Russian Chemical Reviews **65** (8) 709–724 (1996)
© 1996 Russian Academy of Sciences and Turpion Ltd

UDC 577.113.4

Methods for the covalent attachment of nucleic acids and their derivatives to proteins

G Ya Sheflyan, E A Kubareva, E S Gromova

Contents

I.	Introduction	709
II.	Covalent attachment of proteins to nucleic acids by exposure to different types of radiation	709
III.	Chemical methods for the covalent attachment of proteins to nucleic acids	714
IV.	Covalent attachment of nucleic acids to enzymes based on the mode of enzyme action	716
V.	Methods for the identification and isolation of covalent protein–DNA complexes	718
VI.	Determination of peptides in a protein bound covalently to nucleic acid	720
VII.	Probing the DNA recognising proteins mode of action via cross-linking assay	722
VIII.	Conclusion	722

Abstract. The methods for the affinity modification of proteins, specific to nucleic acids, by activated nucleic acid analogues and their components, are considered. The active groups introduced into proteins or nucleic acids and the symmetrical cross-linking agents used to obtain covalent protein–nucleic acid complexes are listed. The methods for the analysis of covalently linked protein–nucleic acid adducts as well as procedures for the identification of the aminoacids in the protein participating in the formation of bonds with nucleic acids are surveyed. The bibliography includes 112 references.

I. Introduction

There are comparatively few methods which make it possible to determine the groups in the protein participating in protein–nucleic acid recognition. The most accurate of these is X-ray diffraction (XRD). However, owing to the poor resolution of X-ray diffraction patterns, the difficulty of isolating sufficient amounts of the protein, and also the differences between the conformations of the macromolecules in solution and in the crystal, this method has not found extensive application. Furthermore, it is not always possible to obtain and isolate specific complexes between proteins and DNA. There are also problems associated with their crystallisation.

Site-specific mutagenesis and chemical modification of the proteins also yield valuable information about the interaction of proteins with nucleic acids, but such experiments require, at least to a first approximation, the knowledge of location of the protein region responsible for this process. A method which makes it possible to identify this region is covalent attachment of proteins to nucleic acids (NA).

G Ya Sheflyan, E A Kubareva, E S Gromova Belozerskii Institute of Physical and Chemical Biology at the Lomonosov Moscow State University, Lomonosov Moscow State University, Chemistry Department, Vorob'evy gory, 119899 Moscow, Russian Federation. Fax (7-095) 939 31 81. Tel. (7-095) 939 54 20 (G Ya Sheflyan), 939 31 48 (E A Kubareva), and 939 31 44 (E S Gromova)

Received 13 December 1995
Uspekhi Khimii **65** (8) 765–781 (1996); translated by A K Grzybowski

Numerous methods for the attachment of proteins to nucleic acids are known at the present time. We have attempted to arrange systematically the available literature data on this question. In our view, the existing methods for the affinity modification of proteins by nucleic acids can be divided into three groups:

– the attachment of proteins to nucleic acids by exposure to different types of radiation;

– chemical methods for the covalent attachment of proteins to nucleic acids;

– covalent attachment of nucleic acids to enzymes based on the mechanisms of the functioning of the latter.

Nucleic acid analogues containing different active groups are usually employed for the affinity modification of the protein. Such groups may be introduced into heterocyclic bases and carbohydrate residues via the terminal or internucleotide phosphate group of the nucleic acid. So, to our mind it is necessary to concentrate attention also on the approaches for the activation of nucleic acids used in each of these methods for the covalent attachment of proteins to nucleic acids.

In the concluding section of the review, the approaches to the identification and isolation of covalent protein–nucleic acid complexes and for the determination of the peptides in the protein bound covalently to the nucleic acid are surveyed. The possibilities for the application of the method involving the covalent attachment of nucleic acid–specific proteins to nucleic acids for the probing of the mechanisms of their interaction are considered.

II. Covalent attachment of proteins to nucleic acids by exposure to different types of radiation

1. Irradiation with UV light

When complexes between proteins and nucleic acids or their analogues are irradiated with UV light, the side chains of the protein become covalently attached to the heterocyclic base residues closest to them (the so-called 'zero length cross-linking').

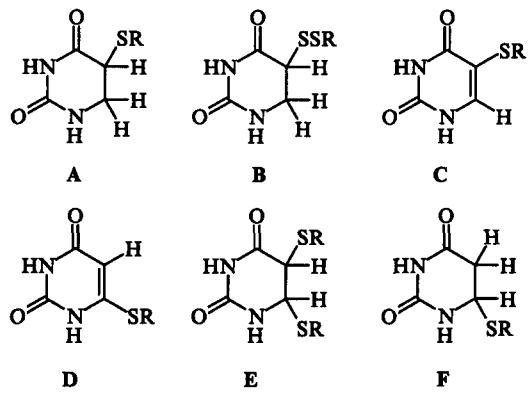
a. Irradiation of complexes between proteins and unmodified nucleic acids with UV light

This method for covalent attachment is used in relation to a wide range of biological systems: from cell nuclei up to isolated and purified proteins in complexes with synthetic oligonucleotides. The conditions for the irradiation of various systems in order to obtain protein — nucleic acid complexes in the appropriate buffer

solutions are surveyed in Table 1. It is seen from the data presented that the irradiation time varies from 1 min to 4 h and the wavelength is as a rule 254 nm. Under these conditions, the yield of the conjugate varies from 0.1% to 70%. Owing to the complexity of the system, in certain cases it was not possible to determine the yield of the final conjugate, so that the stoichiometry of the complexes is presented instead. It reflects the number of protein molecules per DNA molecule or the number of base pairs of the DNA bound covalently to one protein molecule.

On irradiation with UV light of a protein–nucleic acid complex, the 5,6-double bond of the pyrimidines is excited, which leads to its rupture; a biradical is formed at the 5,6-positions in the pyrimidine ring and reacts with the side chains of the protein.[21] This reaction mechanism has been established in a detailed study of the products of the irradiation of uracil in the presence of cysteine (Scheme 1).

Scheme 1



R = $CH_2CH(NH_2)COOH$.

Compounds **A**, **B**, **E**, and **F** are formed on irradiation with light at a wavelength > 290 nm in the presence of acetophenone, which is a 'trap' for the radicals. It has been suggested that the reaction is reversible. The product **C** is formed in solutions containing free radical generators, for example oxygen or riboflavin. The irradiation conditions depend on the energy required for the formation of the radical: when the solution contains oxygen, a wavelength of 254 nm is employed, while in the presence of riboflavin the wavelength is 290 nm. Compound **C** is probably obtained from **E** as the primary product. Compound **D** is apparently formed similarly to **C**, via a free radical in the 5-position in uracil.[22]

At the present time, there is no unanimous view in the literature concerning the mechanism of the covalent cross-linking of unmodified nucleic acids with proteins on exposure to UV light. In a number of studies, for example in that of Welsh and Cantor,[21] it has been suggested that such cross-linking is nonspecific and depends not on the nature of the side chain of the aminoacid residue but on the distance between it and the heterocycle. On the other hand, other workers (see Refs 3, 22, and 23) noted the specificity of the formation of the covalent cross-link in relation to definite aminoacid residues, for example Ser, Ile, Thr, Arg, and Lys.

Saito et al.[3] believe that, when complexes between proteins and nucleic acids are irradiated with UV light, a cross-link is formed between the thymine residue and the ε-amino-group of lysine. The deoxythimidine–L-Lys model system was used to investigate this reaction (Scheme 2).

It was established that, on irradiation with light at a wavelength > 280 nm, the adduct L-Lys–thymidine is not formed. On exposure to UV light, thymidine also interacts with Arg. The reaction involves the α-amino-group of Arg and not the guanidine residue, i.e. Arg in the protein may react with thymidine on irradiation only when it is N-terminal.

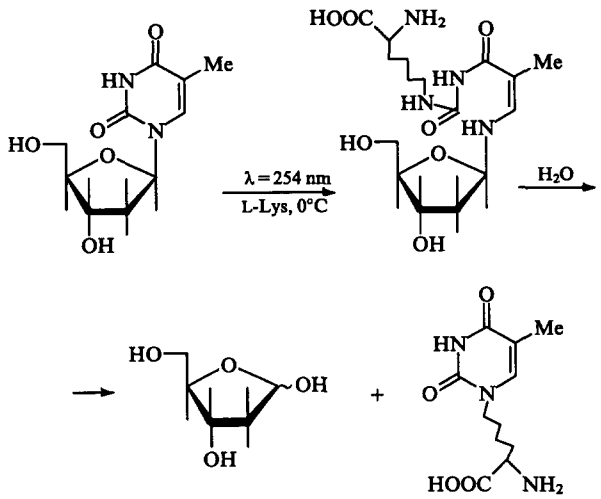
The method for the UV irradiation of proteins and nucleic acids considered in this Section has its advantages and disadvantages.

Table 1. Conditions for the irradiation of complexes between proteins and unmodified nucleic acids and the yields of the reaction products.

Object	Wavelength /nm	Time /min	Yield (%) or stoichiometry of the complex	Ref.
Chick erythrocyte nucleus	254	240	58[a], 70[c]	1
Nucleohistone	230–290	30	see[b]	2
Calf thymus nucleosome	>250	40	see[b]	3
Cells from a *D. melanogaster* tissue culture	not indicated	12	see[b]	4
Trypanosoma brucei cells	254	10	see[b]	5
E. coli RNA polymerase and DNA	254	1–4	0.1	6
E. coli RNA polymerase and phage T7 DNA	254	20–240	see[b]	7
E. coli RNA polymerase and poly-d(AT)	254	not indicated	50 (protein : NA = 3 : 1, 6 : 1 and 12 : 1)	8
σ-Factor of *E. coli* RNA polymerase and phage M13 single-stranded DNA with an oligonucleotide primer	254	20	25[c]	9
The protein snRNP and RNA	254	30	see[b]	10
DNA polymerase and fragments of the phage T7	254	not indicated	protein : NA from 3 : 1 to 10 : 1	11
Ile-tRNA-synthetase and ATP	253.7	30–40	protein : ATP = 6.7 : 1	12
Replication protein A and single-stranded DNA	UV region	10	see[b]	13
DNA-binding adenovirus protein and $p(dT)_{14}$	UV region	20	20[a]	14
Human protein recognising $d(GA)_n$ and $d(GT)_n$ repeats in DNA and $d(GA)_n$	254	100	see[b]	15
Guanidine-acetate methyltransferase and SAM[d]	254	240	17–20[a]	16
DNA methyltransferase CheR from *Salmonella typhymirium* and SAM	254	40	60–70[e]	17
DNA methyltransferase Dam from *E. coli* and SAM	254	5–90	3–4[e]	18, 19
DNA methyltransferase *Mcp*I and SAM	254	60	see[b]	20

[a] The yield calculated relative to the protein is indicated. [b] The yield is not indicated. [c] The yield calculated relative to DNA is indicated.
[d] *S*-Adenosyl-L-methionine. [e] The yield calculated relative to SAM is indicated.

Scheme 2



The former include the possibility of identifying the aminoacid residue closest to the nucleic acid. However, the absence of a covalent adduct after irradiation does not imply the absence of contact between the nucleic acid and protein, since it may be that the activation of the components of the complex by UV irradiation is insufficient for the reaction to occur. A disadvantage of the method is also the damage to the structure of the nucleic acid and the protein for a high intensity and long duration of the irradiation. Under these conditions, the mutations occurring in nucleic acids interfere with the study of the conjugate (for example, owing to a change in mobility in the gel). The inactivation of the protein, which may occur either as a result of its breakdown on irradiation or owing to covalent cross-linking, cannot serve as a test for the formation of a protein–nucleic acid complex.

b. Irradiation with UV light of complexes of proteins with modified nucleic acid analogues

The cross-linking of proteins to photosensitive nucleic acid analogues is used for the determination of the topology of nucleic acid–protein complexes. Nucleic acids where a bromine atom or the azido-group has been introduced in the 5-position in pyrimidine bases and in the 8-position in the purine bases are used as analogues of this kind.[24] The above chemical groups are exposed in the maior groove of the double helix. They do not induce a significant change in its structure and do not prevent protein–nucleic acid recognition.

