, S N Khodyreva, E A Belousova, O I Lavrik *Biokhimiya* **66** 733 (2001) [c]
58. O I Lavrik, R Prasad, R W Sobol, J K Horton, E J Ackerman, S H Wilson *J. Biol. Chem.* **276** 25541 (2001)
59. W A Wlassov, G M Dymshyts, O I Lavrik *Mol. Biol.* **32** 5 (1998) [d]
60. D M Kolpachnikov, N I Rechkunova, M I Dobrikov, S N Khodyreva, N A Lebedeva, O I Lavrik *FEBS Lett.* **448** 141 (1999)
61. N I Rechkunova, D M Kolpashchikov, N A Lebedeva, I O Petruseva, M I Dobrikov, S Kh Degtyarev, O I Lavrik *Biokhimiya* **65** 244 (2000) [c]
62. N A Lebedeva, D M Kolpashchikov, N I Rechkunova, S N Khodyreva, O I Lavrik *Biochem. Biophys. Res. Commun.* **287** 530 (2001)
63. N A Lebedeva, D M Kolpashchikov, N I Rechkunova, S N Khodyreva, O I Lavrik *Biokhimiya* **67** 807 (2002) [c]
64. O I Lavrik, D M Kolpashchikov, R Prasad, R W Sobol, S H Wilson *Nucleic Acids Res.* **30** U73 (2002)
65. N A Lebedeva, N I Rechkunova, S N Khodyreva, A Favre, O I Lavrik *Biochem. Biophys. Res. Commun.* **297** 714 (2002)
66. I Ahmad, D N Rao *J. Mol. Biol.* **259** 229 (1996)
67. L J Duggan, P T Asmann, T M Hill, P A Gottlieb *Biochemistry* **35** 15391 (1996)
68. C W Espelin, K B Kaplan, P K Sorger *J. Cell Biol.* **139** 1383 (1997)
69. S G Liang, S N Maity *J. Biol. Chem.* **273** 31590 (1998)
70. M C Golden, K A Resing, B B Collins, M C Willis, T H Koch *Protein Sci.* **8** 2806 (1999)
71. Q Wang, R Shoeman, P Traub *Biochemistry* **39** 6645 (2000)
72. G Mass, T Nethanel, O I Lavrik, M S Wold, G Kaufmann *Nucleic Acids Res.* **29** 3892 (2001)
73. K M Meisenheimer, P L Meisenheimer, M C Willis, T H Koch *Nucleic Acids Res.* **24** 981 (1996)
74. T M Jenkins, D Esposito, A Engelman, R Craigie *EMBO J.* **16** 6849 (1997)
75. P W Hammond, T N Lively, T R Cech *Mol. Cell. Biol.* **17** 296 (1997)
76. V A Malkov, I Biswas, R D Camerini-Otero, P Hsieh *J. Biol. Chem.* **272** 23811 (1997)
77. D Esposito, R Craigie *EMBO J.* **17** 5832 (1998)
78. D L Wong, J G Pavlovich, N O Reich *Nucleic Acids Res.* **26** 645 (1998)
79. V Pingoud, H Thole, F Christ, W Grindl, W Wende, A Pingoud *J. Biol. Chem.* **274** 10235 (1999)
80. F Christ, S Schoettler, W Wende, S Steuer, A Pingoud, V Pingoud *EMBO J.* **18** 6908 (1999)
81. B Holz, N Dank, J E Eickhoff, G Lipps, G Krauss, E Wienhold *J. Biol. Chem.* **274** 15 066 (1999)
82. U Schweizer, T Hey, G Lipps, G Krauss *Nucleic Acids Res.* **27** 3183 (1999)
83. N Biswas, S K Weller *J. Biol. Chem.* **274** 8068 (1999)
84. A Jeltsch, M Roth, T Friederich *J. Mol. Biol.* **285** 1121 (1999)
85. Q M Eastman, I J Villey, D G Shatz *Mol. Cell. Biol.* **19** 3788 (1999)
86. S Yoo, A Kimzey, W S Dynan *J. Biol. Chem.* **274** 20 034 (1999)
