Russian Chemical Reviews **67** (3) 245 – 262 (1998) © 1998 Russian Academy of Sciences and Turpion Ltd

UDC5 47.963.32; 577.113.4

Double-helical nucleic acids with cross-linked strands: synthesis and applications in molecular biology

S I Antsypovich, T S Oretskaya

Contents

I.	Introduction	245
II.	The main methods for the synthesis of cross-linked DNA duplexes	246
III.	Methods for the synthesis of modified oligonucleotides	246
IV.	The use of psoralens for the synthesis of covalently linked DNA duplexes	247
V.	Cross-linked nucleic acid duplexes synthesised using pyrrole derivatives	250
VI.	The main methods for the synthesis of nucleic acid duplexes with cross-links in predetermined positions of the double helix	252
VII.	The use of sulfhydryl derivatives for the synthesis of cross-linked nucleic acid duplexes	256
VIII.	Synthesis of cross-linked nucleic acid duplexes using sugar moieties of nucleosides	259
IX.	Conclusion	261

Abstract. Data on the methods employed for cross-linking of DNA strands and for the synthesis of oligonucleotide duplexes with cross-links between strands are summarised. Existing methods are systematised; their advantages and drawbacks are discussed. The examples of applications of DNA duplexes with covalently cross-linked chains for the study of protein – nucleic acid recognition and mechanisms of action of nucleic acid-binding proteins for gaining information about the spatial structure of nucleic acids, and for the solution of other problems of molecular biology are given. The bibliography includes 131 references.

I. Introduction

Studies of fine details of cellular processes related to the expression and processing of genetic information require that a new generation of modified oligonucleotide derivatives be created that can be used as tools in these studies.

Therefore, in recent years the chemical approach employing modified oligonucleotides for the study of mechanisms of interaction of proteins with nucleic acids has been widely used in combination with physicochemical methods, such as X-ray structural analysis and NMR spectroscopy. The direct modifications of the nucleic acid chemical structure in protein – nucleic acid complexes and the analysis of properties of such synthetic systems has made it possible to gain information about the spatial structure of these complexes and the nature of protein-nucleic acid contacts. Directed modification allows useful properties to be conferred on oligonucleotides, such as resistance against nuclease degradation, enhanced thermodynamic stability of the duplexes formed by them, low toxicity, and the ability to penetrate cellular membranes. It also favours selective recognition and site-specific (including covalent) binding of DNA with proteins.[1–9]

S I Antsypovich, T S Oretskaya Department of Chemistry, M V Lomonosov Moscow State University, Leninskie Gory, 119899 Moscow, Russian Federation. Fax (7-095) 939 31 81. Tel. (7-095) 939 31 48

Received 11 February 1997
Uspekhi Khimii **67** (3) 274 – 293 (1998); translated by R L Birnova

Oligonucleotides and their reactive derivatives are important tools for the study of protein – nucleic acid interactions in transcription, translation, replication, repair, recombination, and modification of nucleic acids. Until fairly recently, it was mainly derivatives of single-stranded oligonucleotides that were used to this end. Meanwhile, a great variety of processes, such as repair, recombination, initiation of transcription by the corresponding factors, *etc*., occur with the involvement of double-stranded nucleic acids. Consequently, it is the double-stranded structures that should be used as models and affinity reagents in studies of these processes.

Modified nucleic acid duplexes have found use in elucidating the mechanisms of action of enzymes that interact with nucleic acids and in establishing the features of protein – nucleic acid recognition. The method based on the use of cross-linked nucleic acid duplexes is one of the most promising and popular procedures for investigation of protein – nucleic acid interactions. DNA duplexes with covalently cross-linked strands can be used as models for the study of the majority of the aforementioned cell-mediated processes and for the construction of affinity reagents and 'decoys' for proteins binding double-stranded DNA.[10–16]

For example, if the mechanism of action of an enzyme includes the unwinding of the double helix of DNA, covalent cross-links between the strands will prevent this process and thus strongly suppress the action of the enzyme. Duplexes stabilised by cross-linking can be potential inhibitors of nucleic acid-binding proteins.[10, 13] Duplexes with covalently cross-linked chains are promising for the isolation of proteins by affinity chromatography. Covalent cross-links formed in specific positions can stabilise thermodynamically unstable DNA structures.[15, 17]

Two main methods exist for the synthesis of this type of compound. Historically, the first of them consists in non-specific treatment of sufficiently long DNA molecules with bifunctional alkylating agents, such as derivatives of psoralen, pyrroles, and some other compounds.[1–5, 18] The second method, which leads to the formation of DNA duplexes with covalently cross-linked chains, is based on the use of derivatives of modified oligonucleotides containing chemically reactive groups.

One of the first to begin the study of reactions between the strands of DNA duplexes more than 30 years ago was a group of Russian investigators from the Siberian Branch of the Russian Academy of Sciences directed by Academician D G Knorre.[1–3, 19–23] These studies have ushered in the method

of 'targeted chemical modification' which has been used in the structural analysis of nucleic acids. Conceptually, this method is related to the construction of synthetic analogues of nucleases specific for a definite type of DNA base. Later, this method gave way to the method of 'complementarily targeted modification'.[1–3,24–31] The latter consists in the introduction of a reactive alkylating group into an oligonucleotide. After hybridisation, this group interacts site-specifically with the target DNA, *i.e.*, it alkylates the target at a predetermined position.

Thus, the majority of studies connected with the use of alkylating reagents aimed at the modification of nucleic acids dealt with the synthesis of highly reactive derivatives, which reacted with the target nucleic acid resulting in labilisation of the glycosidic bond in one of the nucleosides and, as a consequence, in the cleavage of the phosphodiester bonds of the oligonucleotide.[1–3] Irrespective of the nature of the alkylating reagent, double-stranded complexes with covalently cross-linked chains formed were unstable and underwent spontaneous hydrolysis in aqueous solutions. These investigations gave information about the reactions between the chains of double-stranded nucleic acids; however, the synthesis of DNA duplexes with covalently cross-linked chains was not the aim of these studies, and for this reason they are not considered in detail in the present review.

This review encompasses studies devoted predominantly to the preparation of synthetic oligonucleotide duplexes with covalently cross-linked chains and to the methods of nonspecific cross-linking of DNA chains. It does not consider covalently closed structures of nucleic acids, such as cyclic oligonucleotides and dumbbell-like structures, since this is an independent trend of research, which is also being developed in recent years.

II. The main methods for the synthesis of cross-linked DNA duplexes

It is known that bifunctional alkylating reagents that can form cross-links between the chains (such as 2-chloroethylnitrosourea, *etc.*) can also inhibit nucleic acid biosynthesis. In many cases, such reagents are efficient antitumour drugs.[5,6,32–34] This is due to the fact that nonspecific cross-linking of DNA chains is highly toxic for the cell,[5,34] since it prevents expression of the genetic material. The ability to form cross-links in double-stranded DNA is the reason for the cytotoxicity of many compounds employed in clinical practice.

A great number of compounds that can cross-link nucleic acid chains have been described. Psoralen[7,11,12,18,35–64] and pyrrole[34,65–71] derivatives are the most popular to this end. Other compounds used for the formation of interchain cross-links in double-stranded nucleic acids include aziridinolbenzoquinones (2,5-bisaziridino-1,4-benzoquinone[72,73] or alkylaziridino-1,4-benzoquinones),[74] aziridinonitrobenzamides (5-aziridino-2-nitro-4-nitrosobenzamide and 5-aziridino-4-hydroxylamino-2-nitrobenzamide[75]), anthracyclines (daunomycin,[76] adriamycin,[77] anthramycin,[78,79] mitomycin C[80–83] and its analogues[84]), melamine derivatives, *e.g.*, trishydroxymethyl(trimethyl)melamine,[85] acridine derivatives, *e.g.*, 1-nitro-9-(3-*N*,*N*-dimethylaminopropyl)aminoacridine,[86] chloroambucyl,[87,88] 4-hydroperoxycyclophosphamide,[89] tris(2-chloroethyl)amine,[90,91] 1,3-bis(2-chloro-ethyl)-1-nitrosourea,[92] 2-aminomethylpyrrolidine(cyclo-butane-1,1-dicarboxy)platinum(II),[93] cisplatin,[34,80,92,94] and *cis*-diamminedichloroplatinum(II).[95]

Synthetic cross-linked oligonucleotide duplexes are the most convenient tools for the study of gene structure and functions. Their application has a number of salient advantages over nonspecific alkylation of nucleic acids. The presently known bifunctional electrophilic reagents usually recognise those sequences in nucleic acids the length of which does not exceed two nucleotide residues. As a rule, the reactions with the participation of these reagents yield mixtures of monoadducts and products of intrastrand cross-links. Heterogeneity of the reaction mixture resulting from the alkylation of DNA with non-specific bifunctional reagents[18] is often an obstacle in the studies of effects of individual interstrand cross-links on DNA processing.

At present, the synthesis of oligonucleotide duplexes cross-linked at the predetermined positions of the double helix is complicated because the cross-links between the chains are either unstable[96] or endow significant disturbances in the secondary structure of DNA.[35,36] New approaches to the synthesis of DNA duplexes with interstrand cross-links have recently appeared, which permit cross-links to be created at any predetermined site of the DNA duplex and to introduce several cross-links into the same duplex.

The main requirements for the methods of synthesis of modified oligonucleotides that are used in the preparation of duplexes with covalently cross-linked chains can be formulated as follows: (1) the possibility of targeted incorporation of modified units into any predetermined position of the oligonucleotide chain; (2) the applicability of standard procedures of oligonucleotide synthesis, and (3) the possibility of incorporation of several modified units into the same oligonucleotide for subsequent synthesis of an oligonucleotide duplex with several cross-links between the chains.

III. Methods for the synthesis of modified oligonucleotides

Generally, an oligonucleotide can be modified at heterocyclic bases, phosphate groups (internucleotide or terminal), or carbohydrate residues. The potentialities of modern chemical synthesis of nucleic acid fragments allow targeted incorporation of one or several modified units (identical or different) into the oligonucleotide.

Compounds carrying primary aliphatic amino groups as well as aldehyde, carboxy, sulfhydryl, and phosphate groups are used for the cross-linking of the duplex chains. The general requirement for the formation of a cross-link is the localisation of the reactive group at the end of a spacer (a cross-linking arm). The length of the spacer can vary widely depending on the mutual arrangement of the reactive groups within the duplex. Thus in the case of psoralens, a relatively short, two-carbon-atoms-long spacer is sufficient for the interaction with the opposite chain.[37,38]

A functional group can be incorporated into the oligonucleotide by the following methods:[8]

(a) using a modified nucleotide component carrying the corresponding functional group in the chemical oligonucleotide synthesis;

(b) incorporating a precursor compound into the oligomeric chain followed by treatment with the corresponding reagents that give the target-modified unit either in the course of the oligonucleotide synthesis or after its completion;

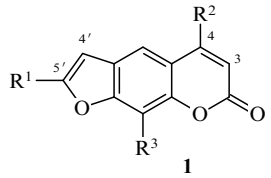
(c) post-synthetic modification of an oligonucleotide.

The potentialities of the latter approach are limited. It can be used only for preparing oligonucleotide analogues carrying numerous statistically modified units (*e.g.*, units with identical heterocyclic bases undergoing modification). In the former two methods, modification is envisaged in the very pattern of the chemical synthesis of oligonucleotides.