The conditions governing the cross-linking of the bromo- and azido-substituted nucleic acid analogues to proteins are summarised in Table 2. The table shows that the irradiation conditions for the modified nucleic acid analogues are milder than for the unmodified nucleic acids. The yields of the covalently bound

Table 2. Conditions for the irradiation of complexes between proteins and modified nucleic acid analogues and the yields of the reaction products.

Object	Wave-length /nm	Time /min	Yield of complex (%)	Ref.
Restriction endonucleases *Eco*RI and *Eco*RV and DNA duplex containing br^{5}dU	>300	1	0.05[a]	25
lac repressor and oligonucleotide with *lac*-operator sequence containing br^{5}dU	300–400	20	60[b]	26, 27
Matrix-binding polypeptide in the transcription complex of pea choroplasts and DNA containing br^{5}dU	300–315	20	not indicated	28
Protein of the bacteriophage R17 coat and RNA containing br^{5}dU	312	3	20[a]	29
Transcription activator GCN4 from yeast and DNA containing br^{5}dU	254	1.5	2	30
Cyclic 3′,5′-phosphate receptor protein and *E. coli* DNA containing br^{5}dU	300–400	not indicated	70[b]	31
Chromosomal protein of the archeobacteria MC1 and *E. coli* DNA containing br^{5}dU	300–400	not indicated	20[b]	32
DNA methyltransferase *Eco*RV and DNA containing 4-thiothymidine	>310	10–12	18–35[a]	33
DNA methyltransferase *Eco*RV and DNA containing 6-thiodeoxyguanosine	>310	10–12	5–8[a]	33
Restriction endonuclease *Eco*RV and DNA containing 4-thiothymidine	>310	10–12	6[a]	33
Restriction endonuclease *Eco*RV and DNA containing 6-thiodeoxyguanosine	>310	10–12	6[a]	33
Terminal deoxynucleotidyltransferase and 40-membered DNA containing N_3^{5}dU	302	1	40–50[a]	34
DNA polymerase I and oligonucleotide duplex containing N_3^{5}dU	302	2	40[b]	35
Carboxy-*O*-methyltransferase and 8-azido-SAM	254	15	42[b]	36
E. coli RNA polymerase and N_3^{8}dA	254	8	not indicated	37
70S Ribosomes and tRNA^{Phe} of *E. coli* containing 3-{*N*-[*N*-*p*-azidobenzoyl)glycyl]-(3-amino-3-carboxypropyl)}uridine	UV region	2–3	25	38
Endonuclease V of phage T4 and 14-membered DNA duplex containing 3-trifluoromethyl-3-(*p*-carboxybenzoyl)diazirine	365	30	15[a]	39

[a] The yield calculated relative to the nucleic acid is indicated. [b] The yield calculated relative to the protein is indicated.

protein–nucleic acid conjugates are fairly large and in certain systems amount to 60%.

The introduction of photosensitive groups into a heterocyclic base alters the reaction mechanism. Irradiation of the complexes of bromo-substituted nucleic acid analogues with a protein is accompanied by the loss of the bromine atom and the formation of an unpaired electron in the 5-position of the uracil residue, which may react in different ways with the protein side chains, giving rise to a covalent cross-link.[21] The only condition for the occurrence of such cross-linking is the proximity of the side chain of the aminoacid residue and of the free radical. Wolfes et al.[25] believe that the interaction of the bromine-containing substrates with the Gly, Ser, Phe, Tyr, Trp, Cys, and Met residues, having aliphatic chains or aromatic rings, as possible free radical acceptors is most likely. The mechanism and specificity of this reaction were not investigated in greater detail. Among the advantages of the method considered one should mention the high sensitivity of the modified nucleic acids to UV irradiation, as a result of which the probability of mutations in the nucleic acids and in the protein diminishes owing to the smaller intensity and duration of the irradiation.

The azido-substituted nucleic acid analogues are even more effective in the covalent binding reaction.[40] In the case of nucleoside residues containing the azido-group in the 8-position in the purine bases or in the 5-position in the pyrimidine bases, the mechanism of the formation of the covalently bound protein–nucleic acid complex is much more complicated. On irradiation with UV light, the photochemical elimination of a nitrogen atom results in the formation not of a radical but of a nitrene, which may react nonselectively with the closest groups regardless of their electrophilicity or nucleophilicity.[37] An indirect demonstration of this mechanism has been achieved in a spectrophotometric study of the consequences of the UV irradiation of $N_3{}^5$dUTP.[40] After only 20 s, the maximum at 288 nm, corresponding to the azido-derivative of dUTP, vanishes completely. The decrease in absorption after photolysis indicates the rupture of the conjugated ring system of the heterocyclic base (possibly as a result of an intramolecular rearrangement initiated by the formation of the nitrene).

The particular value of the azido-derivatives of nucleic acids consists in the fact that the azido-group is fairly small and does not prevent the formation of hydrogen bonds between the protein and the nucleic acid. Furthermore, the introduction of the azido-group in the 5-position in the pyrimidine residues does not alter the *anti*-conformation of the heterocyclic base of the nucleoside and only induces a slight distortion of the nucleic acid structure.[40,41] Reich and Everett[41] compared the kinetic parameters of the interaction of the methyltransferase *Eco*RI with the cofactor SAM (*S*-adenosinyl-L-methionine) and also its analogues containing the azido-group in the 8-position of the heterocyclic base. It was found that the constants for binding to the enzyme in the absence of DNA are of the same order of magnitude for the native cofactor and its azido-analogues, amounting to 4.8 and 12.9 μM respectively. The catalytic constant (k_{Cat}) and the Michaelis–Menten constant (K_M) for SAM and 8-azido-SAM in ternary complexes with the enzyme and with the 12-membered DNA duplex hardly differ, amounting to 4.3 s^{-1} and 0.335 μM for the native cofactor and to 5.0 s^{-1} and 0.710 μM respectively for the modified compound. These results indicate the insignificant nature of the changes in the structure of the nucleoside following the introduction of the azido-group. When acted upon by UV light, 8-azido-SAM becomes covalently attached to the enzyme. The specificity of the covalent attachment has been demonstrated by its inhibition on addition of the unmodified cofactor.[41]

Another advantage of the azido-derivatives is the fact that the conversion of the azido-group into the nitrene occurs at a much lower intensity of the UV light, which makes it possible to avoid damage to DNA on irradiation. The initial rates of disruption of the heterocycle on irradiation with UV light at a wavelength of 254 nm have been compared.[40] It was established that a decrease in maximum absorption for $N_3{}^5$dUMP was 7000 times faster than for dTTP and 1150 times faster than for br^5dUMP. Thus the azido-analogues are much more reactive in relation to proteins. There are no data on the specificity of their action.

Thymidine and deoxyguanosine in which the keto-group has been replaced by the thio-group have been tried as photosensitive analogues of nucleosides.[33] It was established that the DNA duplexes modified in this way become covalently attached to restriction–modification enzymes *Eco*RI and *Eco*RV (the conditions for their irradiation are presented in Table 2). It was shown in control experiments that the enzymes themselves retain fully their activity after irradiation; the irradiation of enzymes with unmodified DNA does not lead to the formation of covalently bound adducts; incubation of the components of the complex without irradiation also does not induce the formation of conjugates.

By introducing 4-thiothymidine residues into genes coding for 5S rRNA and $tRNA^{Tyr}$ of the yeast *Saccharomyces cerevisiae*, the interaction of such DNA with transcription factors — TFIIIC and TFIIIB — was investigated. It was shown that the modification does not prevent the formation of noncovalent specific complexes with the proteins considered. Photoattachment occurred after the preincubation of the complexes on irradiation with UV light at a wavelength of 300 nm for 15 min. The method made it possible to identify the subunits of the transcription factors in contact with particular regions of the transcribed DNA.[42]

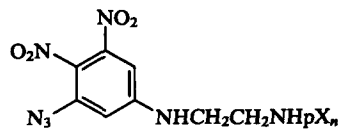
The modification of mRNA and certain ribosomal RNA with the 4-thiouridine residue has been used in the study of the topography of ribosomes in early stages of the transcription.[43–48] It was shown that, in the transcription involving T7-polymerase, 4-thiouridine triphosphate is incorporated in the synthesised RNA strand in the same way as the unmodified nucleoside triphosphates.[43–48] When ribosomes containing the modified components were irradiated with UV light at a wavelength > 300 nm for 10 min, cross-links were formed between mRNA and the ribosomal RNA[43–47] as well as RNA–protein cross-links. The ribosomal proteins bound covalently to a synthetic analogue of mRNA containing the 4-thiouridine residue have been identified.[48] The method has been used to determine the proteins spatially in close proximity to mRNA at different transcription stages.

We shall consider two possible mechanisms of the reaction of proteins with nucleic acids containing a 4-thiopyrimidine base: in the presence of oxygen and in its absence.[33] In the latter case, the protein is attached to the C(6) atom of the pyrimidine ring with the simultaneous reduction of the 5,6-double bond. The reaction proceeds via a radical mechanism, which involves the elimination of a hydrogen atom from 4-thiopyrimidine in the excited state.

In the presence of oxygen, the 4-thiopyrimidines in the nucleic acid are oxidised to the corresponding 4-sulfonates, which are highly reactive and are subjected to nucleophilic attack by the aminoacid residues of the protein with substitution of the sulfonate group.

There are data concerning the mechanism of the photoattachment of nucleic acids containing 6-thiodeoxyguanosine to proteins. It is believed that they form sulfonate derivatives as a result of irradiation.[33]

Another type of activation of nucleic acids involves the formation of oligoribonucleotides or oligodeoxyribonucleotides in which the photoactivated group is not attached directly to the heterocyclic base. Thus oligoribonucleotides containing a dinitrobenzene ring with the azido-group and linked to the 5′-phosphate group via ethylenediamine have been used for the affinity modification of ribosomal proteins.[49]



X = ribonucleoside residue

Mitchell et al.[38] modified tRNA with *N*-(*p*-azidobenzoyl)glycine by attaching it to the 3-(3-amino-3-carboxypropyl)uridine residue (the irradiation conditions are presented in Table 2). The mechanism of the reaction involving the attachment of nucleic acids modified in this way to proteins does not differ from that considered above, since in this case too the main photoactivated reactive group is the azido-group.

Apart from the nitrene, a carbene may be used for the covalent attachment to proteins. The formation of a carbene from the diazirine-derivative of the nucleic acid is induced by irradiation with UV light.[39] The active group — 3-trifluoromethyl-3-(*p*-carboxybenzoyl)diazirine — was attached to the internucleotide phosphate in the DNA duplex. As a result of irradiation with UV light at a longer wavelength (Table 2), this group decomposes, affording a carbene. It is postulated that the carbenes are more reactive than the nitrenes and, furthermore, that irradiation with UV light at a longer wavelength is preferable, since it prevents the degradation of the protein.

c. The use of laser radiation

Not only a mercury or mineral lamp but also a laser may be used as a source of UV light. The difference between laser irradiation and the usual irradiation consists in the extremely high power and brief duration of the light pulse. Laser radiation was first used for the covalent attachment of proteins to nucleic acids in relation to the complex between the *E.coli* RNA polymerase and the phage T7 DNA. The reaction mixture was irradiated with a single pulse having an energy of 120 mJ (irradiation time 20 ns, wavelength 248 nm). Depending on the concentration of mercaptoethanol, the yield of the complex varied from 5% to 20%.

The mechanism of the covalent attachment of proteins to nucleic acids on exposure to laser irradiation does not differ from that of the attachment on irradiation by light from a mercury or mineral lamp. However, the use of a brief pulse and the presence of mercaptoethanol as the 'trap' for the long-lived radicals formed on irradiation made it possible to characterise in greater detail the intermediate compounds formed on irradiation by a laser. The lifetime of the radicals formed on irradiation is calculated from the formula

$$\frac{f_0}{f_q} = 1 + K_q t\,,$$

where f_0 and f_q are the yields of the product of the covalent attachment in the absence and in the presence of mercaptoethanol respectively, K_q is the rate constant for the trapping of radicals, and t the lifetime of the radical.[50] It was established that $t = 25-4000$ ns and that approximately half of the radicals are long-lived. Since the lifetime of a particular conformation of the DNA–protein complex proved to be much greater than the duration of the laser pulse, it became possible to 'freeze' a particular structure of the test object, but this is interfered with by the long-living radicals because the latter may react with the protein already after the transition of the system to another state (the so-called diffusion of radicals). In order to avoid such artefacts, it has been suggested that mercaptoethanol be used as a 'trap' for the long-lived radicals.