87. D L Wong, N O Reich *Biochemistry* **39** 15410 (2000)
88. E A Kubareva, H Thole, A S Karyagina, T S Oretskaya, A Pingoud, V Pingoud *Nucleic Acids Res.* **28** 1085 (2000)

89. E L Lin, S W Lin, A Lin *Nucleic Acids Res.* **29** 2510 (2001)
90. N Biswas, S K Weller *J. Biol. Chem.* **276** 17 610 (2001)
91. Y Mikata, Q He, S J Lippard *Biochemistry* **40** 7533 (2001)
92. H Steen, J Petersen, M Mann, O E Jensen *Protein Sci.* **10** 1989 (2001)
93. O V Babkina, C A Chutko, A A Shashkov, M S Dzhidzoev, R Eritja, E S Gromova *Photochem. Photobiol. Sci.* **1** 636 (2002)
94. E V Koudan, O M Subach, G A Korshunova, E A Romanova, R Eritja, E S Gromova *J. Biomol. Struct., Dynamics* **20** 421 (2002)
95. C Beck, A Jeltsch *Biochemistry* **41** 14 103 (2002)
96. M Mucke, V Pingoud, G Grelle, R Kraft, D H Kruger, M Reuter *J. Biol. Chem.* **277** 14 288 (2002)
97. A S Bernards, J K Miller, K K Bao, I Wong *J. Biol. Chem.* **277** 20960 (2002)
98. Z Wang, T M Rana *Biochemistry* **35** 6491 (1996)
99. A McGregor, M V Rao, G Duckworth, P G Stockley, B A Connolly *Nucleic Acids Res.* **24** 3173 (1996)
100. M J Kovach, R Tirumalai, A Landy *J. Biol. Chem.* **277** 14530 (2002)
101. M A Cahill, A Nordheim, Y-Z Xu *Biochem. Biophys. Res. Commun.* **229** 170 (1996)
102. D R Mernagh, I A Taylor, G G Kneale *Biochem. J.* **336** 719 (1998)
103. Z Y Wang, T M Rana *Biochemistry* **37** 4235 (1998)
104. S Sastry, B M Ross *Proc. Natl. Acad. Sci. U.S.A.* **95** 9111 (1998)
105. G D Prestwich, G Dorman, J T Elliott, D M Marecak, A Chaudhary *Photochem. Photobiol.* **65** 222 (1997)
106. I M Schmitt, S Chimenti, F P Gasparro *J. Photochem. Photobiol. B* **27** 101 (1995)
107. R E Wellinger, R Lucchini, R Dammann, J M Sogo *Methods Mol. Biol.* **119** 161 (1999)
108. R P van Gijlswijk, E G Talman, P J Janssen, S S Snoeijers, J Killian, H J Tanke, R J Heetebrij *Expert. Rev. Mol. Diagn.* **1** 81 (2001)
109. S I Antsypovich, T S Oretskaya *Usp. Khim.* **67** 274 (1998) [*Russ. Chem. Rev.* **67** 245 (1998)]
110. D Bethea, B Fullmer, S Syed, G Seltzer, J Tiano, C Rischko, L Gillespie, D Brown, F P Gasparro *J. Dermatol. Sci.* **19** 78 (1999)
111. J Zehnder, R van Atta, C Jones, H Sussman, M Wood *Clin. Chem.* **43** 1703 (1997)
112. S S Sastry, H P Spielmann, Q S Hoang, A M Phillips, A Sancar, J E Hearst *Biochemistry* **32** 5526 (1993)
113. S S Sastry *Biochemistry* **35** 13519 (1996)
114. S S Sastry, B M Ross, A Parraga *J. Biol. Chem.* **272** 3715 (1997)
115. J T Reardon, A Sancar *Mol. Cell. Biol.* **22** 5938 (2002)
116. C Wylenzek, M Engelmann, D Holten, R van Atta, M Wood, B Gathof *Clin. Chem.* **46** 1853 (2000)
117. G A Korshunova, N V Sumbatyan, A N Topin, M T Mchedlidze *Mol. Biol.* **34** 966 (2000) [d]