The first method is the most versatile. It consists in preliminary formation of modified units and their subsequent incorporation into the oligonucleotide in the process of synthesis. This method does not constrain either the number or the chemical nature of the units to be introduced into the oligonucleotide chain. Moreover, it significantly simplifies isolation of the target product. Side reactions in the process of oligonucleotide synthesis are minimised because all chemically reactive functional groups of the oligonucleotide have been preliminarily protected. The initial modification of the nucleoside to be inserted is usually carried out in non-aqueous media, which permits the use of a broad spectrum of modifying reagents, whereas post-synthetic modifications are usually performed in aqueous solutions.

IV. The use of psoralens for the synthesis of covalently linked DNA duplexes

Psoralens {7*H*-furo[3,2-*g*][1]benzopyran-7-ones} † are polycyclic aromatic compounds of the general formula **1**; they have a planar structure and are capable of intercalating between the base pairs of the double helix.[36, 39–41]

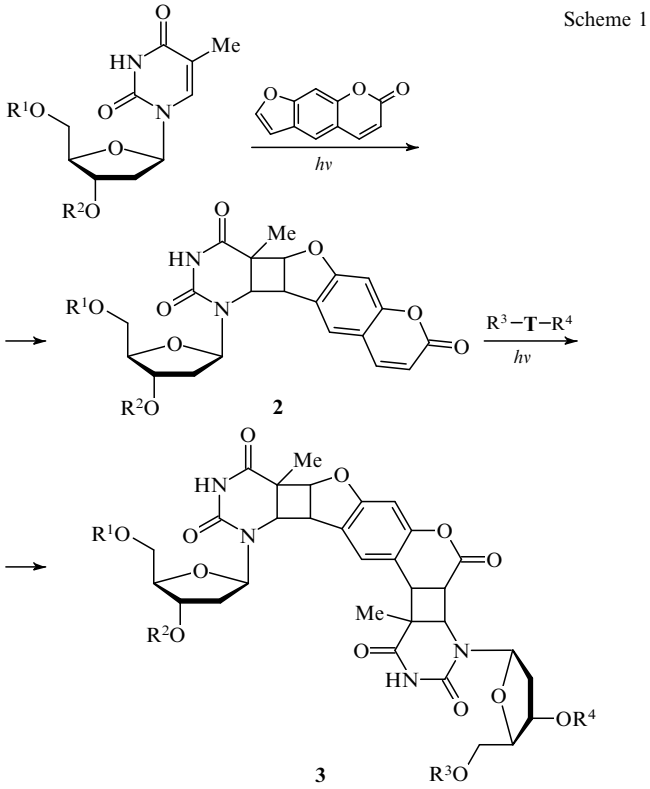


1

R^1, R^2, R^3 = H, Alk.

These compounds are specific to the thymidine and uridine residues. On UV irradiation of DNA duplexes at λ 320–360 nm, psoralen incorporates into the double helix and binds covalently with the pyrimidine bases of DNA, preferentially with thymine, at positions 5 and 6 to form four-membered rings (Scheme 1).[37, 42, 43]

Scheme 1



R^1–R^4 are fragments of oligonucleotide chains.

Psoralens can enter into photoaddition reactions with cytidine. However, this reaction occurs 15 times more slowly than with thymidine.[44] In DNA, psoralens predominantly bind with the 5′-d(TpA) sequence.[40, 43, 45, 46] The efficiency of psoralen photoaddition to this sequence is at least one order of magnitude higher than that to thymidine in any another environment. The same regularities hold for the synthesis of cross-linked RNA duplexes.[11, 12, 47, 48]

Since psoralen is a bifunctional reagent, the photo-induced reaction can lead either to the monoadduct of the type **2**, when the psoralen molecule is added to only one chain of nucleic acid, or to the bisadduct of the general formula **3**. In the latter case, a duplex with covalently cross-linked chains is formed. The formation of the bisadduct **3** requires the absorption of two photons[49] and occurs in two steps (Scheme 1).[7, 11, 45] If the pyrone ring of psoralen enters into the reaction at the first stage, the bisadduct is not formed, since the furan ring of the monoadduct becomes incapable of absorbing UV light.[11]

Intercalation of psoralen into the double helix, which usually increases the thermodynamic stability of the duplex, is a prerequisite for the reaction between the chains. The formation of adducts can be accompanied by photoinactivation of psoralen, which consists in photochemical degradation of the pyrone ring.[38] The adducts of nucleic acids with psoralen are decomposed into the original components on irradiation with UV light with a wavelength that depends on the structure of the psoralen derivative;[39, 40, 42, 45] as a rule, it is equal to 254 nm. Thus, the cross-linking of DNA chains with psoralen is reversible.

In the absence of UV irradiation, psoralens are non-toxic. The ability of these compounds to form cross-links between DNA chains makes possible their application in *in vitro* and *in vivo* studies of the DNA structure. Psoralens are used both in medicine for photochemotherapy of some diseases[39, 40, 50, 51] and in investigations of some molecular-biological processes. Psoralen has been used for the analysis of the secondary structure of RNA[35, 47, 48, 50–52] and the structure of the DNA–RNA-polymerase transcription complex.[12, 53–55] Currently, psoralen derivatives are used for the study of the secondary and tertiary structure of nucleic acids as well as for the analysis of mechanisms of DNA repair.[56–58]

Methods of targeted incorporation of psoralen into oligonucleotides (directly in oligonucleotide synthesis or after its completion and removal of protective groups) are of prime interest. The monoadduct of psoralen with the oligonucleotide can be produced post-synthetically using an activating reagent.[12, 59, 60] Such a monoadduct was obtained in the reaction of a 5′-phosphorylated self-complementary oligonucleotide with 4′-aminoalkyl-4,5′,8-trimethylpsoralen in the presence of 1-ethyl-3-(3-dimethylamino)propylcarbodiimide hydrochloride, $EtN{=}C{=}N(CH_2)_3{-}NMe_2 \cdot HCl$ (water-soluble carbodiimide, CDI). Hybridisation and irradiation yield a DNA duplex, which is cross-linked at the both ends.

The second method consists normally in the addition of psoralen to the 5′-end of the oligonucleotide in the oligonucleotide synthesis. Phosphoramidite derivatives of psoralen are used at the last stage of the synthesis.[37, 38] A typical synthetic scheme for the incorporation of psoralen into oligonucleotides is described in Ref. 38: 4,5′,8-trimethylpsoralen (**4**) was converted into the phosphoramidite **6** (Scheme 2) and the latter was used for the oligonucleotide synthesis.

4,5′,8-Trimethylpsoralen **4** was added to the 5′-end of an 18-membered oligonucleotide; the resulting oligonucleotide **7** was introduced into the cross-linking reaction of DNA duplex chains (Scheme 3). 4,5′,8-Trimethylpsoralen **4** was shown[38] to be stable both under conditions of synthesis and post-synthetic treatments of the oligonucleotide. It is noteworthy that psoralen-containing oligonucleotides of the type **7** are characterised by higher hydrophobicity, which is manifested in the increased retention time (in comparison with non-modified analogues) on HPLC.[38]

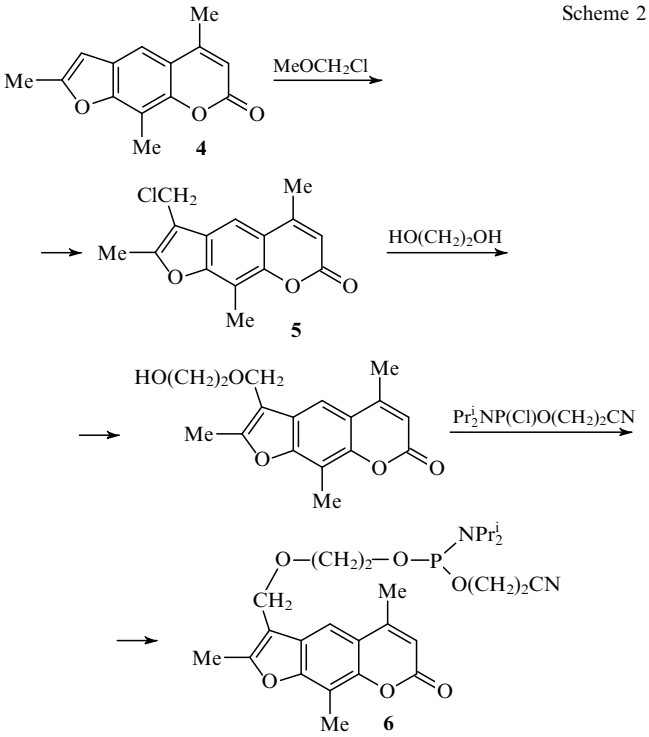
In this case, the nucleotide sequence of the oligonucleotides constituting the duplex was selected so that psoralen **4** intercalated into a double helix and reacted with thymidine in the 5′-d(TpA)/3′-d(ApT) sequence.[61] It was found that the maximum yield of the DNA duplex **8** could be obtained when the $CH_2{-}O{-}CH_2CH_2$ group was used as the spacer for the addition of psoralen to the oligonucleotide (Scheme 3).

As a rule, the yield of the product **8** increases within the first 10–30 min after the beginning of irradiation (360 nm, 20 °C) and reaches 70%–80% after 60 min. A quantitative yield could not be attained even with a 100-fold excess of the modified oligonucleotide **7** with respect to the DNA matrix[38] due to partial inactivation of psoralen in the irradiation.

The 5′-d(TpA)/3′-d(ApT) motif is present in signal sequences of many genes. They represent potential targets for psoralen-

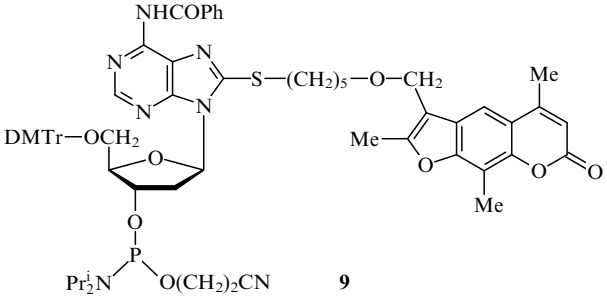
† Henceforth, the nomenclature accepted in the original publication is used.

Scheme 2



modified oligonucleotides of the type **7**. Thus oligonucleotides of this type can serve as convenient probes for the study of gene expression.

A method for the synthesis of psoralen-containing oligonucleotides using a modified deoxyadenosine‡ phosphoramidite **9** has been proposed.[37]



DMTr = $(4\text{-MeOC}_6H_4)_2PhC$.

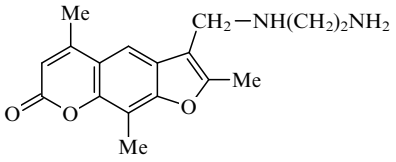
‡ Hereinafter, the term 'deoxynucleosides' will refer to 2′-deoxyribonucleosides.

4,5′,8-Trimethylpsoralen **4** was attached to the C(8) atom of deoxyadenosine through the sulfur atom and the pentamethylene spacer. Subsequent phosphitylation afforded compound **9**. To achieve the maximum yield (90%) in the coupling of the phosphoramidite **9** with an oligonucleotide, 5-(4-nitrophenyl)-1*H*-tetrazole was used instead of tetrazide; the duration of condensation was 8 min.