The yield of the product of the covalent attachment of a protein to the nucleic acid under the influence of laser radiation depends linearly on the irradiation dose, which indicates that only one photon participates in each cross-linking step and that the excitation of only one of the components of the complex is sufficient for the reaction. An advantage of the method is smaller damage to DNA compared with other methods, the protein fully retaining its native state under these conditions.[50]

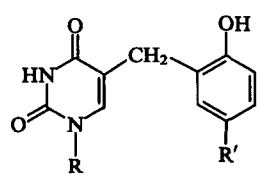
On exposure to laser irradiation, the DNA modified with the psoralen furanoside residue can also be attached to proteins, which has been demonstrated in relation to their attachment to the RNA polymerase of the bacteriophage T7.[51] The irradiation was carried out with the aid of an argon laser having a broad radiation spectrum and a maximum at 366 nm. For an equimolar ratio of DNA and the protein in the reaction mixture, approximately 80% of the DNA is attached covalently to the RNA polymerase. It has been observed by electrophoresis that several isomeric covalent protein–nucleic acid complexes are formed. Degradation of the RNA polymerase under these conditions was not observed. An increase in the irradiation intensity led to a linear growth of the yield of the covalent adducts. The advantages of this method are the short irradiation time and the possibility of 'freezing' the equilibrium states of the DNA–protein complex, which makes it possible to investigate the kinetic constants for this process. Two mechanisms of the attachment of DNA, modified by psoralen furanoside, to a protein have been proposed: (1) via the formation of a singlet oxygen atom, radicals, or radical-ions; (2) by direct photoattachment. It is noteworthy that evidence in support of one of these photoactivation pathways has not been found, but the covalent attachment yield is directly related to the influence of external conditions on the stability of the protein–nucleic acid complex.[51]

2. Irradiation from γ-sources

Irradiation from γ-sources has also been used to obtain DNA–protein conjugates. For this purpose, the corresponding complexes were placed in a saturated solution of nitrous oxide.[52,53] A nucleosome became the object of such studies. A nucleosome consists of a repeating chromatin unit and incorporates DNA consisting of 200 nucleotide pairs and the histones H2A, H2B, H3, and H4.[54] The study of this system is important for the understanding of the mechanisms of the storage of genetic information.

The covalent attachment of histones to DNA occurs on exposure to γ-irradiation. The formation of a covalent bond between the pyrimidine residues of DNA and the tyrosine residues of the protein has been postulated.[53] The tyrosine–cytidine system in an aqueous solution of nitrous oxide has been used to model the formation of the DNA–histone conjugate.[52] The reaction mechanism is as follows. Irradiation of the solution induces the formation of several types of radicals. Hydroxyl radicals react with cytosine and tyrosine, becoming attached to the C(5) carbon atom in the first case and to the aromatic ring with formation of dihydrocyclohexadienyl radicals in the second. Two possible reaction propagation pathways are presented in Scheme 3.

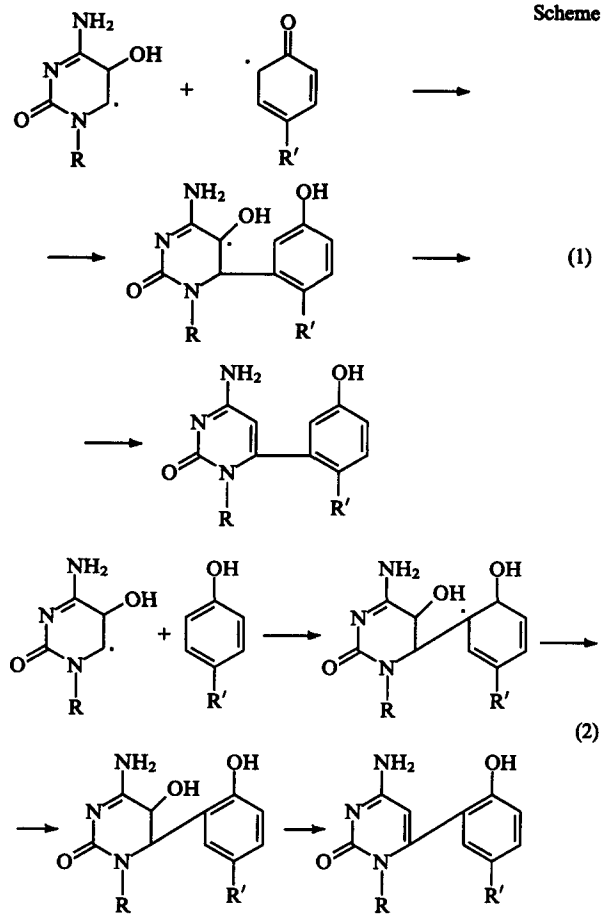
Dizdaroglu et al.[53] considered the possibility of the interaction of tyrosine with thymidine on γ-irradiation. Overall, the mechanism of this reaction is similar to that considered above for cytidine, but the radicals are formed as a result of the abstraction of a hydrogen atom from the methyl group of thymidine and not via the 6-position of the heterocycle (as in the case of cytidine).



R = ribose residue, R′ = $CH_2CH(NH_2)COOH$.

The presence of oxygen molecules diminishes the probability of the formation of a covalent protein–nucleic acid complex, since oxygen forms peroxyl radicals on interaction with the pyrimidine residues.

Scheme 3



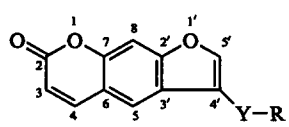
R = ribose residue, R′ = $CH_2CH(NH_2)COOH$.

Arundel et al.[55] suggest that the covalent attachment of DNA to proteins from HCA-1 cells after their treatment with *N*-methylformamide and γ-irradiation takes place via this mechanism. *N*-Methylformamide does not apparently participate directly in the formation of a cross-link and acts only as a radical trap in order to increase the radiosensitivity of the cells.

Irradiation from γ-sources ensures a specific interaction of DNA with aromatic aminoacid residues in close proximity. However, a serious obstacle to its wide scale employment is the need to use a saturated solution of nitrous oxide, which does not promote the formation of protein–nucleic acid complexes. Furthermore, hard γ-radiation induces mutations.

3. Modification of the protein

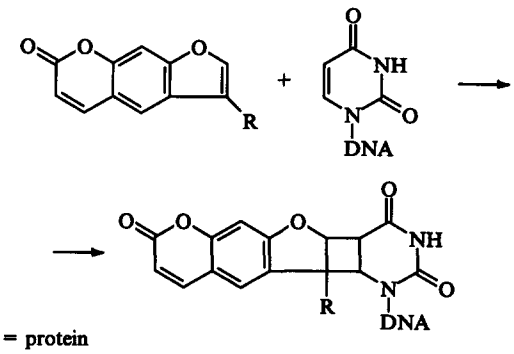
Modification of the protein side chains with psoralen via the corresponding 'stem' (Y) is most frequently employed.



Y = $-NH(CH_2)_2S-$, $-NMe(CH_2)_2NMe-$,
$-NMe(CH_2)_2O(CH_2)_2NMeC(O)(CH_2)_2S-$;
R = protein

In the formation of the protein–nucleic acid complex, psoralen is intercalated in the double helix. When such a complex is irradiated with UV light, monoadducts with an internal cross-link between the furocoumarin and pyrimidine residues are formed. The 3,4- or 4′,5′-double bonds of furocoumarin and the 5,6-double bond of the pyrimidine residue participate in the attachment. This results in the formation of four-membered rings (Scheme 4).[21,56]

Scheme 4



R = protein

By altering the length of the stem connecting psoralen to the protein, it is possible to attach it at both short and long distances from the sites of the interaction of the protein and the nucleic acid. A disadvantage of the method is that it is necessary to know beforehand which site in the protein has to be modified.

III. Chemical methods for the covalent attachment of proteins to nucleic acids

For the covalent attachment of proteins to nucleic acids by means of a chemical reaction not involving the generation of radicals with the aid of irradiation, it is necessary either to increase the activity of one of the components of the protein–nucleic acid complex or to introduce into the system a symmetrical cross-linking agent, capable of forming a covalent bond with both components of the complex. Aldehydes are usually employed as agents of this kind.

1. The use of symmetrical cross-linking agents

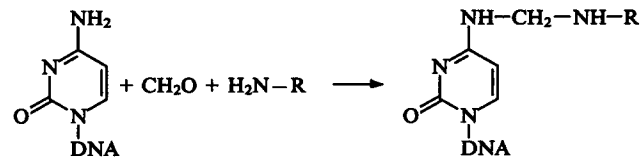
Formaldehyde is used more often than other chemical agents for the covalent attachment of DNA to a protein. The cross-linking of DNA and histones in the nucleosome of the calf thymus nucleus has been achieved in this way. The aim of the study was to identify the sites of contact between DNA and the proteins or to achieve the fixation of protein–nucleic acid complexes in order to prevent their separation during further study on centrifugation and on exposure to other severe treatments.[57,58]

Another example of the application of this method is the probing of the structure of the *E. coli* RNA polymerase and its complex with the lac UV5 promoter.[59]

Formaldehyde has also been used to determine the contacts between mRNA or pBR322 DNA with proteins. In this case, covalent attachment was achieved in 3 h at 37 °C on treatment of exponentially growing *E. coli* cells with a 0.75% solution of formaldehyde.[60]

Formaldehyde reacts with the ε-amino-group of Lys, affording a hydroxy-methyl derivative, which reacts in its turn with the amino- and imino-groups of DNA.[21,57] Formaldehyde also interacts with the DNA adenine and cytosine residues via the exocyclic amino-groups, the protons participating in the formation of Watson–Crick pairs not being involved (Scheme 5).[61]

Scheme 5



R = protein

The endocyclic imino-group of cytosine also reacts with formaldehyde. Furthermore, formaldehyde is attached to the endocyclic imino-groups of the thymine and uracil residues.[62] Thus the reaction involving the covalent attachment of proteins to nucleic acids on treatment with formaldehyde is specific to lysine residues in the protein but is nonspecific in relation to the heterocyclic bases in the nucleic acid. An advantage of the method is the possibility of the covalent attachment of the protein to DNA without rupturing the structure of the double helix of the latter. Under these conditions, the groups of atoms located in both the major and minor grooves of DNA enter into the reaction. However, the method considered leads to a large number of protein–protein cross-links, which complicates the analysis of the results of the covalent attachment in complex systems, for example in *E. coli* cells.[60] Apart from formaldehyde, ketones, for example acetophenone and benzophenone, may be used as the cross-linking agents.

Glutaraldehyde and 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide have also been used as symmetrical cross-linking agents. The former reacts with the primary amino-groups of both the protein and the nucleic acid to form Schiff bases. The carbodiimide interacts with the primary amino-groups only of single-stranded DNA.[21]

Cohen et al.[63] and Hearst[64] indicate the possibility of the formation of a bond between DNA and the protein by coordination via the Cr^{6+} ion in the form of a symmetrical covalent or coordination complex between DNA and the enzyme polyadenosyldiphosphoribosylsynthetase. This type of binding has not been investigated, although the conjugates have already been isolated.