118. Yu Yu Agapkina, D V Agapkin, A V Zagorodnikov, Ya I Alekseev, G A Korshunova, M B Gottikh *Bioorg. Khim.* **28** 324 (2002) [b]
119. M Taranenko, A Rykhlevskaya, M Mtchedlidze, J Laval, S Kuznetsova *Eur. J. Biochem.* **270** 2945 (2003)
120. S A Kuznetsova, M G Ivanovskaya, Z A Shabarova *Bioorg. Khim.* **16** 219 (1990) [b]
121. Z A Shabarova *Nucleic Acids Res. Symp. Ser.* **24** 99 (1991)
122. A A Purmal, Z A Shabarova, R I Gumport *Nucleic Acids Res.* **20** 3713 (1992)
123. G Y Sheflyan, E A Kubareva, E M Volkov, T S Oretskaya, Z A Shabarova *Gene* **157** 187 (1995)
124. M G Ivanovskaya, I A Kozlov, E A Kubareva, E A Taran, T V Serazev, Z A Shabarova *Mol. Biol.* **31** 436 (1997) [d]
125. E A Rudakova, M G Ivanovskaya, M V Kozlov, M V Khoretonenko, T S Oretskaya, V G Nikiforov *Biokhimiya* **65** 753 (2000) [c]
126. Z A Shabarova *Nucleosides Nucleotides* **17** 2063 (1998)
127. I A Kozlov, E A Kubareva, M G Ivanovskaya, Z A Shabarova *Antisense Nucleic Acids Drug. Dev.* **7** 279 (1997)
128. I A Kozlov, E A Kubareva, M G Ivanovskaya, Z A Shabarova *Mol. Biol.* **31** 63 (1997) [d]
129. M Zanetti, P Castiglioni, S Schoenberger, M Gerloni *Ann. N. Y. Acad. Sci.* **987** 249 (2003)
130. J Kucharczak, M J Simmons, Y J Fan, C Gelinas *Oncogene* **22** 8961 (2003)
131. H C Liou *J. Biochem. Mol. Biol.* **35** 537 (2002)
132. S A Kuznetsova, C Clusel, E Ugarte, I Elias, M Vasseur, M Blumenfeld, Z A Shabarova *Nucleic Acids Res.* **24** 4783 (1996)
133. S Kuznetsova, A Rykhlevskaya, M Taranenko, O Sidorkina, T Oretskaya, J Laval *Biochimie* **85** 511 (2003)
134. N A Naryshkin, M G Ivanovskaya, T S Oretskaya, E M Volkov, M J Gait, Z A Shabarova *Bioorg. Khim.* **22** 691 (1996) [b]
135. N A Naryshkin, M A Farrow, M G Ivanovskaya, T S Oretskaya, Z A Shabarova, M J Gait *Biochemistry* **36** 3496 (1997)
136. J Karn *J. Mol. Biol.* **293** 235 (1999)
137. S E Filippova, M G Ivanovskaya, E A Romanova, V L Tunitskaya, S N Kochetkov *Mol. Biol.* **36** 684 (2002) [d]
138. J P Menetski *J. Biol. Chem.* **275** 7619 (2000)
139. B Berkowitz, D-B Huang, F E Chen-Park, P B Sigler, G Ghosh *J. Biol. Chem.* **277** 24694 (2002)
140. K Wecker, M C Bonnet, E F Meurs, M Delepierre *Nucleic Acids Res.* **30** 4452 (2002)
141. C R Escalante, L Y Shen, D Thanos, A K Aggarwal *Structure* **10** 383 (2002)
142. G Y Sheflyan, E A Kubareva, E S Gromova, Z A Shabarova *Bioconjugate Chem.* **9** 703 (1998)
143. O A Borisova *Gorizonty Fiziko-Khimicheskoi Biologii (Shkola-Konferentsiya), Pushchino, 2000* [Horizons of Physicochemical Biology (The School-Conference), Pushchino, 2000] p. 28