The optimum length of the spacer was established by molecular modelling. In this particular case, the optimum length of the spacer needed for the psoralen intercalation into the 5′-d(TA)/3′-d(AT) site is five carbon atoms. An analogous derivative of deoxyadenosine of the type **9** but with a two-carbon-atoms-long spacer manifested a very weak cross-linking capacity. The efficiency of photoaddition to the complementary matrix chain exceeded 90% after 1-hr UV irradiation at λ 345 nm. This resulted in the formation of several products, which correspond to different possibilities of psoralen incorporation into the double helix. It was noted that psoralen was effectively incorporated both into the 5′-d(TA)/3′-d(AT) and the 5′-d(AT)/3′-d(TA) sequences, where it reacted both with the thymidine unit in the opposite chain and the adjacent one.

Owing to the fact that the ends of DNA duplexes of the type **8** with covalently cross-linked chains remain free, various labels (radioactive, fluorescent, or any other) can be easily attached to these ends. In addition, such duplexes can be incorporated into more extended DNA structures.[37]

The reaction of oligonucleotide methylphosphonates (OMP) bearing the psoralen derivative, 4′-[*N*-(2-aminoethyl)aminomethyl]-4,5′,8-trimethylpsoralen **10** (PS-OMP) attached to their 5′-ends with single- and double-stranded DNA, has been studied.

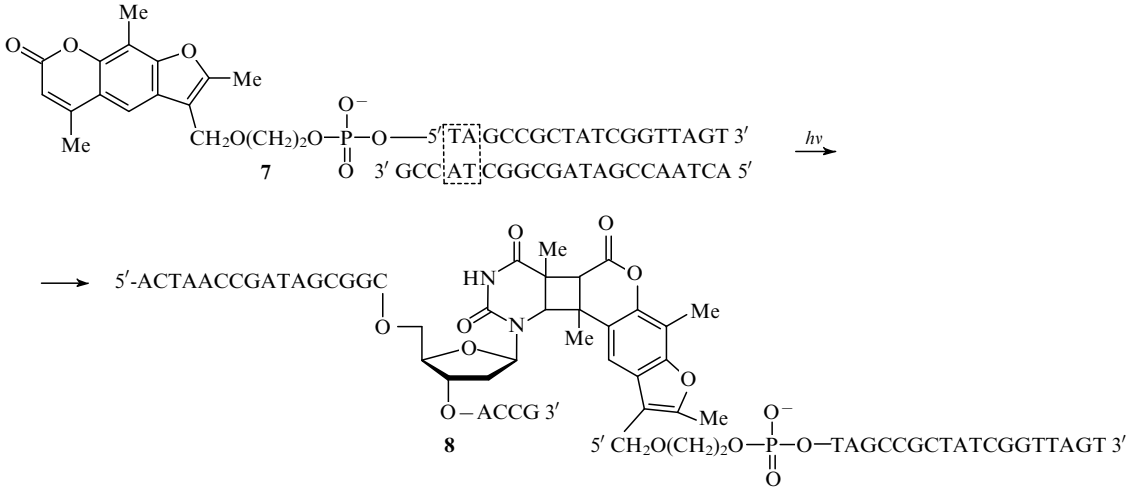


10

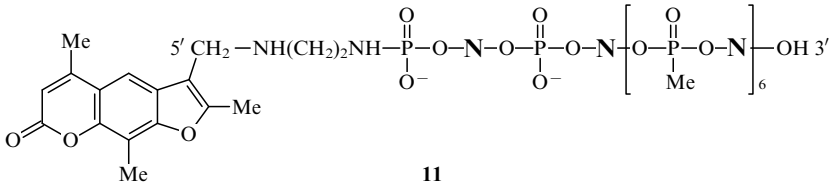
Compound **10** was coupled with the 5′-end of OMP carrying a phosphate group to give compound **11** (Scheme 4). Water-soluble CDI was used for the activation of the phosphate group. It was shown that the P – N bond is resistant to nucleases; however, it is cleaved in an acidic medium to give the original 5′-phosphorylated OMP.

On irradiation, PS-OMP specifically reacted with the template. The presence in the duplex of less stable nucleotide pairs (*e.g.*, G·T or G·A) reduced the yields of the duplex with covalently cross-linked chains.[12] This is due to the decrease in thermal stability of the original duplex. The formation of by-products that result from the interaction between the chains in

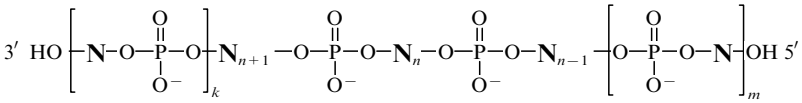
Scheme 3



Scheme 4



11



DNA target

N denotes deoxynucleoside residues, $m \geqslant 6$, $k \geqslant 0$.

non-canonical duplexes depends on temperature: by-products are formed at 20 °C and are absent at 40 °C. This can be explained by dissociation of less stable non-canonical duplexes with an increase in temperature. In the case of a canonical duplex, the temperature dependence of the yield of the covalently cross-linked duplex is not observed below its melting point. At higher temperatures, the yield of the target product drops drastically. The efficiency of the reaction between the chains of a DNA duplex is also influenced by the ratio of the rate of this reaction to the rate of psoralen photoinactivation. The yield of the cross-linked DNA duplex increases with increasing PS-OMP concentration. However, if the PS-OMP concentration is twice as high as that of the matrix, the latter becomes saturated and the yield of the target product does not further increase.[12]

The efficiency of the interaction of various oligonucleotides with the target under identical conditions differs markedly,[12] although the differences in the nucleotide sequences should not be manifested at 4 °C. Molecular simulation has shown [12] that psoralen can be inserted between the nucleotides of the target in two modes: either between the terminal nucleosides $N_{(n+1)}$ and $N_{(n)}$ or between the nucleosides $N_{(n)}$ and $N_{(n-1)}$, where $N_{(n+1)}$ is the 3′-terminal nucleoside of the target and $N_{(n)}$ is the nucleoside that is complementary to the 5′-terminal nucleoside of PS-OMP (Scheme 4). In the former case, the spatial structure of the reaction centre is the most favourable for the photoaddition of psoralen. The reaction between the chains is the most effective, when thymidine occupies position $(n + 1)$ of the template and position $(n-1)$ in OMP, which precedes the psoralen-modified unit, and adenosine is in the positions n. In this case, the reaction occurs with two possible sites and the yield of the duplex with covalently cross-linked chains reaches 95%.

Oligonucleotides of the type **11** were used for the analysis of the secondary structure of nucleic acids.[12] It was shown, in particular, that the cross-linking with the single-stranded region of DNA occurs 10–30 times more efficiently than with the double-stranded one. Since OMP are characterised by enhanced resistance in the cell, these oligonucleotide derivatives can be used in *in vivo* studies.[12]

A series of PS-OMP that are complementary to the template and non-template strands of a 57-membered DNA duplex containing a 17-membered T7 promotor of RNA-polymerase were synthesised.[12] The effects of PS-OMP on *in vitro* transcription were investigated. It was shown that PS-OMP, which is covalently cross-linked with the encoding region of DNA, effectively suppressed transcription, whereas the oligonucleotide cross-linked with the non-coding region had little inhibitory effect. Thus, the oligonucleotide that is covalently cross-linked to the template in the promotor region prevents binding of polymerase with the double helix of DNA either directly or by changing its conformation in the promotor region.

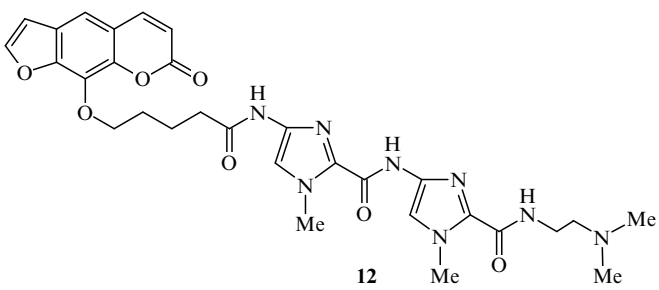
Being an agent that can directly react with DNA, psoralen can induce various mutations.[11] Study of mutagenesis is necessary for the understanding of regularities of the origin of cancer, hereditary diseases, and evolutionary processes.[11, 56, 57] It remained unclear which of the psoralen-induced mutations were generated by the monoadduct and which resulted from the presence of cross-links between the chains.

The use of a plasmid containing a single, definite damaged site has made it possible to relate the initial damage of DNA induced by psoralen to the origin of subsequent mutations.[11] The use of statistically modified DNA did not allow a correlation between the nature of mutations and the type of the adduct. The cyclic closed double-stranded plasmid containing the site-specifically incorporated psoralen as a mono- or bis-adduct (in the latter case, the cross-link is formed between the strands of the plasmid), has been synthesised *in vitro* and incorporated into the cells of *E. coli*. The recombinant DNA contained only one site with thymine in the form of a monoadduct with psoralen. Synthesis of the psoralen-containing plasmid was carried out using T4-DNA-polymerase and an oligonucleotide with psoralen at the 3′-end as the primer. In this case psoralen did not influence either the elongation of the chain or its subsequent ligation. The monoadduct was converted into the bisadduct by UV irradiation; the plasmid with covalently cross-linked chains was isolated taking advantage of its acquired ability to undergo reversible thermal denaturing. In contrast to the plasmid with cross-linked chains, denaturing of an ordinary plasmid is irreversible. The type and frequency of mutations was followed using two different DNA molecules. It was found that both types of psoralen-induced damages (the formation of the both monoadduct and the cross-links between the chains) can lead to severe genotoxic effects.[11] This finding is consistent with the results obtained by other authors.[58, 62, 63]

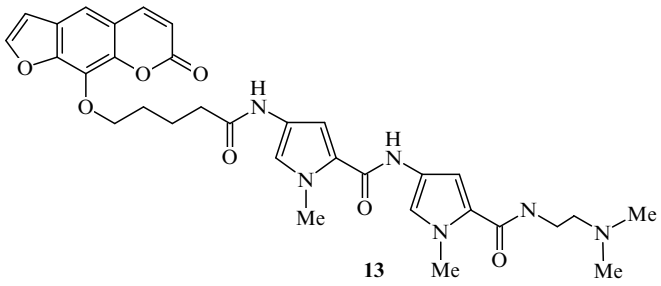
The mutation induced by the formation of the monoadduct was eliminated by the repair mechanism of the cell. In the case of the bisadduct, a great number of secondary mutations (deletions, substitutions, inversions) evolved. The activity of DNA with cross-linked chains was 1.6% of the native one.[11]

Thus, it was shown that it is by bisadducts that the toxicity of psoralen is determined, and their formation results in severe damage to DNA. Treatment of nucleic acids with psoralen was shown to prevent replication and transcription processes.[11]

It was proposed [64] to couple psoralen with distamycin-like imidazole **12** and pyrrole **13** derivatives in order to promote the delivery of psoralen to specific sequences of DNA. Like the oligopeptide distamycin, these compounds bind with DNA in the region of the small groove.



12



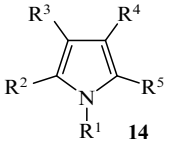
13

The use of oligonucleotides for the delivery of psoralen to specific sequences of DNA has posed a number of problems. They are connected with the low ability of the oligonucleotides to penetrate through hydrophobic cellular membranes and their susceptibility to nuclease hydrolysis. This circumstance restricts the application of such compounds.