2. Activation of nucleic acids

One of the methods for the activation of nucleic acids is acid treatment, which leads to the elimination of purines from DNA. The aldehyde formed as a result of the opening of the furanose ring may react with the amino-groups of the protein to form a Schiff base.[65] However, this reaction cannot be used directly for the preparation of covalent protein–nucleic acid complexes, since the low pH values and the high ionic strength during the acid treatment interfere with the formation of specific protein–nucleic acid complexes. Levina et al.[66] therefore proposed that DNA be subjected to preliminary alkylation at the guanine and adenine residues with the aid of dimethyl sulfate and subsequent heating in a neutral medium. The aldehyde groups formed in the DNA interact with the amino-groups of the protein. Next, β-elimination results in the rupture of one of the DNA strands at the site of attachment to the protein. This approach made it possible to determine the sites of contact between DNA and the protein in nucleosomes[67] and in the complexes between RNA polymerase and the lac UV5 promoter.[68]

A similar method has been used for the probing of the structure of the hyperacetylated region in nucleosomes. The cell nuclei were treated with dimethyl sulfate for 12 h, after which DNA was depurinated with piperidine with subsequent reduction by sodium tetrahydroborate. This resulted in the formation of covalently bound nuclear particles, in which DNA was attached to the H1/H5 histones.[69]

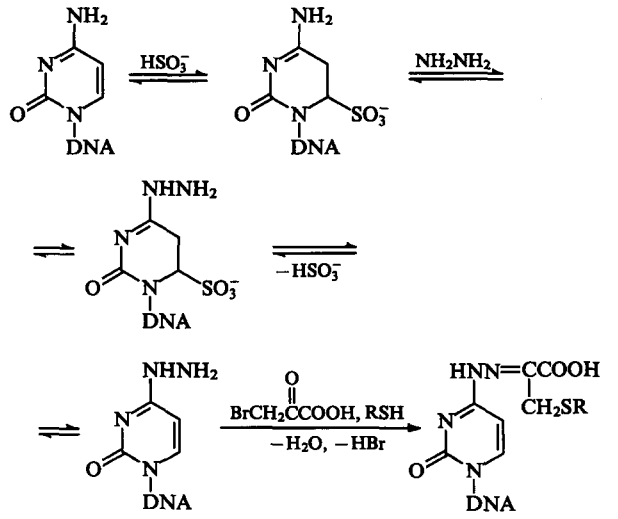
This method has been used in combination with two-dimensional diagonal electrophoresis for the identification of the proteins forming discrimination contacts with the DNA of the cowpox virus.[70]

A disadvantage of the DNA depurination method, used for the covalent attachment of DNA to DNA-specific proteins, is the distortion of the steric structure of the DNA double helix arising from the multiple methylation of the heterocyclic bases.

Another method for the preparation of protein–nucleic acid conjugates is the modification of the nucleic acid fragments at the heterocyclic bases and the phosphate groups by active alkylating agents.

A procedure for the introduction of a bromoacyl residue into cytidine has been proposed (Scheme 6).[71] The hydrazino-group,

Scheme 6

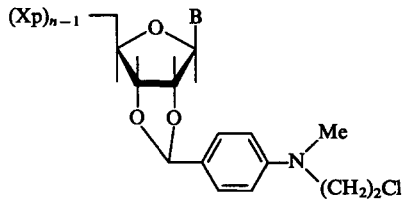


R = protein

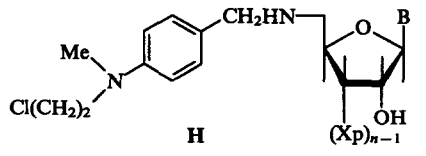
introduced into the 4-position of the cytosine ring, interacts with bromopyruvate, which combines with an SH group of the protein. Glutathione has been used as the donor of the SH group in a model system. The optimum conditions for this reaction have been selected. The covalently bound complex of RNA with the protein in the 30S ribosomal subparticle of *E. coli* has been obtained in this way.[71]

In another study[72], the bromoacetyl residue was attached to the 5′-phosphate group of an oligodeoxyribonucleotide. The active derivative of the oligonucleotide obtained — $BrCH_2CONH(CH_2)_2NH$-d(pTCTAG) — was used for the affinity modification of the protein of gene 5 of the bacteriophage f1. The conditions in the cross-linking reaction were optimal for the formation of a specific protein–nucleic acid complex. The maximum yield of the covalent complex was 20% for a reaction time of 24 h. The decrease in the yield of the protein–nucleic acid complex in the presence of sodium chloride (to 1.3%) and in the presence of competing attachment inhibitors — oligonucleotides not containing an alkyl group (to 9.4%) — demonstrates the specificity of the cross-linking. Since bromoacetate interacts with the nucleophilic centres of the protein, it has been suggested[72] that the SH group of the cysteine residue of the protein of gene 5 of the bacteriophage f1 is alkylated.

Derivatives of oligoribonucleotides containing alkylating groups at the 2′,3′-end (**G**) or the 5′-end (**H**) have been proposed for the study of the topography of ribosomes.



G



H

X = ribonucleoside residue,
B = heterocyclic base residue.

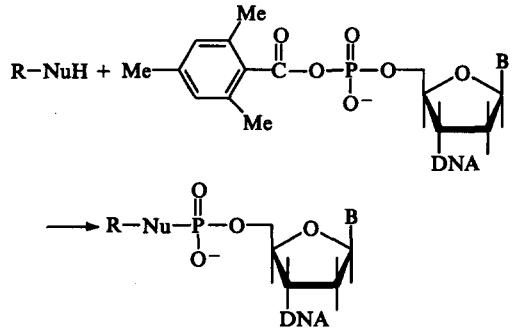
Compounds G and H are active alkylating agents and react with ribosomal proteins. The optimum reaction time is believed to be the time required for the half-conversion of the *N*-chloroethyl-*N*-methylamino-group into the ethyleneimmonium cation as a result of the elimination of the chloride ion. The reaction mixtures are incubated at 25 and 37 °C. The yield of the covalent protein–nucleic acid complex is between 10% and 30%.[49]

This method for the introduction of an active group suffers from a disadvantage associated with the modification of only the terminal residues of the oligoribonucleotides, but it has many advantages because of the reactivity of the alkylating groups.

Nucleic acid analogues containing activated terminal phosphate groups are also used for the affinity modification of proteins. The cyclic substituted trimetaphosphate and *N*-cyclohexyl-*N'*-(4-methylmorpholino)ethylcarbodiimide are used as activators. The reaction time is 10 min and the yields of the product of cross-linking to ribosomal proteins are between 1% and 7.5%.[49]

The mixed anhydrides of phosphoric and mesitylenecarboxylic acids in an aqueous medium can phosphorylate different compounds containing nucleophilic groups, for example primary and secondary amines and mercapto-compounds.[73] This permits their application in the reaction with a protein. Nucleic acid analogues mesitoylated at the terminal phosphate have been used for the covalent attachment to RNase A,[74] mitochondrial ATPase, myosin,[73,75,76] and other proteins (Scheme 7).

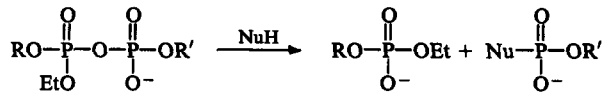
Scheme 7



R = protein, Nu = nucleophilic group of the protein,
B = heterocyclic base residue.

The advantages of the mesitoyl derivatives of oligonucleotides are their stability in aqueous media and low reactivity, which reduces the probability of a nonspecific modification of the protein. According to the results of Sokolova and Tret'yakova,[77] the reactions of the mixed anhydrides of mesitylenecarboxylic and phosphoric acids with strong nucleophiles proceed in several hours, whilst those with weak nucleophiles take several days. The results of the covalent attachment of the mixed anhydride of mesitylenecarboxylic acid and ATP or AMP to myosin were followed from the irreversible inhibition of the enzyme. The reaction time was 1–2 h. In order to obtain the conjugate of RNase A with an oligodeoxyribonucleotide, use was made of reagents such as d[MsC(O)pApApA], d(ApApApMsCO), and d(MsC(O)pTpApGpT), where Ms is the 2,4,6-trimethylbenzoic acid residue. For complexes with a high association constant, the yield of the conjugate was 2%–4% relative to the protein for a reaction time of 7 days. The decrease in the yield of the covalent protein–nucleic acid complex to 0%, which has been noted following a gradual increase in the ionic strength of the solution and when complexes with a low association constant were employed, demonstrates the specificity of the cross-linking.[74]

Shabarova and coworkers[78] proposed DNA duplexes, containing, in a specified position in the sugar-phosphate backbone, a substituted pyrophosphate group instead of the natural phosphodiester linkage, as affinity reagents for covalent attachment to nucleic acid-recognising proteins. It was established that, in single-stranded oligodeoxyribonucleotides in an aqueous medium, the anhydride component of the substituted pyrophosphate group is readily and quantitatively cleaved at the phosphorus atom on treatment with nucleophiles via a nucleophilic substitution mechanism. The substitution is accompanied by the formation of a covalent bond between the fragment of the oligodeoxyribonucleotide, bound to the disubstituted phosphate group, and the nucleophile.



NuH = nucleophile,
RO and R'O = fragments of oligodeoxyribonucleotides.

It was shown that the functional groups of the side chains of glutamic acid and arginine may also serve as nucleophiles.[79]

The DNA duplexes containing a substituted pyrophosphate group have been used successfully for the affinity modification of the restriction–modification fragments of *Rsr*I and *Eco*RI[80] and of the restriction endonuclease *Eco*RII.[79] These enzymes 'recognise' definite sites in DNA and hydrolyse (restriction endonucleases) or methylate (DNA methyltransferases) them in strictly defined positions. The specificity of their covalent attachment to DNA duplexes with a monosubstituted pyrophosphate internucleotide bond has been demonstrated.[79,80]

The advantages of reagents of this type consist in the fact that the substituted pyrophosphate group does not introduce serious changes into the structure of the DNA duplexes; the chemical reaction between the nucleophilic group of the protein and the reactive group of the substrate proceeds at a 'zero distance' between them and, in contrast to photoactive reagents, does not require external activation (for example UV or γ-irradiation). This precludes the possibility of the distortion of the real pattern of the protein–nucleic acid contacts in the interpretation of the results. The disadvantages of the use of DNA duplexes containing a substituted pyrophosphate internucleotide group for the affinity modification of proteins include the need for a close proximity of the nucleophilic group of the protein and the activated group of the nucleic acid and also the lability of the bond formed between the 5'-phosphate group of the oligodeoxyribonucleotide and certain aminoacid residues.

IV. Covalent attachment of nucleic acids to enzymes based on the mode of enzyme action

In the course of the interaction of an enzyme with a nucleic acid, there is a possibility of the formation of a covalently linked protein–nucleic acid intermediate, which then decomposes with formation of the reaction products. In order to achieve the covalent attachment of the protein to the nucleic acid for the study of the structure of the complex by X-ray diffraction or for the determination of the aminoacid interacting with the substrate, it is sufficient to fix this complex at the stage involving the formation of a covalent compound between the enzyme and the nucleic acid.

One of the examples of the use of an enzymic reaction for the preparation of covalent adducts between enzymes and nucleic acid substrate analogues (substrates) is the study of thymidylate synthetase in the interaction with 5-fluoro-2'-deoxyuridine monophosphate (fl^{5}dUMP). Thymidylate synthetase is an enzyme which transfers a methyl group from 5,10-methylenehydrofolate to uridine with formation of thymidine. The formation of a noncovalent ternary complex, which is converted into a covalent adduct between the protein and the uridine, has been postulated in the interaction of the protein with two substrates. The thiol group of the enzyme attacks the 6-position of the uracil residue in such a way that the C(5) atom can react as a carbanion with methylenetetrahydrofolate.[81] After the transfer of the methyl group from

methylenehydrofolate to the 5-position of uridine, the protein group is eliminated from the 6-position simultaneously with a proton from the 5-position. If a fluorine atom is introduced in place of the hydrogen atom removed, the reaction is blocked at the protein elimination stage.[82]

The decrease in the growth of lymphocyte cells in a medium containing fl^{5}dUMP is also evidence in support of the formation of a covalent adduct, the quantitative characteristics of the process being the same as those of the inactivation of thymidylate synthetase, i.e. the inhibition of cell growth can be accounted for by the interaction of the modified nucleotide with this enzyme.[82]