144. S A Kuznetsova, M G Ivanovskaya, Z A Shabarova *Bioorg. Khim.* **17** 1633 (1991) [b]
145. A V Kachalova, D A Stetsenko, E A Romanova, V N Tashlitsky, M J Gait, T S Oretskaya *Helv. Chim. Acta* **85** 2409 (2002)
146. A V Kachalova, E M Zubin, T S Oretskaya *Usp. Khim.* **71** 1173 (2002) [*Russ. Chem. Rev.* **71** 1041 (2002)]
147. Z A Shabarova, M G Ivanovskaya, M B Gottikh, in *Nucleic Acid Chemistry Improved and New Synthetic Procedures Methods and Techniques* Pt. 4 (Eds L B Tomnsend, R S Tipson) (New York: Wiley, 1991) p. 386
148. M G Ivanovskaya, M B Gottikh, Z A Shabarova *Nucleosides Nucleotides* **6** 913 (1987)
149. E A Taran, M G Ivanovskaya, M J Gait, Z A Shabarova *Mol. Biol.* **32** 832 (1998) [d]
150. N Naryshkin, A Revyakm, Y Kim, V Mekler, E H Ebright *Cell* **101** 601 (2000)
151. N M Luscombe, R A Laskowski, J M Tharnton *Nucleic Acid Res.* **29** 2860 (2001)
152. O A Borisova, A S Romanenkov, V G Metelev, E A Kubareva, E M Zubin, M A Timchenko, T S Oretskaya *Mol. Biol.* **37** 534 (2003) [d]
153. M V Reed, D Fraga, D E Schwartz, R D Kinrichsen *Bioconjugate Chem.* **6** 101 (1995)
154. O M Gritsenko, E S Gromova *Usp. Khim.* **68** 267 (1999) [*Russ. Chem. Rev.* **68** 241 (1999)]
155. B S Ermolinskii, S N Mikhailov *Bioorg. Khim.* **26** 483 (2000) [b]
156. E V Efimtseva, S N Mikhailov *Biokhimiya* **67** 1374 (2002) [c]
157. A V Kachalova, T S Zatsepin, E A Romanova, D A Stetsenko, M J Gait, T S Oretskaya *Nucleosides Nucleotides Nucleic Acids* **19** 1693 (2000)
158. T S Zatsepin, A V Kachalova, E A Romanova, D A Stetsenko, M J Gait, T S Oretskaya *Bioorg. Khim.* **27** 45 (2001) [b]
159. T S Zatsepin, D A Stetsenko, A A Arzumanov, E A Romanova, M J Gait, T S Oretskaya *Bioconjugate Chem.* **13** 822 (2002)
160. D V Turutin, T S Zatsepin, M A Timchenko, E A Kubareva, T S Oretskaya *Mol. Biol.* **36** 877 (2002) [d]
161. O V Vorob'eva, O V Yamskova, T S Zatsepin, in *Proceedings of IQN Spring Meeting 'Nucleic Acids and Nucleo-Protein-Complexes', Giessen, Germany, 2003*
162. M Hashimoto, M M Greenberg, Y W Kow, J-T Hwang, R P Cunningham *J. Am. Chem. Soc.* **123** 3161 (2001)
163. M S DeMott, E Beyret, D Wong, B C Bales, J-T Wang, M M Greenberg, B Demple *J. Biol. Chem.* **277** 7637 (2002)
164. K M Kroeger, M Hashimoto, Y W Kow, M M Greenberg *Biochemistry* **42** 2449 (2003)

165. R Singh, G M Whitesides *Chemistry of Sulphur-Containing Functional Groups* (Eds S Patai, Z Rappoport) (New York: Wiley, 1993) p. 634
166. R T Raines *Nature Struct. Biol.* **4** 424 (1997)
167. S R Paalman, D M Noll, N D Clarke *Nucleic Acids Res.* **25** 1795 (1997)
168. M H Moore, J M Gulbis, E J Dodson, B Demple, P C Moody *EMBO J.* **13** 1495 (1994)