The derivatives **12** and **13** are devoid of these drawbacks and can deliver psoralen to both GC-rich (imidazole analogues of distamycin, **12**) and AT-rich (pyrrole analogues, **13**) regions of DNA. It was found that the derivatives **12** and **13** modify DNA 10 and 1000 times more efficiently than psoralen. Compound **13** is 100 times more efficient than compound **12**. Hence, psoralen, being attached to reagents that possess affinity with respect to DNA, manifests enhanced efficiency in the formation of multiple cross-links between the chains of extended DNA.

V. Cross-linked nucleic acid duplexes synthesised using pyrrole derivatives

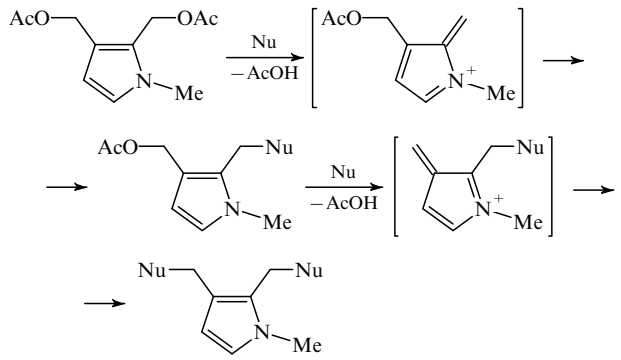
Bifunctional pyrrole derivatives **14**, together with psoralen derivatives, are widely used for cross-linking of chains of double-stranded nucleic acids.[34, 65–71] Only bifunctional pyrrole derivatives can cross-link the nucleic acid chains.



14 R^1-R^5 = H, Alk, CH_2OH, $CH_2OC(O)Alk$.

Cross-linking with pyrrole derivatives **14** involves exocyclic amino groups of heterocyclic bases of DNA (usually, deoxyguanosine) and those functional groups of pyrrole derivatives that can undergo nucleophilic attack. Hence, pyrrole derivatives are alkylating agents. A hypothetical mechanism of this reaction has been proposed (Scheme 5).[65]

Scheme 5



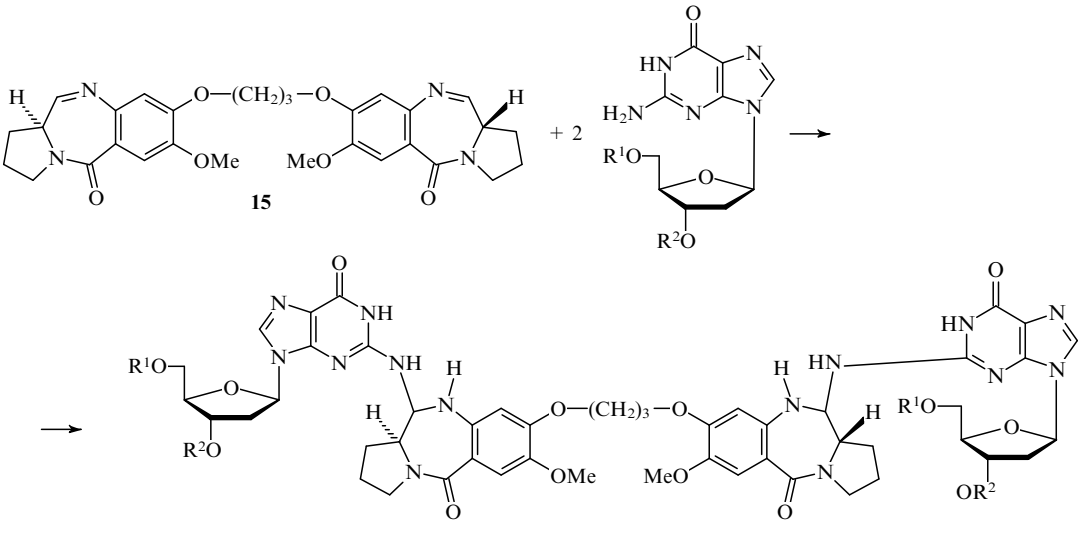
Nu is the exocyclic amino group of guanine.

Pyrrole derivatives as such are not specific reagents and can react with many reaction centres of DNA. Therefore, treatment of DNA with bifunctional pyrrole derivatives of the type **14** cannot be used to introduce cross-links at predetermined positions of DNA. The main drawback of pyrrole derivatives as antitumour agents is their low selectivity (*i.e.*, the affinity for sequences with the length of no more than two nucleotides). Such limited selectivity of pyrrole derivatives is due to the high probability of occurrence of certain short repeated nucleotide sequences in the genome.

Enhancement of selectivity of bifunctional pyrrole derivatives is an important problem, particularly as regards their application as therapeutic agents. This can be achieved through a more targeted delivery of pyrrole derivatives to a definite genome region, *i.e.*, to a more rare nucleotide sequence. However, the solution of this task demands that the affinity of such agents for nucleic acids be increased.[2, 34, 65–71] As a rule, it is sufficient to achieve selective interaction of the cross-linking reagent with sequences more extended than four nucleotide pairs. The main route for attacking the selectivity problem is the addition of a bifunctional alkylating agent (which has little or no affinity for nucleic acids) to compounds manifesting high affinity for definite sequences of DNA, *e.g.*, to oligopeptides.[34, 65–70]

Synthesis of a new irreversible reagent **15** has been described.[66] Compound **15** belongs to the group of pyrrolobenzodiazepine dimers (PBD dimers). It binds with DNA through the small groove and reacts with the amino group of guanine (Scheme 6). The preferential sequence for the binding of this compound is the Pu-G-Pu triplet, where Pu is the purine nucleotide (A or G). The yield of the interstrand reaction product is about 50%.

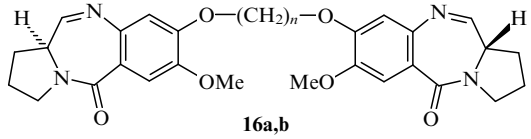
Scheme 6



15

R^1, R^2 are fragments of oligonucleotide chains.

The synthesis of C(8)-linked PBD dimers of the type **16**, which are homologous to compound **15**, has been described by the same authors,[67] and their relatively high affinity for DNA has been demonstrated. The highest efficiency of the interchain reaction was achieved with compounds **15** and **16b** (three- and five-carbon-atoms-linking groups, respectively).

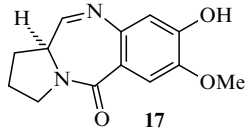


16a,b

n = 4(**a**), 5(**b**).

Treatment of double-stranded DNA with any PBD dimers increases the melting point of the duplexes.[67] The presence of cross-links in the linear double-stranded DNA prepared from the ring plasmid accounts for its rapid renaturation into a double-stranded form as the temperature decreases following thermal denaturing.[67] The ability of extended DNA to undergo rapid renaturation is a sign of cross-linking between the chains. Thus, PBD dimers can stabilise double-stranded DNA.

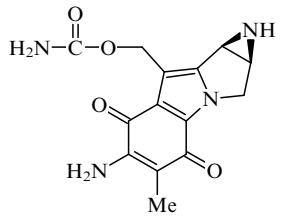
The cytotoxicity data [67] correlate with the yields of duplexes with covalently cross-linked chains. The PBD dimer (n = 3), *i.e.*, compound **15**, possesses the highest toxicity. On the other hand, the monomer of the type **17** hardly exerts any noticeable effect on the stability of DNA duplexes and is non-toxic.



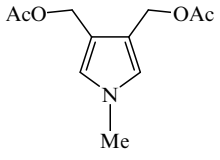
17

Thus a possibility exists for the application of these compounds for the treatment of tumours. Their main shortcoming is relatively low affinity for DNA.

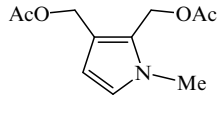
The mechanism of cross-linking of the DNA chains by bifunctional pyrrole derivatives was investigated.[65] It was found that all the pyrrole derivatives including natural (*e.g.*, the antitumour antibiotic mitomycin C **18**), and synthetic pyrroles **19** and **20** react with DNA by the same mechanism, they are directed to a common binding site and an alkylation site. Mitomycin C is converted into a bifunctional alkylating agent following reduction with specific enzymes. A binding site in DNA molecules, which is common for all the pyrrole derivatives in question, has been identified.



18

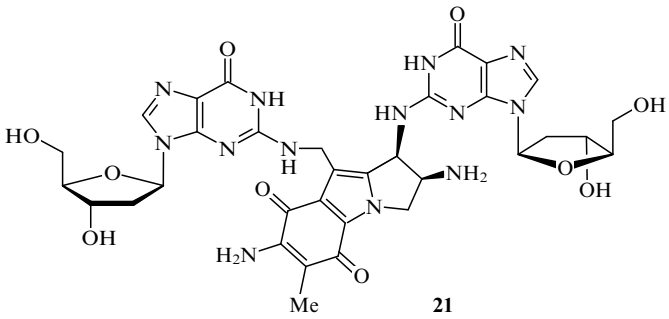


19



20

The exocyclic amino group of deoxyguanosine in the 5′-d(CG)/3′-d(GC) sequence is the common target site in DNA for all the alkylating pyrrole derivatives.[65] The formation of a cross-linked duplex requires the reaction of exocyclic amino groups of both deoxyguanosines with pyrrole derivatives. The structure of the adduct **21**, the reaction product of two molecules of deoxyguanosine and mitomycin C, is given below.



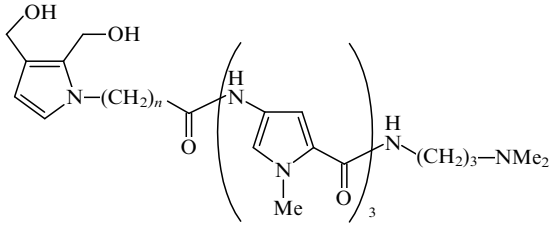
21

DNA duplexes containing only adenosine and thymidine residues have been studied by Wou *et al.*[65] These duplexes differed in the nature of nucleosides in several nucleotide pairs. To obtain evidence for the involvement of the exocyclic amino group of guanine in the reaction between the chains, deoxyguanosine was substituted by inosine. This gave the duplexes containing the 5′-d(ACGT)/3′-d(TICA) and 5′-d(ACIT)/3′-d(TICA) fragments. When even one deoxyguanosine was substituted by inosine, the interchain cross-linking product disappeared completely. It was also shown [65] that the reaction in the 5′-d(CG)/3′-d(GC) sequence occurs at least ten times more efficiently than in any other sequence.

The structure of compounds thus obtained was established by an analysis of the products of total enzymic hydrolysis of the cross-linked duplexes and by mass spectroscopy. It is by this method that the structure of compound **21** has been established.[65] The same method has been used for structural elucidation of the adducts of the synthetic pyrrole derivatives **19** and **20**. It should be emphasised that computer simulation of a cross-linked DNA duplex revealed a relatively significant distortion of the native B-form of the DNA helix irrespective of the nature of the pyrrole derivatives, which is associated with the torsional rearrangement of the DNA structure.

Unfortunately, the attempts to rationalise the reason for such specificity failed. The reaction between the chains is a multistage process, which includes the formation of a mono- and then of a bis-adduct. Each of these stages can be responsible for the specificity. Since both the synthetic pyrrole derivatives **19** and **20** and natural compounds having a more complicated structure manifest identical specificity, it was concluded [65] that the specificity of pyrrole derivatives is inherent in the common elements of their structure. On the other hand, the significant differences in the efficiency of pyrrole derivatives as the reagents that cross-link DNA chains point to the important role of individual structural features of pyrrole derivatives.