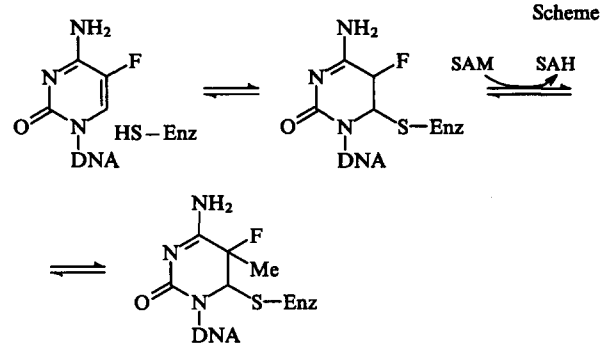
The use of the covalent attachment method is of interest because it makes it possible to determine exactly the stage in which the complex is fixed and to discover the role of the covalently bound aminoacid residue. Under these conditions, the modified nucleotide undergoes the same conformational changes as the substrate. Hence the application of reagents of this kind opens a path to hitherto uninvestigated stages of the catalysis.[83] Thus studies have been made[84,85] of the stereochemical mechanism of the methylation of deoxyuridylate in relation to the covalent ternary complex of thymidylate synthetase with fl^{5}dUMP and 5,10-methylenetetrahydrofolate or 10-propargyl-5,8-dideazafolate. The high stability of the complexes made it possible to obtain crystals and to investigate their structures by X-ray diffraction. It was shown that, in the formation of the ternary complex, thymidylate synthetase undergoes conformational rearrangements involving four aminoacid residues at the C-end of the protein, which are in immediate proximity to the aromatic ring of the fl^{5}dUMP inhibitor.[85] It has been suggested that before elimination the reaction product undergoes a conformational isomerisation and the pathways leading to its occurrence have been discussed.[84]

The enzymes forming covalent conjugates with the substrate include also DNA [C(5)-cytosine]methyltransferases. It has been virtually demonstrated that they all function via a universal mechanism, according to which in the first stage the nucleophilic aminoacid is attached to the 6-position in the cytosine being modified. In the second stage, the methyl group is transferred from the cofactor SAM to the C(5) carbon atom of cytosine. The introduction of a substituent, which is bound to the carbon atom via a stronger bond than hydrogen, in the 5-position of the base being modified fixes the protein–nucleic acid complex at the covalent adduct stage. It has been suggested that a fluorine atom be used as a substituent of this kind.[88]

The interaction of the modification enzyme *Eco*RII with a DNA duplex containing an azacytidine residue has been investigated.[86] The formation of the adduct was detected electrophoretically. It was shown that the introduction of azacytidine into the cell system leads to a sharp increase in the expression of the given gene. Bearing in mind the inhibiting effect of this reagent on the methylase *Eco*RII, this fact can be regarded as evidence for the negative regulation by this enzyme of the expression of its gene.[87]

DNA duplexes containing the 5-fluoro-2′-deoxycytidine (fl^{5}dC) residue are used for the determination of the aminoacid residues participating in the catalytic step in the methylation of DNA by DNA [C(5)-cytosine]methyltransferases. The conjugate of the methylase *Hae*III with a DNA duplex containing fl^{5}dC has been isolated. The substrate in the conjugate was methylated at the C(5) atom of the heterocyclic base. Owing to the presence of fluorine in the 5-position of cytosine, the protein was not split off from the C(6) atom.[88] Analogous conjugates of DNA duplexes containing fl^{5}dC with the enzymes *Eco*RII[89,90], *Hha*I,[91] and *dcm*-methyltransferase of *E. coli* K-12[92] have been isolated. The presence of a covalent complex was demonstrated electrophoretically and detected by autoradiography. The stability of the adduct in denaturing solutions — 6 M urea and 1% solution of sodium dodecyl sulfate (SDS) — was tested. After the denaturation of the protein, the bond between the modified DNA and the methylase in the presence of the cofactor was retained. Kinetic study of the reaction showed that both the enzymic methylation process and the reaction involving the covalent attachment of the protein to DNA are first-order processes. The results confirm the reaction mechanism proposed previously (Scheme 8).

Scheme 8



SAM = *S*-adenosyl-L-methionine,
SAH = *S*-adenosyl-L-homocysteine,
Enz = enzyme.

Analysis of the adducts obtained showed that in all cases the Cys residue forms a covalent bond with the carbon atom in the 6-position in cytosine.[89,90,92] Further studies on these lines lead to the possibility of a detailed description of the stereochemical pathway of the methyl group transfer reaction.[91]

The universality, postulated for DNA [C(5)-cytosine]methyltransferases, extends also to the methyltransferases which interact with RNA. tRNA (m^{5}U54)methyltransferase was investigated by Gu and Santi.[93] A covalent adduct of this enzyme with tRNA, containing the fl^{5}dU residue, was obtained. It proved possible to detect a covalent complex between tRNA (m^{5}U54)methyltransferase and RNA, which was stable when treated on a nitrocellulose filter and in electrophoresis under denaturing conditions. It was shown that such adducts are capable of participating in a further methylation reaction and that their formation reaction is reversible.[94]

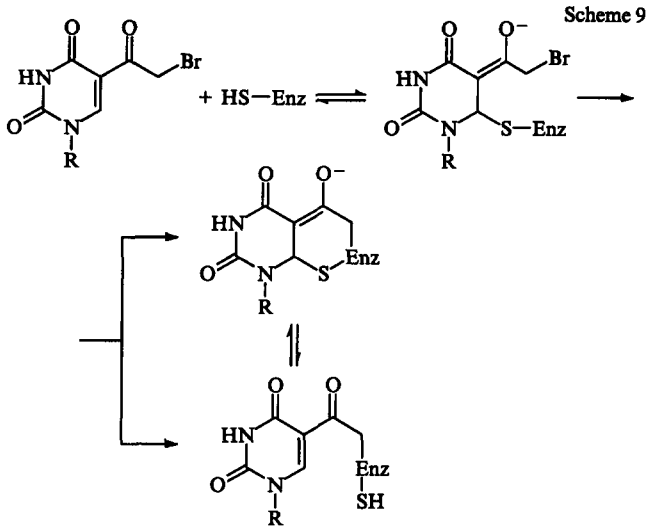
The preparation of intermediate covalent complexes between topoisomerases and DNA has been described. Topoisomerases are enzymes which alter the number of superturns in cyclic covalently closed DNA, initiating single-stand breaks in DNA.[95] The topoisomerase of the type I cowpox virus binds specifically to DNA and forms covalent adducts by cleaving the phosphodiester linkage and forming a new bond with the terminal phosphate group produced at the break point. The enzyme recognises in DNA the conservative sequence 5′-(C/T)CCTT-3′. After the uncoiling of the DNA strand, the enzyme transfers the end of the DNA linked to it to the 5′-terminal region of the same strand with regeneration of the covalently closed plasmid ring. The presence of covalent conjugates at an intermediate stage has been demonstrated by transferring the oligonucleotide residue from the complex to the 5′-end of the oligonucleotide and also by blocking the enzymic reaction by an antibiotic inhibiting the activity of topoisomerases, as a result of which the DNA strand remains broken.[96]

It has been suggested[97] that irreversible inhibitors of the enzymic reaction be employed. Thymidylate synthetase was inhibited by an analogue of uridine triphosphate containing in the 5-position one of the reactive groups: $COCH_2Br$, $CH_2NHCOCH_2I$, or NO_2. Such compounds convert reversible enzyme–inhibitor complexes into irreversibly covalently bound complexes. The mechanism of their activity has not been determined accurately, but a possible pathway has been proposed for this process (Scheme 9).

By employing the reagents considered, it is also possible to probe the groups of atoms in the protein which participate in the methylation of the heterocyclic base.

An interesting method involving the employment of tRNA, containing the br^{5}U residue, for covalent attachment to amino-

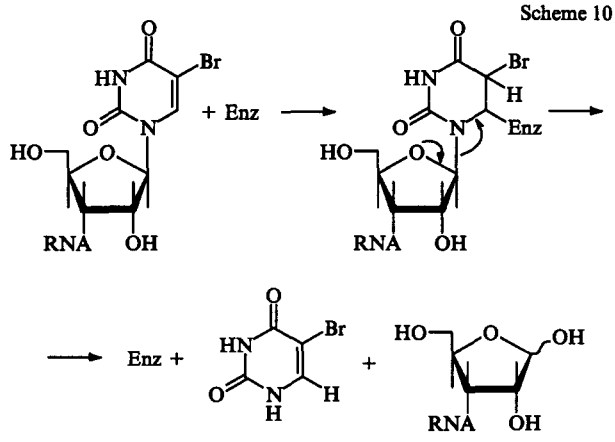
Scheme 9



R = ribose triphosphate residue

acyl-tRNA synthetase without activation by UV light has been described.[98] Although numerous pieces of evidence for the formation of an intermediate covalent adduct between aminoacyl-tRNA synthetase and the uracil ring exist, such compounds had not been obtained earlier. The introduction of a halogen atom, which is highly electronegative, into the heterocyclic base increases the sensitivity of the carbon atom in the 6-position to nucleophiles. The adducts with halogenated uracil derivatives are therefore capable of further rearrangements, leading to the formation of a covalent enzyme–substrate complex. The development of this process leads to the formation of bromouracil and the appearance of a ribose residue at the site occupied previously by 5-bromouracil and of the free enzyme in accordance with Scheme 10.

Scheme 10



Optimum conditions for the prevention of the decomposition of the covalent complex were selected: the use of a sulfhydryl reducing agent and of a tenfold excess of 5-bromouridine. The success of the new method was demonstrated in relation to alanine aminoacyl-tRNA synthetase. The inactivation of the enzyme by br^{5}U containing tRNA was investigated. It was shown that the elimination of uracil leads to the regeneration of the activity of the protein. The results obtained demonstrate the existence of an intermediate covalent adduct in the functioning of aminoacyl-tRNA synthetases.

V. Methods for the identification and isolation of covalent protein–DNA complexes

For the identification and isolation of covalent protein–nucleic acid adducts, use is made, for example, of the ability of one of the components of the complex to be adsorbed on a carrier and also of the difference between the chemical or physicochemical properties of nucleic acid component of the covalent adduct and of its protein component. Thus, in contrast to the protein, DNA binds to a nitrocellulose filter only under certain conditions. On the other hand, DNA, attached covalently to the protein molecule having an affinity for some kind of polymeric carrier (for example, nitrocellulose), acquires the ability to be precipitated on such filters. The new property can be used to separate the reacted DNA from the free DNA and also to demonstrate that covalent attachment has taken place.

A property of the protein which may be used to identify the covalent adduct is the immune response, i.e. the formation of a specific complex of the given protein with an antibody. In this case, the immunochemical approach may be used to isolate protein–DNA conjugates and for their purification.

The protein bound to the nucleic acid also acquires new features, in its turn, characteristic of the nucleic acid component. One of the applications of this phenomenon is the sedimentation method for the isolation of the attached protein in the case where it is fairly small and does not alter the sedimentation constant of the free nucleic acid in a density gradient. When cells or cell nuclei are irradiated in vivo, this property proves very useful for the isolation of conjugates from complex cell systems. For example, the first stage in the purification of covalent adducts of chromosomal proteins with DNA from a cell culture of *D. melanogaster* tissues has been centrifugation in a CsCl gradient in order to isolate DNA-containing cell fractions.[4] Gel filtration can also be used for such purposes.

Extraction methods for the isolation of conjugates are associated with the solubility of one of the components in an organic substance, for example of the protein in phenol. The solubility of the protein component in an organic medium promotes the transfer of the covalently bound nucleic acid into the organic phase, whereas the unreacted nucleic acid remains in the aqueous solution.[21]

Examples of the application of methods most frequently used in practical investigations are presented below.

1. Binding to a polymeric carrier

In order to isolate covalent complexes of the DNA adenine methyltransferase (Dam methylase) from *E. coli* with SAM, the protein component of the conjugate was precipitated on a glass filter with 10% trifluoroacetic acid. The presence of a covalent bond between the methyltransferase and the cofactor was demonstrated with the aid of a radioactive label (^{3}H or ^{14}C) in SAM. As a control, a similar procedure was carried out with a noncovalent complex between the methyltransferase and SAM.[18]

The covalent adducts of a DNA-recognising protein from the type 2 adenovirus with ^{32}P-p(dT)$_{14}$ were filtered through nitrocellulose filters and washed with solutions containing NaCl or SDS in order to decompose noncovalent specific complexes. The yield of the covalent protein–nucleic acid complex was determined from the radioactivity of the residue on the filter. The same approach (binding on a nitrocellulose filter) was used to purify the covalent adducts of tRNA (m^{5}U54)methyltransferase and RNA.[93] Toyo Roshi 51 paper filters were used in another study[16] instead of the nitrocellulose filters.