169. K Gao, S L Butler, F Bushman *EMBO J.* **20** 3565 (2001)
170. H Huang, R Chopra, G L Verdine, S C Harrison *Science* **282** 1669 (1998)
171. H Huang, S C Harrison, G L Verdine *Chem. Biol.* **7** 355 (2000)
172. E N Peletskaya, P L Boyer, A A Kogon, P Clark, H Kroth, J M Sayer, D M Jerina, S H Hughes *J. Virol.* **75** 9435 (2001)
173. S G Sarafianos, A D Clark, S Tuske, C J Squire, K Das, D Sheng, P Ilankumaran, A R Ramesha, H Kroth, J M Sayer, D M Jerina, P L Boyer, S H Hughes, E Arnold *J. Biol. Chem.* **278** 16280 (2003)
174. C A He, G L Verdine *Chem. Biol.* **9** 1297 (2002)
175. N Dolinnaya, V Metelev, T Oretskaya, D Tabatadze, Z Shabarova *FEBS Lett.* **444** 285 (1999)
176. N G Dolinnaya, V G Metelev *Bioorg. Khim.* **26** 306 (2000) [b]
177. V G Metelev, O A Borisova, E M Volkov, T S Oretskaya, N G Dolinnaya *Nucleic Acids Res.* **29** 4062 (2001)
178. V G Metelev, E A Kubareva, E A Romanova, T S Oretskaya *Bioorg. Khim.* **29** 57 (2003) [b]
179. V G Metelev, E A Kubareva, O V Vorob'eva, A S Romanenkov, T S Oretskaya *FEBS Lett.* **538** 48 (2003)
180. A S Romanenkov, O A Borisova, E A Kubareva, T S Oretskaya, V G Metelev *Vestn. Mosk. Univ., Ser. 2: Khim.* **44** 208 (2003) [e]
181. Z Shaked, R P Szajewski, G M Whitesides *Biochemistry* **19** 4156 (1980)
182. V Pingoud, C Conzelmann, S Kinzebach, A Sudina, V Metelev, E Kubareva, J M Bujnicki, R Lurz, G Lueder, S-Y Xu, A Pingoud *J. Mol. Biol.* **329** 913 (2003)
183. R H Abeles, T A Alston *J. Biol. Chem.* **265** 16705 (1990)
184. S Cox, J Harmenberg *J. Biochem. Biophys. Methods* **25** 17 (1992)
185. F B Fauman, R M Blumenthal, X Cheng, in *S-Adenosylmethionine-Dependent Methyltransferases: Structure and Functions* (Eds X Cheng, R M Blumenthal) (Singapore, London, Hong Kong: World Scientific, 1999) p. 1
186. X Gu, D V Santi *Biochemistry* **31** 10295 (1992)
187. S Nunes-Duby, L Matsumoto, A Landy *Cell* **59** 197 (1989)
188. S Nunes-Duby, M Azaro, A Landy *Curr. Biol.* **5** 139 (1995)
189. R S Tirumalai, E Healey, A Landy *Proc. Natl. Acad. Sci. U.S.A.* **94** 6104 (1997)
190. M R Redinbo, L Stewart, P Kuhn, J J Champoux, W G Hol *Science* **279** 1504 (1998)
191. B L Staker, K Hjerrild, M D Feese, C A Behnke, A B Burgin, L Stewart *Proc. Natl. Acad. Sci. U.S.A.* **99** 15387 (2002)
192. D M Noll, N D Clarke *Nucleic Acids Res.* **29** 4025 (2001)
193. A D Chheda, G W Teebor, R P Cunningham *Methods* **22** 180 (2000)
194. A Gogos, J Cillo, N D Clarke, A-L Lu *Biochemistry* **35** 16665 (1996)
195. S D Willams, S D David *Nucleic Acids Res.* **26** 5123 (1998)
196. D O Zharkov, A P Grollman *Biochemistry* **37** 12384 (1998)
197. X Liu, R Roy *Biochemistry* **40** 13617 (2001)