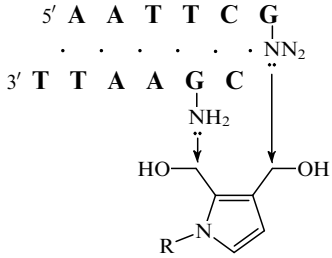
Recent developments in this area [68, 69] have been mainly focussed on the enhancement of the selectivity of pyrrole derivatives. In order to increase the efficiency of affinity binding of the alkylating reagent with the nucleic acid target, it was proposed [68, 69] to prepare its conjugate with an oligopeptide that can specifically recognise the small groove of the double helix. The affinity reagent **22** containing the oligopeptide distamycin, which recognises the 5′-d(AATT) sequence (through the small groove), and pyrrole derivative, which recognises the 5′-d(CG) sequence, was obtained.[68, 69]



22 – 24

n = 2 (**22**), 3 (**23**), 4 (**24**).

Thus, the affinity reagent **22** selectively recognises the nucleotide sequence 5′-d(CGAATT)/3′-d(GCTTAA) and produces a cross-link between the chains. The activity of this reagent was tested both with a synthetic duplex and a linearised plasmid. It was found that the derivative **22** is three orders of magnitude more efficient as a cross-linking reagent than a pyrrole derivative nonbound with the peptide. After thermal denaturation and cooling, the plasmid treated with this reagent underwent rapid renaturation into a double-stranded form, which can be explained by the presence of cross-links between the chains.



R is an oligopeptide.

The effect of compound **22** on DNA samples with different nucleotide sequences was studied. It was found that only the DNA containing the 5′-d(CGAATTCG)/3′-d(GCTTAAGC) sequence efficiently reacted with compound **22** (the yield of the cross-linked product is about 70% with respect to DNA). This sequence contains both 5′-d(AATT), the binding site for distamycin, and 5′-d(CG), the consensus binding site for pyrrole derivative. By-products were formed in all the cases studied as a result of the reaction of compound **22** with two deoxyguanosine residues of the same DNA chain.

The synthesis of a series of bifunctional reagents that are homologous to compound **22** has been described.[69] In these reagents (**23**, **24**), 2,3-bis(hydroxymethyl)pyrrole is also linked to the oligopeptide distamycin. They differ from compound **22** only in the length of the linker connecting pyrrole derivative with the oligopeptide. Compounds **22**–**24** are three orders of magnitude more efficient and more selective than ordinary pyrrole derivatives. They can form both intrachain and interchain cross-links in DNA depending on the nucleotide sequence in the cross-linking site of pyrrole derivatives. The interchain cross-link is formed in the duplex containing the 5′-d(CGAATT)/3′-d(GCTTAA) sequence, while the intrachain cross-link is formed in the case of the 5′-d(GGAATT)/3′-d(CCTTAA) sequence.

The structures of nucleosides linked through pyrrole derivatives and isolated following total enzymic hydrolysis of internucleotide bonds are identical for both the interchain and intrachain reaction products. Comparison of binding efficiency of various substrates with each of compounds **22**–**24** revealed that both binding sites [5′-d(CG)/3′-d(GC) (pyrrole derivative) and 5′-d(AATT)/3′-d(TTAA) (distamycin)] play an important role. Nucleotide substitution in any of these sites (CG for GC or A·T for G·C) sharply decreased the efficiency of the reaction. Thus, the 5′-d(CGAATTCG)/3′-d(GCTTAAGC) sequence was found to be the consensus sequence.[69]

Compounds **22**–**24** were brought into reaction with a linearised plasmid containing twenty seven 5′-d(AATT)/3′-d(TTAA) sites conjugated with 5′-d(GC)/3′-d(CG).[69] The efficiency of the interchain reaction was estimated according to Hartley.[5, 70, 71] This method is based on the ability of DNA with covalently cross-linked chains to undergo rapid renaturation into a double-stranded form after thermal denaturation. The efficiency of this reaction increased [69] with an increase in the reagent : DNA ratio up to 5 : 1 (these conditions favour saturation of the peptide-binding sites). The highest yield (80%) was observed on treatment of the oligonucleotide duplexes containing the 5′-d(CGAATTCG)/3′-d(GCTTAAGC) sequence with the compound **22**. For the 5′-d(CGAAAAACG)/3′-d(GCTTTTTGC) sequence, the yield of the cross-linked product was about 50%, whereas that for the 5′-d(GC...)/3′-d(CG...) sequences was no more than 10%. In the case of compound **23**, the yield for the 5′-d(CGAATTCG)/3′-d(GCTTAAGC) sequence was 40%. Sigurdsson and Hopkins [69] could not rationalise the decreased efficiency of this reaction with an increase in the linker length. Probably, this is due to the geometrical disorientation of pyrrole derivative in the reaction centre.

The structures of compounds thus obtained were analysed by studying the composition of products of total enzymic hydrolysis of duplexes with covalently cross-linked chains. Even prolonged hydrolysis of DNA with covalently cross-linked chains by a mixture of snake venom phosphodiesterase and alkaline phosphatase did not result in its complete disintegration into nucleosides. Increased amounts of snake venom phosphodiesterase in combination with endonucleases (DNAase I and DNAase II) were used for complete hydrolysis of the phosphodiester bonds of the duplex. The cytotoxicity of the compounds in question was relatively low and comparable with that of natural pyrrole derivatives.

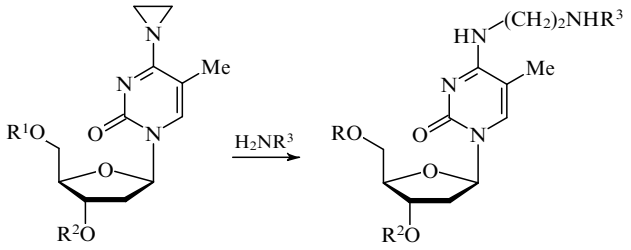
The use of compounds of the type **22**–**24** for cross-linking of DNA chains has a number of advantages over psoralen derivatives. Psoralen recognises the large groove of DNA and binds thymidine residues in 5′-d(TA)/3′-d(AT) sequences through intercalation into the double helix. In contrast, pyrrole derivatives coupled with peptides (compounds **22**–**24**) recognise the small groove. It is noteworthy that a more extended (six nucleotides) sequence is recognised in this case, and the reagent itself remains outside the double helix of DNA without causing its destabilisation. However, the use of such targeted-action agents is now limited due to the lack of the general concept of the design of oligopeptides that bind to a predetermined nucleic acid sequence.[34]

Thus, a series of compounds capable of cross-linking nucleic acid chains has been described.[11, 12, 18, 34–95] Each of them binds with the duplex at a definite site and produces covalent cross-links with two nucleotides from the opposite chains. However, the majority of compounds described in the literature have a number of disadvantages. They are devoid of high binding specificity with a unique site in nucleic acids. This results, as a rule, in the formation of multiple interchain and intrachain cross-links. These reagents are not versatile, because they cannot form cross-links at predetermined positions between nucleic acid chains, but interact only with definite heterocyclic bases. In some cases, they destabilise the double helix at the cross-linking region, which results in disturbances in the geometry of the canonical spatial structure of nucleic acids. Psoralens, pyrrole derivatives, and other compounds mentioned above have found use both in clinical practice (as antitumour agents) and in the study of mechanisms of functioning of some molecular-biological systems. Nevertheless, it is a new generation of chemically designed synthetic systems devoid of the above drawbacks that will be helpful for the study of peculiarities of structural organisation and mechanisms of functioning of intracellular structures at the molecular level.

VI. The main methods for the synthesis of nucleic acid duplexes with cross-links in predetermined positions of the double helix

The use of nucleic acid duplexes with cross-links at predetermined positions of the double helix is the most promising for the study of protein–nucleic acid recognition and mechanisms of action of enzymes. As a rule, synthesis of such duplexes requires the use of reactive oligonucleotide derivatives containing modified nucleosides. The use of these derivatives has a number of advantages over the above-described methods. In particular, they provide a full control over the localisation of cross-links to be incorporated between the chains and the completion of the synthetic process. Various types of reactive derivatives have been proposed.

The synthesis of a series of oligonucleotides **25** containing N^4,N^4-ethano-5-methyldeoxycytidine dC^e has been described.[97–99] These oligonucleotides interact with the complementary chain of DNA, and N^4,N^4-ethano-5-methyldeoxycytidine reacts with the oppositely located nucleotide immediately after hybridisation with the complementary target oligonucleotide under mild conditions and without treatment with activating reagents.[97] This method was named 'hybridisation-triggered alkylation'.

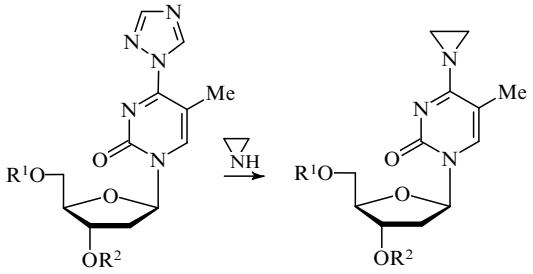


25

R^1 – R^3 are fragments of oligonucleotide chains.

A phosphoramidite derivative of 5-methyldeoxycytidine carrying a triazole residue at position 4 was used in the synthesis of oligonucleotides of the type **25**. Oligonucleotides **26** containing this residue are stable under the conditions of the phosphoramidite procedure of oligonucleotide synthesis.

Treatment of the oligonucleotide **26** with ethyleneimine converts it into the oligonucleotide **25** in quantitative yield.[97] This makes it possible to incorporate a reactive acid-labile group into the oligonucleotide following all acidic treatments needed for the removal of the dimethoxytrityl group in the process of oligonucleotide synthesis.



26 **25**

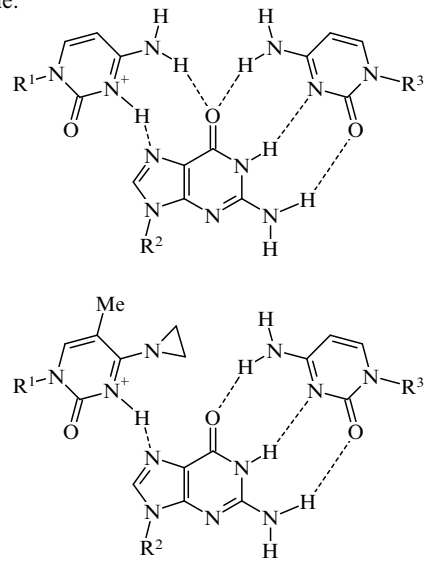
R^1 and R^2 are fragments of oligonucleotide chains.

N^4,N^4-Ethano-5-methyldeoxycytidine is unstable under conditions of standard treatment of oligonucleotides with ammonia (5 h, 55 °C, concentrated NH_4OH). Therefore, the 9-fluorenylmethoxycarbonyl group[97] (Fmoc) was used to protect the exocyclic amino groups of the nucleosides; this protective group is removed faster and under milder conditions of β-elimination than the benzoyl group routinely used for this purpose. Fmoc-protected deoxyadenosine and deoxycytidine phosphoramidites were synthesised. Attempts to obtain a Fmoc-derivative of deoxyguanosine in satisfactory yield failed. Conditions have been selected for removal of the protective groups with retention of the ethyleneimine ring of compound **25**.