In order to isolate the proteins from the nuclear extract of mouse hybridoma B-cells, participating in the specific recognition of definite DNA sequences, their complexes with oligonucleotides, incorporating not only the binding site but also a 10-membered dA sequence at the 3′-end, were irradiated with UV light. The covalent adducts obtained were isolated on oligo-dT-cellulose. In this case, the elimination of 'unattached' proteins

from the complex was achieved by the formation of an oligo-dA · oligo-dT duplex on a polymer matrix.[99]

Yet another example of the use of a polymeric carrier for the isolation of protein–nucleic acid conjugates is the hybridisation of the products of the treatment with formaldehyde of *E. coli* cells with single-stranded DNA immobilised on cellulose and complementary to the pBR322 plasmid. The conjugates of mRNA and pBR322 DNA with proteins are then separated from the remaining cell material. After being split off from the covalent RNA complex, the protein components were analysed by electrophoretic methods (Laemmli's methods).[59]

2. Electrophoretic isolation methods

Electrophoresis is the most effective method for the isolation of covalently bound protein–nucleic acid conjugates. In the electrophoretic methods of analysis, there is a possibility of the simultaneous determination of both the protein component (by staining with an appropriate dye) and of the nucleic acid (from the ^{32}P radioactivity). There are two versions of this method differing in the presence or absence of a denaturing agent in the supporting medium.

Thus the isolation of covalent protein–nucleic acid complexes by the 'retardation' method in 1% agarose gel not containing denaturing agents has been described.[100] Covalent adducts of *E. coli* topoisomerase with an oligonucleotide were also identified in a nondenaturing 4.5% polyacrylamide gel.[96]

Two-dimensional electrophoresis in a nondenaturing medium has a greater resolving power. In this case, the reaction mixture is separated initially in one direction, after which one of the components of the protein–nucleic acid complex is subjected to degradation by a chemical or enzymic method and electrophoresis is carried out in the perpendicular direction.[21] DNA-binding proteins of the cowpox virus, attached covalently to the viral DNA, were identified in this way. The electrophoretic separation in the first direction was carried out in a polyacrylamide gel (the linear acrylamide concentration gradient was from 5% to 15%). In the second direction, the concentration of acrylamide was constant at 12.5%. The proteins were detected from the radioactivity of the ^{125}I introduced with the aid of the Bolton–Hunter reagent. The set of marker proteins, used in the electrophoresis in the second direction, made it possible to determine the molecular masses of the attached proteins.

However, the use of nondenaturing gels does not usually afford the complete certainty that the protein–nucleic acid complex is indeed covalent. Denaturing agents, for example SDS or urea, are therefore introduced into the gel. Gel electrophoresis in Laemmli's version with SDS as the denaturing agent is usually employed. The gel concentration varied in different investigations from 10% to 15%.[10,17,19,25,29,33,34,87,101,102] Preliminary treatment of samples from the SDS procedure with 2-mercaptoethanol, including maintenance for several minutes at 95 °C, guaranteed the complete decomposition of the noncovalent protein–nucleic acid complexes and also the denaturation of the protein. The attachment of nucleic acid fragments to the protein has only a slight influence on the electrophoretic mobility of the latter, because it is abolished by the large negative charge on SDS, while the increase in the mass of the protein after the attachment of the nucleotide or oligonucleotide to the latter is negligible compared with its intrinsic mass. After the electrophoretic separation, the gels are usually stained with a Coomassie dye in order to determine the protein bands and are then dried and autoradiography is carried out in order to visualise the nucleic acid fragments. In the study of Pelle and Murphy,[5] before determining the positions of the complexes in the gel the latter were transferred by electroblotting on to a nitrocellulose membrane.

After the irradiation of the complexes of a polypeptide present in the transcription complex of the pea and DNA containing br^{5}dU the samples were treated with DNase I (in certain experiments an additional treatment with snake venom phosphodiesterase was performed).[28] After this, they were analysed by electrophoresis in a dilute (5%) polyacrylamide gel.

Nucleic acid was decomposed similarly by micrococcal nuclease and bacterial alkaline phosphatase after the covalent attachment of DNA to histones. A ^{32}P label was then introduced into the residual oligonucleotide with the aid of polynucleotide kinase (before deposition on a 15% polyacrylamide denaturing gel).[69] The need for the degradation of DNA is due to its large size and hence low electrophoretic mobility.

The protein component of the complex with nucleic acid can also be subjected to partial cleavage. *N*-Chlorosuccinimide is used for this purpose. After the treatment of the complexes with *N*-chlorosuccinimide, the reaction mixture is deposited on a 10% SDS-containing polyacrylamide gel. The positions of the bands of the partly degraded protein–nucleic acid complex are determined autoradiographically. The set of products of the partial decomposition of such compounds by *N*-chlorosuccinimide is specific to each protein and may be used for the rapid determination of the macromolecules recognising a definite sequence in DNA or RNA.[103]

After covalent attachment to DNA, GCN4 activator of the transcription from yeast was analysed electrophoretically in a gel with another type of denaturing agent — urea. For the denaturation of the 'unattached' complexes, the reaction mixtures were treated with alkali and deposited on a 10% polyacrylamide gel containing 8 M urea. The covalent protein–nucleic acid complex was detected autoradiographically.[30]

3. Chromatographic isolation methods

Chromatographic methods are more rarely used for the analysis of covalently bound protein–DNA complexes because large amounts of substances are required for the spectrophotometric detection of the reaction products. Nevertheless, for example, in the study of Barbier et al.,[26] a mixture of the *lac* repressor with br^{5}dU-containing DNA was separated chromatographically on a Sephadex column after irradiation with UV light. When a 200 mM potassium phosphate buffer containing 500 mM KCl was used, the unbound repressor separated from the covalent complex, which was indicated by a well resolved peak. The separation process was followed spectrophotometrically from the absorption at 260 and 280 nm.

The covalent complex of the T4 endonuclease V with an oligonucleotide containing a phenyldiazirine derivative has been isolated on a phosphocellulose column.[39]

The method for the isolation of the covalent complexes of the methyltransferase *Hae*I with an oligonucleotide containing the fl^{5}dC residue by fast protein liquid chromatography (FPLC) on a MonoQ column in a linear NaCl concentration gradient from 0 to 1 M proved effective.[88]

4. Immunochemical methods for the isolation of protein–nucleic acid covalent complexes

Several examples of the use of immunochemical methods for the isolation of protein–nucleic acid covalent complexes have been described in the literature.

The replication protein A was irradiated together with oligonucleotides. After the reaction, the protein and the covalent protein–nucleic acid complex were separated from the unreacted oligonucleotides with the aid of monoclonal antibodies. The reaction mixture was then analysed by the retardation method in a gel.[13]

The proteins from the extract of cell cytoplasm, attached covalently in vitro to nucleic acids, were isolated with the aid of antibodies obtained by the vaccination of mice.[5]

After the covalent attachment of histones to DNA, high-molecular-mass complexes were isolated with the aid of antibodies specific to the acetylated histones.[69]

The protein component of the covalent adduct was determined by the immune precipitation of one of the proteins of the transcription factor II (TBP subunit) after its covalent attachment to DNA as a result of irradiation with UV light.[42]

VI. Determination of peptides in a protein bound covalently to nucleic acid

The final aim of the covalent attachment of a protein to nucleic acid is the determination of the aminoacids or domains of the protein coming into contact with nucleic acids. To do this, the protein is fragmented after its covalent attachment and its component forming a covalent adduct with the nucleic acid is determined. Numerous methods, both chemical and enzymic, are employed for this purpose.[104] Depending on the conditions, mild acid hydrolysis leads to the formation of large or small protein fragments, but its effect is as a rule nonspecific in relation to particular aminoacid residues and the set of peptides obtained is not reproduced when the hydrolysis of the same protein is repeated. The secondary and tertiary structures of the protein exert some influence on this process. There are also reagents which cleave the peptide bond at a particular aminoacid when the latter is C- or N-terminal. For example, BrCN hydrolyses a specific bond in proteins when methionine is NH_2-terminal.[105, 106] Bromocyanogen has also been used to determine the DNA-binding domains of the restriction enzymes *Eco*RI and *Eco*RV after their irradiation with UV light in a complex with DNA duplexes containing br^{5}dU residues. Since methionine is rarely encountered in proteins, this made it possible to locate only those regions in the enzymes to which DNA is bound.[25] A disadvantage of the method is the severe conditions in the treatment (80% formic acid), under which only the C–C bonds formed on irradiation of protein–nucleic acid complexes with UV light are stable.

The enzymic fragmentation of the protein component of the covalent adduct with the nucleic acid proceeds under milder conditions and is fairly specific when particular proteinases are employed. Rarely and frequently cleaving proteinases are distinguished, which is determined primarily by their specificity and also by the frequency with which certain aminoacids are encountered in proteins. For example, the endopeptidase Lys-C cleaves proteins into fairly large fragments, which are easier to analyse. However, when this enzyme is employed, it is not possible to locate accurately the active region of the protein. Trypsin is the most frequently employed enzyme. It cleaves the protein specifically at the peptide bonds formed by the carboxy-groups of lysine or arginine.[104] Since the range of aminoacids recognised by them is greater than for the endopeptidase Lys-C, the size of the peptides formed on cleavage is smaller and their number is greater. This complicates the isolation of the peptide bound covalently to the nucleic acid, on the one hand, but it permits a more accurate location of the attached aminoacid, on the other.

When covalent protein–nucleic acid adducts are cleaved by proteolytic enzymes, a large set of peptides is formed, including the protein fragments unattached to the nucleic acid, one or several peptides bound covalently to the nucleic acid, and the products of the hydrolysis of the proteinase itself, because it too may hydrolyse its own peptide bonds (autoproteolysis). In order to avoid autoproteolysis, specially modified trypsin, which hardly undergoes any self-decomposition, is used. For a more complete and exhaustive hydrolysis, the protein is also denatured by 8 M urea, but, since this may prove insufficient, it is treated further with iodoacetamide.[107] After treatment with iodoacetamide, the concentration of urea may be reduced to 2 M, whereupon the effectiveness of the cleavage by trypsin increases sharply, which makes it possible to reduce the trypsin : protein mass ratio to 1 : 25. For this ratio, the autoproteolysis of trypsin is no longer a problem.[108] Table 3 presents the most frequently used methods for the fragmentation of proteins in covalent protein–nucleic acid complexes. In each specific case, the choice of the proteolysis conditions is determined by the effectiveness of the operation of the proteinase (or the chemical reagent) and by the stability of the protein–nucleic acid conjugates. Since the nucleic acid may protect the protein against proteolysis, an individual selection of the protein : proteinase (or reagent) ratio and of the method for

Table 3. Methods for the cleavage of the protein component of the protein–nucleic acid conjugates.

Enzyme or reagent	Enzyme (reagent) concentration or enzyme : substrate mass ratio	Buffer solution for proteolysis	Time /h	Incubation temperature /°C	Ref.
Trypsin	1 : 3.7	83.3 mM tris-HCl, pH 8.0, 16 mM $CaCl_2$, 2 M urea	72	37	51
	1 : 10	2 M urea	4	37	34
	1 : 5000, 1 : 500, 1 : 100	25 mM tris-HCl, pH 8.0, 1 mM EDTA, 0.1 M NaCl	0.3	20	29
	1 : 25	10 mM tris-HCl, pH 7.5, 4.8 M urea, 0.1 M NaCl	18	25	31
TPCK-trypsin	1 : 20	50 mM NH_4HCO_3, pH 8.0, 2 M urea	12 (or 48)	37 (or 20)	17
Chymotrypsin	1 : 50, 1 : 100, 1 : 200	25 mM tris-HCl, pH 8.0, 1 mM EDTA, 0.1 M NaCl	0.3	20	14
	1 : 100 [a]	100 mM NH_4HCO_3	16	37	16
Staphylococcal proteinase V8	1 : 1, 1 : 25	25 mM tris-HCl, pH 8.0, 1 mM EDTA, 0.1 M NaCl	1	20	14
	100 μg ml^{-1} [a]	50 mM NH_4HCO_3	16	37	89
	36 μg ml^{-1}	50 mM NH_4HCO_3, pH 8.0, 0.1 % SDS	12	37	17
Lysine endopeptidase	5 pmol	10 mM K_3PO_4, pH 6.5, 10% ethylene glycol, 10 mM mercaptoethanol, 2 mM EDTA	4	30	39
o-Iodobenzoic acid	15 mg ml^{-1} [b]	80 % AcOH, 4 M guanidine, 1% cresol	24	20	17

[a] Preliminary treatment with iodoacetamide. [b] Preliminary treatment with 4-vinylpyridine.

the denaturation of the protein–nucleic acid complex (urea or SDS) is necessary.