198. A J Kurtz, M L Dodson, R S Lloyd *Biochemistry* **41** 7054 (2002)
199. J C Fromme, S D Bruner, W Jang, M Karplus, G L Verdine *Nature Struct. Biol.* **10** 204 (2003)
200. O M Gritsenko, E V Koudan, S N Mikhailov, B S Ermolinsky, A van Aerschot, P Herdewijn, E S Gromova *Nucleosides Nucleotides Nucleic Acids* **21** 753 (2002)
201. D E Kamashev, A V Balandina, V L Karpov *J. Biol. Chem.* **275** 36056 (2000)
202. U K Laemmli *Nature (London)* **227** 680 (1970)
203. L J Deterding, R Prasad, G P Mullen, S H Wilson, K B Tomer *J. Biol. Chem.* **275** 10463 (2000)
204. R A Rieger, M M McTigue, J H Kycia, S E Gerchman, A P Grollman, C R Iden *J. Am. Soc. Mass Spectrom.* **11** 505 (2000)
205. H Urlaub, K Hartmuth, S Kostka, G Grelle, R Luhrmann *J. Biol. Chem.* **275** 41458 (2000)
206. S Raveh, J Vinh, J Rossier, F Agou, M Veron *Biochemistry* **40** 5882 (2001)
207. S Sastry, B M Ross *Biochemistry* **38** 4972 (1999)
208. P Edman *Arch. Biochem. Biophys.*, **22** 475 (1949)
209. M Karas, K F Hillenkamp *Anal. Chem.* **60** 2299 (1988)
210. J B Fenn, M Mann, C K Meng, S F Wong, C M Whitehouse *Science* **246** 64 (1989)
211. P F Crain, J A McCloskey *Curr. Opin. Biotechnol.* **9** 25 (1998)
212. D R Owens, B Bothner, Q Phung, K Harris, G Siuzdak *Bioorg. Med. Chem.* **6** 1547 (1998)
213. M L Nielsen, K L Bennet, B Larsen, M Moniatte, M Mann *J. Proteome Res.* **1** 63 (2002)
214. J E Redman, K M Wilcoxen, M R Ghadiri *J. Comb. Chem.* **5** 33 (2003)
215. A E Ashcroft *Nat. Prod. Rep.* **20** 202 (2003)
216. O N Jensen, D F Barofsky, M C Young, P H von Hippel, S Swenson, S E Seifried *Rapid Commun. Mass Spectrom.* **7** 496 (1993)
217. S F Bennet, O N Jensen, D F Barofsky, D W Mosbaugh *J. Biol. Chem.* **269** 21870 (1994)
218. O N Jensen, S Kulkarni, J V Aldrich, D Barofsky *Nucleic Acids Res.* **24** 3866 (1996)
219. H Urlaub, B Thiede, E C Muller, B Wittmann-Liebold *J. Protein Chem.* **16** 375 (1997)
220. N Tretyakova, B Matter, A Ogdie, J S Wishnok, S R Tannenbaum *Chem. Res. Toxicol.* **14** 1058 (2001)
221. J Qin, B T Chait *Anal. Chem.* **69** 4002 (1997)
222. H Urlaub, B Thiede, E C Muller, R Brimacombe, B Wittmann-Liebold *J. Biol. Chem.* **272** 14547 (1997)
223. D Craft, A Doucette, L Li *J. Proteome Res.* **1** 537 (2002)
224. J L Beck, M L Colgrave, S F Ralph, M M Sheil *Mass Spectrom. Rev.* **20** 61 (2001)
225. S A Hofstadler, R H Griffey *Chem. Rev.* **101** 377 (2001)
226. H Steen, O N Jensen *Mass Spectrom. Rev.* **21** 163 (2002)

[a] — *Russ. Chem. Bull., Int. Ed. (Engl. Transl.)*
[b] — *Russ. J. Bioorg. Chem. (Engl. Transl.)*
[c] — *Biochemistry (Moscow) (Engl. Transl.)*
[d] — *Mol. Biol. (Engl. Transl.)*
[e] — *Moscow Univ. Bull. (Engl. Transl.)*