The reaction with the oligonucleotide template is carried out for several days (the semi-conversion period is about 30 h at 24 °C).[98] The efficiency of the reaction between the chains substantially depends on the nature of the nucleotide that reacts with the dC^e unit of compound **25**. The following order of the efficiency of this reaction was established: A ≈ T ≫ C ≈ G. The reaction occurs with each of the four bases to a certain degree. In some cases, only one by-product is formed in the interchain reaction. It is assumed[98] that under these conditions the reaction involves not only the opposite base in the other chain, but also one of the neighbouring bases of the duplex. The melting point of duplexes with cross-linked chains is increased, on the average, by 15 °C (from 22 °C to 37 °C for a 21-membered AT-enriched duplex consisting of fifteen A·T pairs and six G·C pairs).

The main disadvantage of this method is the necessity of substitution of standard nucleoside phosphoramidites on Fmoc-containing ones. The Fmoc-group is extremely labile, therefore the yields of Fmoc-protected nucleosides are low and deoxyinosine must be used instead of deoxyguanosine. Moreover, the rate of the interchain reaction is too low (semiconversion period is 30 h) and therefore these oligonucleotides cannot be used as inhibitors of mRNA expression *in vitro*. On the other hand, the proposed method of interchain cross-linking[97] is devoid of such shortcomings as the necessity of activation of the alkylating function and low level of specificity. The latter can be due to the excessive reactivity of the alkylating group.

The majority of the known methods for the preparation of duplexes with covalently cross-linked chains are based on the interaction of a single-stranded DNA with oligodeoxyribonucleotide containing a reactive group. Recent reports suggest an alternative variant, which consists in the cross-linking of an oligonucleotide to double-stranded DNA within a triple helix (triplex).[7, 88, 100–109] Thus, a method has been proposed for specific cross-linking of dC^e oligonucleotides with units of the type **25** described above to double-stranded DNA within the triplex.[100] In this case, the oligonucleotide is recognised by Hugsteen's pairs and intercalates into the large groove of the target DNA duplex. In the dC^{e+}·G·C triplex, the electrophilic methylene group of N^4,N^4-ethano-5-methyldeoxycytidine comes close to the N(7)- and O(6)-nucleophilic centres of deoxyguanosine.



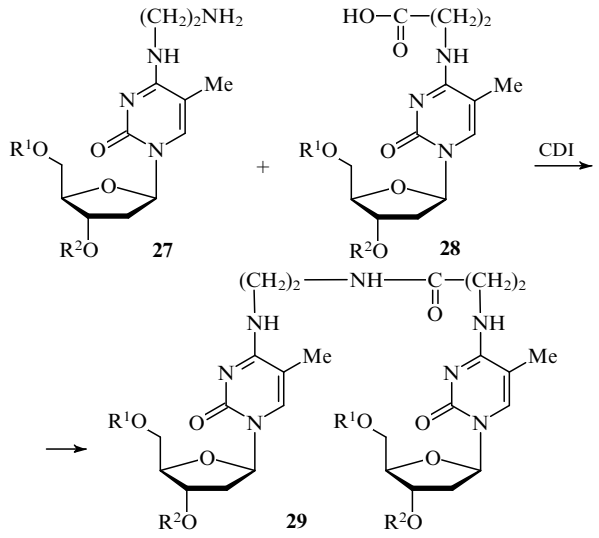
R^1 – R^3 are oligonucleotides.

The ethyleneimine ring interacts with one of the guanine bases of the double helix of the target DNA. The reaction between the chains occurs under physiological conditions in more than 95% yield and is completed within 16 h. This method can be promising for specific *in vivo* inhibition of DNA transcription and replication due to the formation of an untwistable triplex.[100, 101]

As has been mentioned above, one of the main problems connected with the introduction of new functional groups into the oligonucleotide is that under conditions of automated synthesis the DNA chain that is elongated is exposed to the action of acids, bases, and electrophilic reagents. The functional group to be introduced must withstand these treatments either as such or after protection. Since N^4,N^4-ethano-5-methyldeoxycytidine is sensitive to nucleophiles,[101] it can be incorporated only into oligothymidylates but not into an oligonucleotide containing nucleosides of all the four types, because only the former do not require post-synthesis ammonia treatment to remove protective

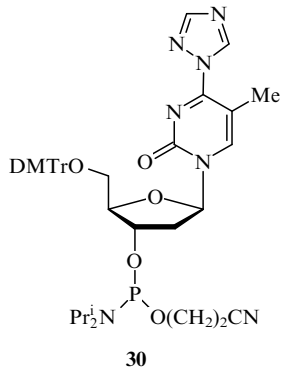
groups. This restricts the possible application of such oligonucleotides as substrates that can covalently bind to the DNA target. An attempt to apply this synthetic scheme to deoxyadenosine was even less successful.[110] It was found that N^6,N^6-ethanodeoxyadenosine was a weaker electrophile than N^4,N^4-ethano-5-methyldeoxycytidine.

In a recent study, devoted to the synthesis of oligonucleotide duplexes with covalently cross-linked chains Shabarova *et al.*[13] also described modification of the heterocyclic bases of deoxycytidin. Two approaches to the design of DNA duplexes with cross-links between oligonucleotides in the predetermined positions of the double helix were proposed. The first one consisted in the formation of an amide bond between the amino and carboxy groups of modified bases localised in the opposite chains, while the second one entails the formation of a covalent 'bridge' as the Schiff's base between the two amino groups using dialdehydes of the aliphatic series. To this end, deoxycytidine in the self-complementary oligonucleotide 5′-d(TAATGCATTA) was replaced by a 5-methyldeoxycytidine derivative containing an aliphatic amino group (oligonucleotide 10^n, **27**) or a carboxy group (oligonucleotide 10^c, **28**). These functional groups occupied terminal positions at the spacer linked to the exocyclic amino group of 5-methyldeoxycytidine. In the formation of the DNA duplex **29**, the carboxy and amino groups turned to be close to each other, which enabled the reaction resulting in the formation of an amide bond between the chains to occur.



R^1 and R^2 are fragments of oligonucleotide chains.
CDI is $EtN{=}C{=}N(CH_2)_3NMe_2 \cdot HCl$.

Synthesis of a self-complementary oligonucleotide precursor, which is common for compounds **27** and **28**, *viz.*, 5′-d(TAATGC*ATTA) **29**, where C* is the cytidine residue containing triazole at position 4, was carried out by the phosphoramidite method, and the triazolyl derivative of thymidine **30** was introduced in the reaction.

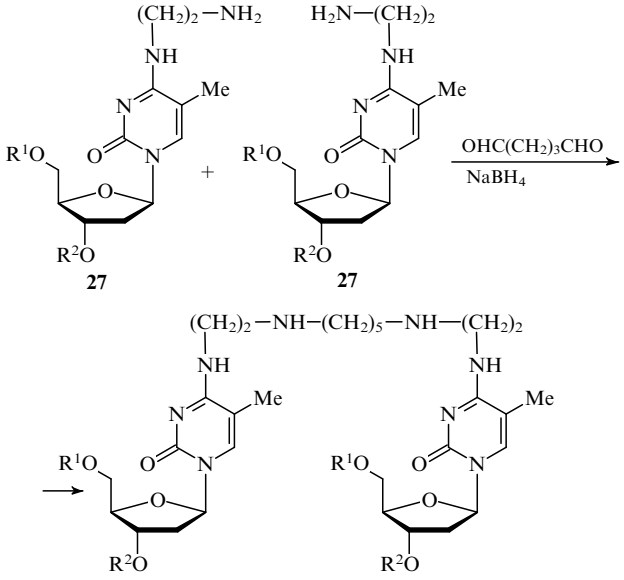


30

After completion of the synthesis, the polymer with the immobilised oligonucleotide was divided into two equal parts. One of them was treated with ethylenediamine to give compound **27**, while the other one was treated with β-aminopropionic acid to give compound **28**. The reaction between the chains was carried out with water-soluble CDI as the condensation reagent at 0 °C, which facilitated the formation of a double-stranded complex.

It must be noted that this system favours the formation of three types of oligonucleotide duplexes: $(10^n):(10^c)$, $(10^n):(10^n)$, and $(10^c):(10^c)$, since the original oligonucleotide is self-complementary. It was found that the $(10^n):(10^n)$ complex was the most stable, presumably owing to the electrostatic contacts between the amino and phosphate groups of the opposite chains. The predominance of the $(10^n):(10^n)$ complex in solution is responsible for the low yield of the target product, which can be formed from the $(10^n):(10^c)$ complex. The optimum pH value for this reaction lies in the range of 5.5 – 6.0. Any deviation from this interval increases the amount of by-products as a result of modification of heterocyclic bases with carbodiimide. In addition, the formation of an adduct of the oligonucleotide (10^c) with carbodiimide involving the carboxy group took place as a side process.

In the alternative version, the reaction between the chains of the 5′-d(TAATGC*ATTA)/3′-d(ATTAC*GTAAT), $(10^n):(10^n)$, duplex was carried out with glutaraldehyde. The Schiff's base formed initially was reduced with sodium borohydride to give a stable secondary amine. The yield of the target, cross-coupled duplex was brought to 20%.



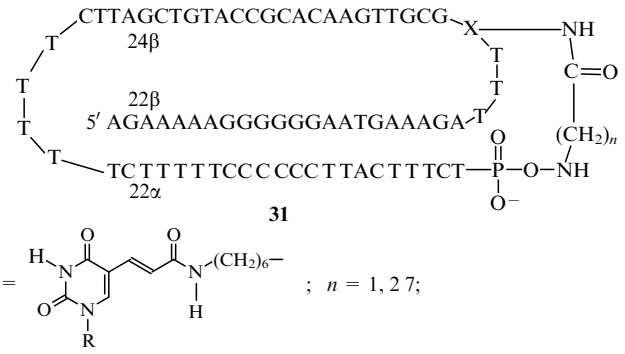
R^1 and R^2 are fragments of oligonucleotide chains.

The low yield can be explained by several reasons. First, besides the aliphatic amino groups that have been introduced, oligonucleotides contain other reaction centres capable of reacting with glutaraldehyde (*e.g.*, exocyclic amino groups of heterocyclic bases). Second, it is difficult to select conditions where two aliphatic amino groups of the modified duplex will react with the same molecule of glutaraldehyde, since the reaction is carried out with an excess of the aldehyde. Third, most probably, the optimum mutual arrangement of reactive amino groups was not realised, since they were localised in adjacent base pairs. Preliminary molecular modelling of the optimum mutual arrangement of reactive groups in the reaction centre has now become a standard procedure.

At the same time, it was demonstrated that oligonucleotides containing an aliphatic amino and a carboxy group can be used for the synthesis of DNA duplexes with an amide bond between the chains.[13] It must be noted that the system was designed in such a way as to enable polymerisation of the duplex with covalently

cross-linked chains under the action of DNA-ligase. This gives a DNA-like polymer containing a repeating consensus Pribnow sequence, which is the main binding site for RNA-polymerase in the initiation of transcription. This requires unwinding of the DNA double helix with the formation of an open complex. As the presence of cross-links between DNA chains prevents unwinding, these compounds can be used for the study of initiation of transcription and, apparently, for influencing this process.

Gottikh *et al.*[111] have synthesised an unusual oligonucleotide derivative **31** with covalently cross-linked chains. This compound has a 'snail-like' structure and represents a chimeric oligonucleotide comprising an α-nucleotide cluster. Unlike ordinary DNA, the α- and β-oligonucleotides form a parallel duplex, which allows the chimeric oligonucleotide to occur as an intramolecular duplex. The cross-link is formed between the amino group attached through a spacer to a thymidine analogue in a single-stranded region and the 3′-terminal phosphate of the α-fragment. As expected, no reaction occurred directly between these groups because of steric hindrances. Therefore, an amino acid $NH_2-(CH_2)_n-COOH$ (n = 1, 2, 7) was coupled with the terminal phosphate of the α-fragment activated in the form of *N*-hydroxybenzotriazole ester. Then the reaction was carried out between the incorporated carboxy and amino groups using water-soluble CDI as the coupling reagent.