After the successful fragmentation of the proteins in their covalent complexes with nucleic acids a new problem arises — the isolation of one or several peptides which were involved in the formation of the covalent adduct with the nucleic acid. Several approaches to the solution of this problem have been described in the literature: the use of electrophoretic methods, extraction into organic solvents, and a different kind of chromatographic purification. Conditions for electrophoresis designed to separate peptides with molecular masses ranging from 1 to 5 kDa have been selected.[109] Phenol extraction of peptides is also used. One purification stage is usually insufficient for the separation of the target compound from other substances, so that column chromatography is almost always used at one of the purification stages.[108, 109]

Table 4 presents the chromatographic conditions used for the isolation and purification of peptides bound covalently to nucleic acids. For the selection of the column, sorbent, eluent, and other chromatographic separation conditions, the investigator may use the literature data (Table 4) taking into account the individual features of the conjugate investigated.

After isolation and purification, the aminoacid sequence of the peptide is determined.[108, 110]

Cases are possible where the aim — the determination of the peptide bound covalently to DNA — proves to be unattainable because of the instability of the adduct, insufficient amounts of the protein, and the lack of the necessary equipment. Indirect data are then employed to find the aminoacid cross-linked to the nucleic acid. These include the determination of the nature of the bond between the nucleic acid and the protein. Such determination is based on the difference between the reactivities of the functional groups and their hydrolytic stabilities.

If the nucleic acid is attached to the protein via the terminal phosphate group, then, owing to the instability of the covalent complex, problems involving the determination of the peptide usually arise. The phosphodiester or phosphotriester bond formed as a result of the attachment of the protein to hydroxyaminoacids

Table 4. Conditions for the isolation by chromatographic methods of peptides (products of the proteolysis of protein–nucleic acid conjugates) attached covalently to nucleic acids (the separation temperature is not indicated).

Column	Initial eluent	Gradient[b]	Separation time /min	Consumption of eluent /ml min^{-1}	Detection method[c]	Ref.
Delta-Pak C_{18}; 7.8×300 mm[a]	10 mM CH_3COONH_4, pH 6.5	70% CH_3CN	60	1	SP (214)	18
Delta-Pak C_{18}; 2×150 mm[a]	1.16% CF_3COOH	50% CH_3CN	60	0.2	SP (214)	18
NAP-10	50 mM NH_4HCO_3	50mM NH_4HCO_3	not indicated	not indicated	SP	34
TSK ODS 120T; 4.6×250 mm	0.05% CF_3COOH	48% CH_3CN	50	0.8	SP (220)	16
	5 mM NH_4HCO_3	80% CH_3CN	50	not indicated	SP (220), RA	16
SynChropak AX300 4.1×250 mm	20 mM K_3PO_4, pH 6.8, 5% C_2H_5OH, 50 mM KCl	1 M KCl	90	not indicated	SP (254), RA	14
Waters C_{18}	100 mM $(C_2H_5)_3N \cdot CH_3COOH$, pH 6.8	60% CH_3CN	50	not indicated	SP (254), RA	14
Aquapore RP-300; 4.6×250 mm	0.1% CF_3COOH	50% Pr^iOH	20	1	SP (220)	39
Beckmann ODS; 4.3×250 mm	50 mM $(C_2H_5)_3N \cdot CH_3COOH$	70% CH_3CN	45	0.5	SP (222), RA	89
Hydropore-SAX; 4.6×100 mm	50 mM tris-HCl, pH 7.5, 0.5 mM EDTA, 10 mM mercaptoethanol	1 M NaCl	60	1	RA (^3H and ^{32}P)	51
μBondapak C_{18}	10 mM CH_3COONH_4, pH 6.0	80% CH_3CN	60	1	SP (214), RA	17
	10 mM CH_3COONH_4, pH 6.0, 15% CH_3CN	45% CH_3CN	60	not indicated	SP (214), RA	17

[a] Separation temperature 40 °C. [b] Only the gradient-forming component, the concentration of which in the initial eluent increases to the value indicated, is quoted. [c] SP = spectrophotometric method (the wavelength in nm is indicated in brackets); RA = detection based on radioactivity.

is labile under alkaline conditions, the stability of the bond being influenced by the nature of the aminoacid. Thus the adducts with hydroxyaminoacids are more sensitive to alkaline hydrolysis than the adducts with tyrosine. By measuring the dependence of the stability of the covalent protein–nucleic acid complex on the pH, it is possible to reach a conclusion about the nature of the aminoacid bound to the nucleic acid. The phosphodiester nature of the bond may be demonstrated with the aid of snake venom and calf spleen phosphodiesterases, which are able to hydrolyse such bonds.[111]

The nature of the bond between the modification enzyme *Eco*RI and an oligonucleotide has been determined.[80] The conjugate proved stable in 0.1 M sodium hydroxide solution, but was hydrolysed in 0.1 M hydrochloric acid or in 4 M solution of hydroxylamine at pH 5. This suggested the formation of a covalent bond between the phosphate group of the oligonucleotide and the amino-group of Arg or Lys or the imino-group of His.

VII. Probing the DNA recognising proteins mode of action via cross-linking assay

The covalent attachment method makes it possible to investigate indirectly the protein–nucleic acid interaction because a specific complex is fixed at the binding stage. By varying the parameters of the cross-linking, it is possible to fix different recognition stages.

The complex of the restriction endonuclease *Nae*I and an oligodeoxyribonucleotide duplex containing the br^{5}dU residue has been irradiated with UV light in order to determine the number of enzyme subunits and substrate molecules in the catalytically active enzyme–substrate complex.[112] On photoactivation of the substrate in the complex with the protein, a covalent complex was produced. An increase in the reaction time led to the appearance of yet another conjugate. The mobility of the new covalent complex in the polyacrylamide gel corresponds to a macromolecule with a molecular mass approximately twice as great as that of a single *Nae*I subunit. The covalent attachment after additional irradiation time probably resulted in the attachment of a second enzyme molecule to the substrate. This finding constitutes yet another confirmation of the fact that the restriction endonuclease *Nae*I binds to two substrate molecules in the form of a dimer.

The covalent attachment of nucleic acids to proteins on exposure to laser radiation provides even greater possibilities for the indirect study of the mechanism of the protein–nucleic acid interaction. Because of the very short duration of the pulse, it is possible to determine the physicochemical parameters of the protein–nucleic acid recognition — the binding constant.[101] The method based on the 'freezing' of the equilibrium states has been used[101] for the protein coded for by the 32 gene of the bacteriophage T4 (gp32). This protein binds preferentially to single-stranded DNA. Since during the duration of the pulse there is insufficient time for the protein to pass from one conformational state to another, the yield of the covalent attachment is proportional to the amount of the protein–nucleic acid complexes existing at the given instant, which is determined by the complex formation constant. The dependence of the yield of the product of the covalent attachment on the formation constant of a complex between gp32 and oligo(dT) (known from the literature) was found and the equilibrium constants for other analogous systems were calculated (for example, for the equilibria between gp32 and shorter oligonucleotides, for the same complexes in other buffer solutions, and for the analogous proteins of the same system of the bacteriophage T4). Since the yield of the product of the covalent attachment of gp32 to $(dT)_n$ increases in jumps as a function of the length of the oligomer (up to $n = 7$, there is virtually no covalent attachment of gp32 to the oligodeoxyribonucleotide, while at $n = 7$ the yield of the adduct increases sharply), it has been suggested that the minimum area for the deposition of the protein gp32 on DNA corresponds to seven nucleotide residues. The repeated increase in the yield of the covalent adducts for n which is a multiple of seven confirms this hypothesis, since the deposition of two gp32 molecules on the oligodeoxyribonucleotide now becomes possible.[101]

Hockensmith et al.[102] chose as the objects of study the proteins coded for by the genes 44 (gp44), 45 (gp45), and 62 (gp62) of the bacteriophage T4, which require nucleotide cofactors as well as Mg^{2+} ions. In the absence of gp45 and also at its low concentrations, DNA forms preferentially covalent adducts with gp44, but, starting from the equimolecular ratio of gp45 and the complex gp44/gp62, the effectiveness of the affinity modification of gp62 increases. Since the protein gp45 has a DNA-dependent ATPase activity, the influence of gp62 on the kinetic parameters of the functioning of gp45 was demonstrated. The nature of the dependence of the covalent attachment of gp44 and gp62 on the gp45 concentration indicates a conformational rearrangement of the DNA-recognising system of the bacteriophage T4 for the equimolecular ratios of all the components. Knowing the yields of the products of the covalent attachment of the proteins considered to $(dT)_{16}$, $(dT)_{20}$, $(dT)_{16}\cdot(dA)_{20}$, $(dA)_{16}\cdot(dT)_{20}$, and $(dT)_{20}\cdot(dA)_{20}$, simulating single-stranded DNA, the complexes of the primer and the template, and DNA duplexes with 'blunt' ends, it is possible to determine the disposition of gp32, gp45, and gp62 on the primer, on the template, or on the primer–template complex.

Summarising the foregoing, one should note that a change in the effectiveness of the affinity modification of nucleic acids by proteins reflects conformational transitions and also more complex multisubunit interactions of the latter.[101]

VIII. Conclusion

Thus the application of the affinity modification method leads to extensive possibilities for the analysis of the structures and mechanisms of the functioning of the proteins interacting with nucleic acids. The principal requirements which must be met by the procedure involving the covalent attachment of a protein to nucleic acids is the creation of optimum conditions for the protein–nucleic acid recognition. The modifications carried out must not distort seriously the structures of the components of the complexes and hence affect the process involving the recognition of nucleic acid analogues by nucleic acid-specific proteins. In our view, the most promising in this sense are the bromo- and azido-substituted nucleic acid analogues as well as nucleic acids containing an activated internucleotide phosphate group. The moderate reactivity of such compounds ensures the specificity of the covalent attachment, since the reaction with the protein becomes possible by virtue of the mutual approach of the reacting centres of both components of the complex.

References

1. T M Cao, M T Sung *Biochemistry* **21** 3419 (1982)
2. G R Kunkel, U G Martinson *Nucl. Acids Res.* **5** 4263 (1978)
3. I Saito, H Sugiyma, T Matsuura *J. Am. Chem. Soc.* **105** 6989 (1983)
4. S V Belikov, A I Belgovsky, O V Preobrazhenskaya, V L Karpov, A D Mirzabekov *Nucl. Acids Res.* **21** 1031 (1993)
5. R Pelle, N B Murphy *Nucl. Acids Res.* **21** 2453 (1993)
6. D S Gilmour, J T Lis *Proc. Natl. Acad. Sci. USA* **81** 4275 (1984)
7. Z Hillel, C-W Wu *Biochemistry* **17** 2954 (1978)
8. G F Strniste, D A Smith *Biochemistry* **13** 485 (1974)
9. M Buck, W Cannon *Nucl. Acids Res.* **22** 1119 (1994)
10. R Mital, Y Albrecht, D Schumperli *Nucl. Acids Res.* **21** 1049 (1993)
11. C S Park, Z Hillel, C-W Wu *Nucl. Acids Res.* **8** 5895 (1980)
12. V T Yue, P R Schimmel *Biochemistry* **16** 4678 (1977)
13. E Seroussi, S Lavi *J. Biol. Chem.* **268** 7147 (1993)
14. V Cleghon, D F Klessig *J. Biol. Chem.* **267** 17872 (1992)
15. A Aharoni, N Baran, H Manor *Nucl. Acids Res.* **21** 5221 (1993)
16. Y Takata, M Fujioka *Biochemistry* **31** 4369 (1992)
17. K Subbaramaiah, S A Simms *J. Biol. Chem.* **267** 8636 (1992)
18. C Wenzel, W Guschlbauer *Nucl. Acids Res.* **21** 4604 (1993)

19. C Wenzel, M Moulard, A Løner-Olesen, W Guschlbauer *FEBS Lett.* **280** 147 (1994)
20. C Taylor, K Ford, B A Conolly, D P Hornby *Biochem. J.* **291** 493 (1993)
21. J Welsh, C R Cantor *Trends. Biochem. Sci.* **9** 505 (1984)
22. A J Varghese *Aging Carcinogenesis Radiat. Biol.* 207 (1976)
23. A Havron, J Sperling *Biochemistry* **16** 5651 (1977)
24. S Maruta, M Burke, M Ikebe *Biochemistry* **29** 9910 (1990)
25. H Wolfes, B A Fleiss, F Winkler, A Pingoud *Eur. J. Biochem.* **159** 267 (1986)
26. B Barbier, M Charlier, J-C Maurizot *Biochemistry* **23** 2933 (1984)
27. R Ogata, W Gilbert *Proc. Natl. Acad. Sci. USA* **74** 4973 (1977)
28. N C Khanna, S Lakhani, K K Tewari *Nucl. Acids Res.* **20** 69 (1992)
29. J M Gott, M C Willis, T H Koch, O C Uhlenbeck *Biochemistry* **30** 6290 (1991)
30. E E Blatter, Y W Ebright, R E Ebright *Nature (London)* **359** 650 (1992)
31. M Katouzian-Safadi, B Blazy, J-I Cremet, J-P Le Caer, J Rossier, M Charlier *Biochemistry* **32** 1770 (1993)
32. M Katouzian-Safadi, B Laine, F Chartier, J-I Cremet, D Belaiche, P Sautiere, M Charlier *Nucl. Acids Res.* **19** 4937 (1991)
33. T T Nikiforov, B A Conolly *Nucl. Acids Res.* **20** 1209 (1992)
34. Y J K Farrar, R K Evans, C M Beach, M S Coleman *Biochemistry* **30** 3075 (1991)
35. C E Catalano, D J Allen, S J Benkovic *Biochemistry* **29** 3612 (1990)
36. S K Syed, S Kim, W K Paik *Biochemistry* **32** 2242 (1993)
37. G A Young, M Woody, C R Vader, R W Woody, B E Haley *Biochemistry* **23** 2843 (1984)
38. P Mitchell, K Stade, M Osswald, R Brimacombe *Nucl. Acids Res.* **21** 887 (1993)
39. N Hori, S Iwai, H Inone, E Ohtsuka *J. Biol. Chem.* **267** 15591 (1992)
40. R K Evans, J D Johnson, B E Haley *Proc. Natl. Acad. Sci USA* **83** 5382 (1986)
41. N O Reich, E A Everett *J. Biol. Chem.* **265** 8929 (1990)
42. M A Sypes, D S Gilmour *Nucl. Acids Res.* **22** 807 (1994)
43. J Rinke-Appel, N Jünke, R Brimacombe, S Dokudovskaya, O Dontsova, A Bogdanov *Nucl. Acids Res.* **21** 2853 (1993)
44. O Dontsova, V Tishkov, S Dokudovskaya, A Bogdanov, T Döring, J Rinke-Appel, B Greuer, R Brimacombe *Proc. Natl. Acad. Sci. USA* **91** 4125 (1994)
45. O Dontsova, A Kopylov, R Brimacombe *EMBO J.* **10** 2613 (1991)
46. J Rinke-Appel, N Jünke, R Brimacombe, I Lavric, S Dokudovskaya, O Dontsova, A Bogdanov *Nucl. Acids Res.* **22** 3018 (1994)
47. O Dontsova, S Dokudovskaya, A Kopylov, A Bogdanov, J Rinke-Appel, N Jünke, R Brimacombe *EMBO J.* **11** 3105 (1992)
48. S S Dokudovskaya, O A Dontsova, S L Bogdanova, A A Bogdanov, R Brimacombe *Biotechnol. Appl. Biochem.* **18** 149 (1993)
49. G G Karpova, Doctoral Thesis in Chemical Sciences, Moscow State University, Moscow, 1990
50. C A Harrison, D H Turner, D C Hinkle *Nucl. Acids Res.* **10** 2399 (1982)
51. S S Sastry, H P Spielmann, Q S Hoang, A M Phillips, A Sancar, J E Hearst *Biochemistry* **32** 5526 (1993)
52. E Gajewski, M Dizdaroglu *Biochemistry* **29** 977 (1990)
53. M Dizdaroglu, E Gajewski, P Reddy, S A Margolis *Biochemistry* **28** 3625 (1989)
54. L Stryer, in *Biochemistry* Vol. 3 (Translated into Russian; Moscow: Mir, 1985) p. 129
55. C M Arundel, C M Vines, P J Tofilon *Cancer Res.* **48** 5669 (1988)
56. V Jackson *Cell* **15** 945 (1978)
57. G Christiansen, J Griffith *Nucl. Acids Res.* **4** 1837 (1977)
58. K L Brodolin, V M Studitsky, A D Mirzabekov *Nucl. Acids Res.* **21** 5748 (1993)
59. J P Shouten *J. Biol. Chem.* **260** 9929 (1985)
60. J D McGhee, P H von Hippel *Biochemistry* **14** 1281 (1975)
61. J D McGhee, P H von Hippel *Biochemistry* **14** 1297 (1975)
62. M Sugiyama, M Costa, T Nakagawa, T Hidaka, R Ogura *Cells Cancer Res.* **48** 1100 (1988)
63. M D Cohen, C A Miller, L S Xu, E T Snow, M Costa *Anal. Biochem.* **186** 1 (1990)
64. J E Hearst *Annu. Rev. Biophys. Bioeng.* **10** 69 (1981)
65. J W Levinson, L F Leibes, J J McCormic *Biochim. Biophys. Acta,* **447** 260 (1976)
66. E S Levina, S G Bavykin, V V Shick, A D Mirzabekov *Anal. Biochem.* **110** 93 (1981)
67. A D Mirzabekov, S G Bavykin, V L Karpov, O V Preobrazhenskaya, K K Evralidze, V M Tuneev, A F Melinkova, K G Goguadze, A A Chenchick, R S Beabealashvili *Cold Spring Harbor Symp. Quant. Biol.* **47** 503 (1982)
68. A A Chenchick, R S Beabealashvili, A D Mirzabekov *FEBS Lett.* **128** 46 (1981)
69. K K Ebralidse, T R Hebbes, A L Clayton, A W Thorne, C Crane-Robinson *Nucl. Acids Res.* **21** 4734 (1993)
70. A N Chkheidse, G G Prikhod'ko, V L Karpov, S K Vassilenko, A D Mirzabekov *FEBS Lett* **336** 340 (1993)
71. N Nitta, O Kuge, S Yui, A Tsugawa, H Negishi, H Hayatsu *FEBS Lett.* **166** 194 (1984)
72. M N Vinogradova, V L Drutsa, M G Ivanovskaya, E S Gromova, Z A Shabarova *Biokhimiya* **48** 286 (1983)
73. E V Petushkova, V M Risnik, N I Sokolova, S S Tret'yakova, Z A Shabarova *Biokhimiya* **45** 726 (1980)
74. V L Drutsa, Candidate Thesis in Chemical Sciences, Moscow State University, Moscow, 1978
75. V M Kodentsova, E V Petushkova, S S Tretyakova, N I Sokolova, Z A Shabarova *Biochem. Int.* **10** 195 (1985)
76. V M Kodentsova, Candidate Thesis in Biological Sciences, Moscow State University, Moscow, 1982
77. N I Sokolova, S S Tret'yakova *Bioorg. Khim.* **9** 1157 (1983)
78. S A Kuznetsova, M G Ivanovskaya, Z A Shabarova *Bioorg. Khim.* **16** 219 (1990)
79. Z A Shabarova, G Ya Sheflyan, S A Kuznetsova, E A Kubareva, O N Sysoev, M G Ivanovskaya, E S Gromova *Bioorg. Khim.* **20** 413 (1994)
80. A A Purmal, Z A Shabarova, R I Gumport *Nucl. Acids Res.* **20** 3713 (1992)
81. R H Abeles, T A Alston *J. Biol. Chem.* **265** 16705 (1990)
82. S Cox, J Harmenberg *J. Biochem. Biophys. Methods* **25** 17 (1992)
83. K M Ibanetich, D V Santi, in *The Expanding Role of Folates and Fluoropyrimidines in Cancer Chemotherapy* (Eds Y Rustum, J J McGuire) (New York: Plenum, 1988) p. 113
84. D A Matthews, J E Villafranca, C A Janson, W W Smith, K Welsh, S Freer *J. Mol. Biol.* **214** 937 (1990)
85. D A Matthews, K Appelt, S J Oatley, N H Xuong *J. Mol. Biol.* **214** 923 (1990)
86. S Friedman *Nucl. Acids Res.* **14** 4543 (1986)
87. S Som, S Friedman *EMBO J.* **12** 4297 (1993)
88. L Chen, A M McMillan, W Chang, K Ezaz-Nikpay, W S Lane, G L Verdine *Biochemistry* **30** 11018 (1991)
89. S Friedman, N Ansari *Nucl. Acids Res.* **20** 3241 (1992)
90. M W Wyszynski, S Gabbara, E A Kubareva, E A Romanova, T S Oretskaya, E S Gromova, Z A Shabarova, A S Bhagwat *Nucl. Acids Res.* **21** 295 (1993)
91. D S Osterman, G D DePillis, J C Wu, A Matsuda, D V Santi *Biochemistry* **27** 5204 (1988)
92. T Hanck, S Schimdt, H-J Fritz *Nucl. Acids Res.* **21** 303 (1993)
93. X Gu, D V Santi *Biochemistry* **31** 10295 (1992)
94. D V Santi, L W Hardy *Biochemistry* **26** 8599 (1987)
95. F S Spirin (Ed.) *Molekulyarnaya Biologiya. Struktura i Biosintez Nukleinovykh Kislot* (Molecular Biology. Structure and Biosynthesis of Nucleinic Acids) (Moscow: Vysshaya Shkola, 1990)
96. S G Morham, S Shuman *J. Biol. Chem.* **267** 15984 (1992)
97. C B Brouilette, C T-C Chang, M P Mertes *Biochem. Biophys. Res. Commun.* **87** 613 (1979)
98. R M Starzyk, S W Koontz, P Schimmel *Nature (London)* **298** 136 (1982)
99. U Wichser, C Brack *Nucl. Acids Res.* **20** 4103 (1992)
100. J Czichos, M Koehler, B Reckmann, M Renz *Nucl. Acids Res.* **17** 1563 (1989)
101. J W Hockensmith, W L Kubasek, W R Vorachek, P H von Hippel *J. Biol. Chem.* **268** 15712 (1993)
102. J W Hockensmith, W L Kubasek, E M Evertsz, L D Mesner, P H von Hippel *J. Biol. Chem.* **268** 15721 (1993)
103. M Mirfakharai, A M Weiner *Nucl. Acids Res.* **21** 3591 (1993)

104. S M Avaeva *Khimiya Belka* (Protein Chemistry) Part 1 (Moscow: Replication Laboratory, Faculty of Chemistry, Moscow State University, 1983)
105. E Cross *Methods Enzymol.* **11** 238 (1967)
106. W A Schroeder, J B Shelton, J R Shelton *Arch. Biochem. Biophys.* **130** 551 (1969)
107. A M Crestfield, S Moore, W H Stein *J. Biol. Chem.* **238** 622 (1963)
108. P T Matsudaira (Ed.) *A Practical Guide to Protein and Peptide Purification for Microsequencing* (New York: Academic Press, 1989)
109. R J Simpson, R L Moritz, G S Begg, M R Rubira, E C Nice *Anal. Biochem.* **177** 221 (1989)
110. M W Hunkapiler, K Granlund-Moyer, N W Whiteley, in *Methods in Protein Sequence Analysis* (Ed. J E Shively) (Clifton, NJ: Humana Press, 1986)
111. B Juodka, M Pfuetz, D Werner *Nucl. Acids Res.* **19** 6391 (1991)
112. B B Baxter, M D Topal *Biochemistry* **32** 8291 (1993)