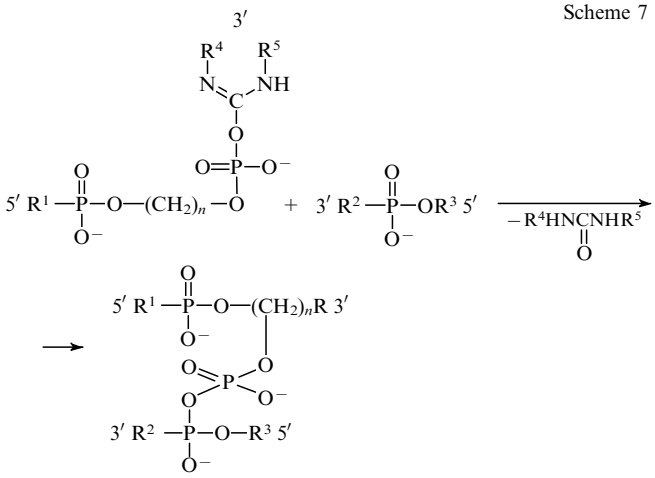
R is the deoxyribose residue.

Owing to the presence of an interchain cross-link in compound **31**, this is highly resistant to nuclease degradation. It has been shown that RNAase H cleaves an oligoribonucleotide in the chimeric duplex involving compound **31**. The study [111] illustrates the possibility of conferring useful properties to a compound by cross-linking of oligonucleotide chains. Presumably, the oligonucleotide derivative **31** can be effectively used for the regulation of RNA expression.[111]

A method for the synthesis of DNA duplexes in which the chains are cross-linked through a substituted pyrophosphate group was developed (Scheme 7).[96] This involves the reaction between the activated terminal phosphate group linked to an oligonucleotide through a spacer and the spatially close internucleotide phosphate group of the complementary oligonucleotide. The activation of the terminal phosphate group was carried out with water-soluble CDI.

The DNA duplexes containing a trisubstituted pyrophosphate internucleotide group in one of the chains are stable in aqueous buffer solutions (pH 6.0–8.75) in the absence of strong nucleophiles. At the same time, they are readily and selectively cleaved at this group with nucleophilic reagents, such as *N*-methylimidazole, amines, and some amino acids. In this case, the nucleophilic substitution reaction involves both phosphate groups of the trisubstituted pyrophosphate group and is accompanied by a cleavage of the interchain bond and transfer of one of the duplex residues to the nucleophile.[96] In contrast with DNA duplexes containing a disubstituted pyrophosphate internucleotide group in one of the chains, the reaction of nucleophilic substitution at the phosphorus atom in the DNA duplexes containing a trisubstituted pyrophosphate interchain internucleotide group can occur with

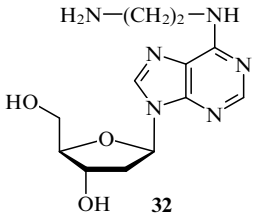
Scheme 7

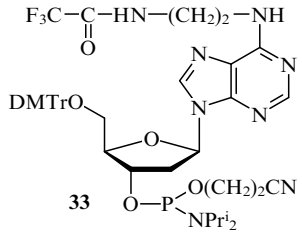


R^1-R^3 are fragments of oligonucleotide chains; n = 3, 4, 7, 12.

the involvement of both di- and tri-substituted phosphate groups to yield a complex mixture of oligonucleotides. Obviously, such DNA duplexes will be unstable in cell cultures. The method considered allows the synthesis of DNA duplexes with a reactive group between the chains. The interest in the synthesis of such compounds is connected with the possibility of their application for binding with proteins.

A method of preparation of cross-linked duplexes involving deoxyadenosine residues was proposed.[112] N^6-(2-Aminoethyl)-deoxyadenosine (**32**) is converted into the phosphoramidite derivative **33**, which can be incorporated at any predetermined position of an oligonucleotide. N^6-(2-Aminoethyl)deoxyadeno-sine **32** was obtained by treatment of 6-iododeoxyadenosine with ethylenediamine, and the amino group was protected by subsequent treatment of the reaction product with *S*-ethyltrifluorothioacetate. The trifluoroacetyl protection is ideally suited for the incorporated aliphatic amino group.





A dinucleoside monophosphate containing the N^6-modified deoxyadenosine **32** with an aliphatic amino group was synthesised.[112] The synthesis was carried out by the phosphoramidite method in solution; the yield was 95%. Higher oligonucleotides were not synthesised; however, this method is a model study aimed at the incorporation into oligonucleotides of compound **33** as a potential tool for interchain cross-linking.

The oligonucleotide synthesis involving the phosphoramidite derivative **33** was free of side reactions.[112] However, the effectiveness of this modification for the synthesis of duplexes with covalently cross-linked chains is unclear. It was proposed [112] to perform the reaction between the chains in the following way:[112] after hybridisation with the template, the aliphatic amino group has to be localised in the vicinity of the pyrimidine base of the complementary chain. The reaction of this base with the bisulfite ion at positions 5 and 6 facilitates the subsequent nucleophilic attack at position 4. The aliphatic amino group of N^6-(2-aminoethyl)deoxyadenosine **32** will act as the nucleophile. This method of interchain cross-linking must not bring significant disturbance to the structure of the double helix. A disadvantage of this method is its inapplicability to cross-linking of oligonucleotides to extended DNA, since in the presence of the bisulfite anion

deoxycytidine in single-stranded regions can undergo side conversions into 2′-deoxyuridine, which is undesirable.

VII. The use of sulfhydryl derivatives for the synthesis of cross-linked nucleic acid duplexes

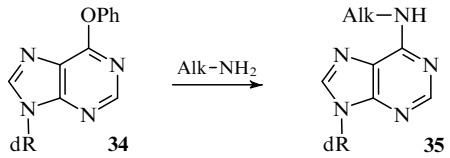
Oligonucleotides containing SH-groups have received wide acceptance in recent years. They not only allow the synthesis of various types of oligonucleotide derivatives, but also of cross-linked DNA duplexes in the predetermined positions of the double helix. The so-called 'convertible nucleoside approach' of oligonucleotide synthesis has become widespread.[113] It consists in the incorporation of a precursor of the modified nucleoside carrying an SH group into the oligonucleotide that is synthesised. This precursor carries a different group, which is usually more resistant to synthetic and post-synthetic treatments. Later (sometimes even after hybridisation with the template), the precursor is converted into the target nucleoside with a thio group, which becomes involved in the reaction between the chains.

This strategy has considerable promise for it permits an access to various types of modified oligonucleotides from the same precursor because the incorporation of a functionalised linker is performed after completion of the oligonucleotide synthesis. This approach does not envisage the synthesis of a new starting monomer or a new modified oligonucleotide in order to change the length of the linker or the nature of the functional group. Another advantage of this method is that it allows a post-synthetic addition to oligonucleotide derivatives of ligands (oligopeptides, *etc*.) unable to withstand the treatments used in the oligonucleotide synthesis.

A characteristic feature of this method is that the monomer used in the synthesis contains a leaving group, which undergoes nucleophilic substitution. When a bifunctional nucleophile is used, reactive oligonucleotide derivatives are formed. At the same time, the bifunctional nucleophile fulfils the function of a linker.

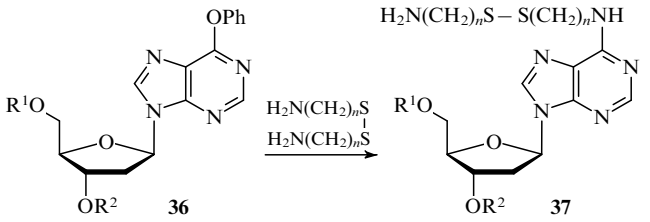
In order to attain the maximum efficiency, a number of conditions must be observed.[113] The nucleoside in question must be stable under conditions of automated synthesis of DNA and be rapidly and completely converted into a reactive derivative by a nucleophile. Incorporation of a linker must cause minimum disturbance to the structure of the DNA double helix. As a rule, this process should not involve the terminal oligonucleotide units so as to retain the possibility of further enzymic reactions, such as radioactive labelling, elongation of the oligonucleotide, or its incorporation into a plasmid.

Ferentz and Verdine[113] have 'converted' deoxyinosine into modified deoxyadenosine. Earlier, these authors had shown[114] that O^6-phenyldeoxyinosine **34** was quantitatively converted into N^6-alkyldeoxyadenosine **35** on aminolysis with alkylamines.



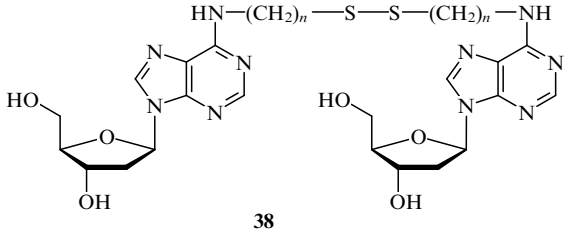
dR is the 2′-deoxyribose residue.

O^6-Phenyldeoxyinosine 3′-phosphoramidite was synthesised and introduced into automated synthesis to give the oligonucleotide 5′-d(GCGAI*TTCGC) **36**, where I* is the residue of the nucleoside **34**. After completion of the synthesis, the oligonucleotide **36** was treated with bis(2-aminoethyl) disulfide or bis(3-aminopropyl) disulfide, resulting in compound **37**.



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

The oligonucleotide **37** obtained was self-complementary and formed a duplex in which the modified bases are localised in neighbouring units. Treatment of the duplex with dithiothreitol led to the conversion of the disulfide into thiol and then the disulfide bond between the chains was closed by oxidation with the atmospheric oxygen. This binding is reversible. Repeated treatment of the duplex with dithiothreitol results in the cleavage of the interchain cross-link. It was shown[113] that the products of enzymic hydrolysis of the DNA duplex with covalently cross-linked chains contain the dinucleoside **38**; the shape of the helix of the DNA duplex corresponds to the B-form of DNA, while the modified bases form ordinary Watson – Crick pairs.



n = 2, 3.

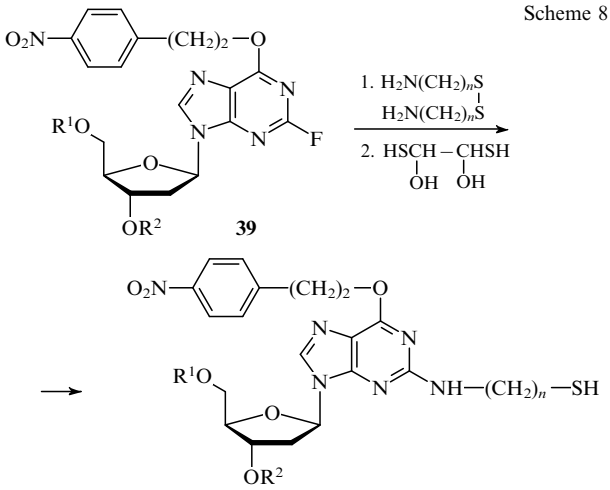
Computer simulation has revealed that the linker is localised in the large groove of DNA and does not cause any significant distortions in the spatial structure of the duplex. The original duplex and the duplex with covalently cross-linked chains have similar melting curve profiles. This indicates that the cross-linking between the chains does not disturb the duplex structure itself. Under these conditions, the melting point of the duplex with covalently cross-linked chains (75 °C) was by 18 °C higher than that for the original duplex in 1 M NaCl, whereas in 0.1 M NaCl it was by 30 °C higher.

Thus, the formation of a disulfide bond between the chains has a number of advantages. First, it increases significantly the stability of the DNA duplex. Second, it does not result in noticeable distortions in the DNA structure. DNA duplexes with chains linked by disulfide bonds can be used for the study of processes involving the unwinding of DNA by enzymes, *e.g*., transcription, replication, and recombination.

Yet another approach to the synthesis of a DNA duplex with covalently cross-linked chains by the 'convertible nucleoside approach' has been described.[14] O^6-[2-(*p*-Nitrophenyl)ethyl]-2-fluorodeoxyinosine[115,116] was introduced into a self-complementary 18-membered oligonucleotide in the process of oligonucleotide synthesis. Treatment of this oligonucleotide with bis(aminoalkyl) disulfides (Scheme 8) or *N*-(2-methylthioethyl)-amine **39** led to the substitution of the fluorine atom for *N*-(ω-aminodithioalkyl)amino group or the *N*-(2-methylthioethyl)-amino group.[14]

Following treatment with dithiothreitol, the reaction between the chains was carried out in the presence of atmospheric oxygen. The presence of a cross-link was proved by an analysis of a correlation between the melting point of the duplex with covalently cross-linked chains and the ability of these cross-links to induce entropic stabilisation of DNA.[117]

Scheme 8



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

The melting point of the duplex cross-linked by a bis(2-aminoethyl) disulfide linker ($n = 2$) was lower than that of the duplex cross-linked with a bis(3-aminopropyl) disulfide linker ($n = 3$). Presumably, a shorter linker causes greater torsional tension in the DNA structure. Comparison with an oligonucleotide duplex containing the (2-methylthioethyl)amino groups revealed that this modification as such has little effect on the melting point of the duplex.[14]

The DNA duplexes thus synthesised with covalently cross-linked chains were used for the study of DNA methylation. This process plays an important role in the development of mammals and in the resistance of bacteria to viral infections.[14] DNA-5-cytosine-methyltransferase (CMTase) catalyses transfer of the methyl group from *S*-adenosyl-L-methionine to cytosine. The enzyme recognises the symmetrical site of DNA, *viz*. 5′-d(GGCC)/3′-d(CCGG), and transfers the methyl group to the internal deoxycytidine residue of each chain. It was assumed[14] that if the methylation site contains an interchain cross-link, the unwinding of the DNA double helix as an element of the enzyme action mechanism will be hampered considerably during the transfer of the methyl group.

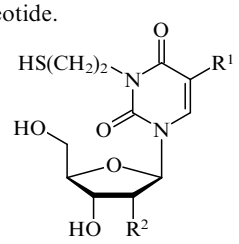
It turned out that the enzyme actually induces a distortion of the double helix. In order to establish a correlation between the effect of interchain cross-linking and the recognition of the methylation site by the enzyme, the equilibrium binding has been studied under non-catalytic conditions. It was found that the duplex with a dithiaoctamethylene bridge between the chains binds to the enzyme with the same efficiency as does the non-modified duplex, whereas the duplex with a dithiahexamethylene bridge binds 43 times more weakly. Thus, the presence of a cross-link between the chains has no effect on the enzyme binding with DNA (at least, for a longer linker). Moreover, in the latter case cross-linking between the chains results in allosteric activation of the protein binding. Most probably, this is due to the fact that the presence of an interchain cross-link generates torsional tension, which decreases the thermodynamic barrier to DNA unwinding. However, in contrast to cross-linking, methylation occurs much more slowly in comparison with the natural duplex. Nevertheless, under conditions of substrate saturation both duplexes with covalently cross-linked chains serve as substrates for CMTase and the degree of methylation is 10% relative to that of the non-modified duplex. This study illustrates the applicability of DNA duplexes with covalently cross-linked chains for gaining insight into the mechanisms of action of DNA-recognising enzymes.

Erlanson *et al.*[14] have carried out computer simulation of the spatial structure of DNA duplexes with covalently cross-linked chains. The data obtained were compared with the results of X-ray analysis.[118] The molecular dynamics methods were used to study the effect of the interchain cross-linking on the secondary structure of DNA. This made it possible to estimate the extent of distortions induced by cross-linking in the duplex structure and to evaluate its effect on the dynamic behaviour of DNA. Spatial structures of the self-complementary decanucleotide 5′-d(GCGAATTCGC) in the form of a duplex and of two modified duplexes with the same primary structure and a cross-link between the amino groups of deoxyadenosine residues in the central A·T pair were studied. The lengths of the linkers were six (dithiahexamethylene) or eight dithiaoctamethylene) atoms (the nitrogen atoms were not taken into account).

It was shown[118] that the canonical B-form of DNA and the Watson–Crick type of base pairing are preserved by all DNA duplexes in question. The parameters of the duplex helices with covalently cross-linked chains, of the large and small grooves, and of the base pairs, as well as the torsional angles of the sugar–phosphate backbone are also mostly similar to the corresponding parameters for the crystalline structure of the non-modified duplex. Yet, the presence of cross-links between the chains endows certain disorder in the geometry of the double helix. The deviations observed are insignificant: no more than 1.2 Å from the canonical B-form of DNA and no more than 1.3 Å from the experimentally determined parameters for the crystal of a non-modified duplex. These changes are more substantial with the shorter interchain linker. The dynamic properties of the duplexes with covalently cross-linked chains are also similar to those of the non-modified duplex. Certain changes are observed in a region adjacent to the cross-linking site and are especially pronounced in the case of a shorter linker. Structural distortions of general character are manifested in decreased mobility at the cross-link site irrespective of the linker length.

Thus, the changes in structural parameters originate mostly from the modification of the amino groups of deoxyadenosines rather than from the presence of a cross-link between the chain. On the whole, the presence of a cross-link between the chains induces insignificant changes in the duplex structure;[118] therefore, these duplexes may be used for the study of proteins that interact with nucleic acids.

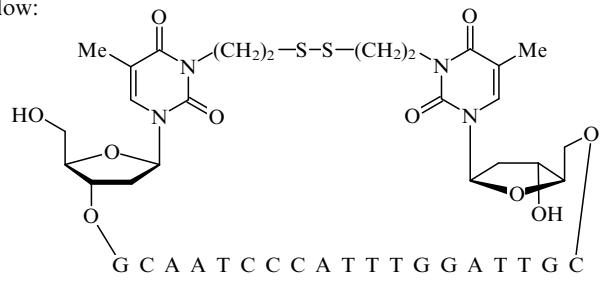
Yet another original procedure for obtaining DNA duplexes with covalently cross-linked chains on the basis of thiols has been proposed.[119] Two residues of N^3-(2-mercaptoethyl)thymidine **40** were incorporated into an oligodeoxyribonucleotide and two residues of N^3-(2-mercaptoethyl)uridine **41**, into an oligoribonucleotide.



40, **41**

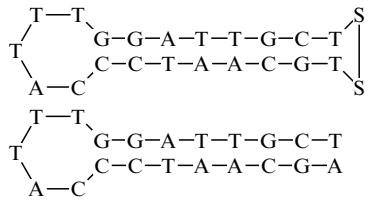
40: R^1 = Me, R^2 = H;
41: R^1 = H, R^2 = OH.

The phosphoramidite derivative of compound **40** was attached to the 5′- and 3′-ends of an oligodeoxyribonucleotide in the automated synthesis. The disulfide bond between the ends of a hairpin was closed by oxidation of the thio groups with atmospheric oxygen. The structure of the interchain cross-link at the end of the hairpin formed by two residues of compound **40** is shown below:



The structures of a covalently closed hairpin and its natural analogue are shown in Scheme 9. Their conformational structure was studied by NMR spectroscopy. It was found that the 8-membered double-stranded region in both hairpins has a canonical B-form; structurally both hairpins are completely isomorphic.

Scheme 9



In this case, the presence of an interchain cross-link neither influences the spatial structure of the hairpin nor interferes with the native B-conformation of the double helix. It should be noted that although in this study the cross-link was formed at the end of the DNA duplex, the modified nucleoside could be incorporated into any predetermined position of an oligonucleotide. This method can thus be applied for obtaining a set of various cross-linked oligonucleotide duplexes.

Thus, the synthesis of duplexes with covalently cross-linked chains based on sulfhydryl derivatives has a number of advantages over the use of alkylating agents, which do not ensure sequence-specific cross-linking of chains. The latter circumstance restricts their application in the engineering of interchain cross-links.

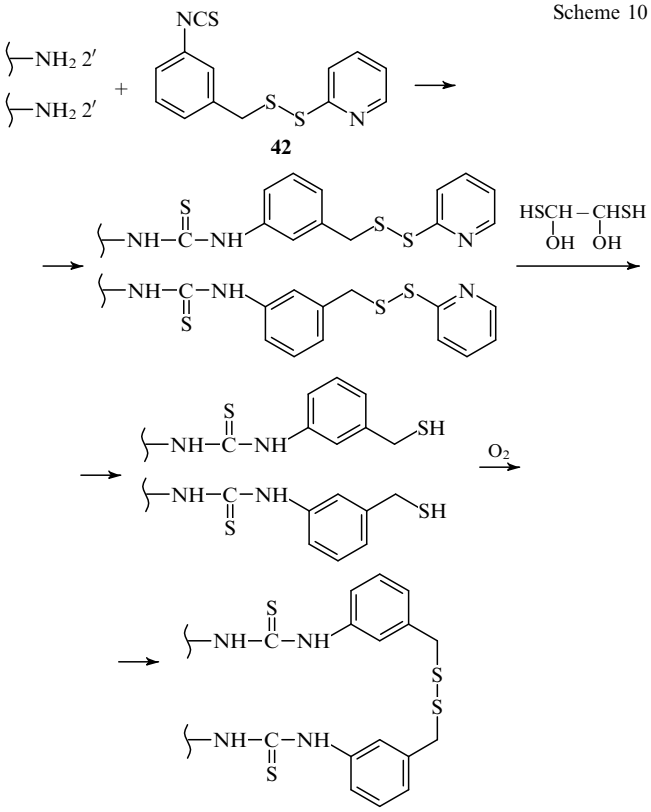
The disulfide method was used to cross-link fragments of the ribozyme RNA chain.[17] The reaction was carried out between spatially close (but not complementary) regions of the RNA chain. Hence, here we do not deal with the formation of cross-links between the RNA chains within the duplex; however, this method is so elegant that it deserves special mention. Eckstein *et al.*[17] took advantage of modification of the 2′-positions of certain sugar residues. Oligoribonucleotides with incorporated 2′-amino-2′-deoxyribonucleotides were synthesised for this purpose. Synthesis of 2′-amino-2′-deoxyribonucleosides has been described earlier.[120–126]

The sulfhydryl groups were introduced as follows. First, the oligoribonucleotides were treated with the bifunctional reagent **42** (Scheme 10) containing the isothiocyanate and a protected sulfhydryl group. It was shown that compound **42** selectively reacts with the 2′-amino groups of the modified ribonucleotides.[17]

The reaction between the fragments of the ribozyme chain was initiated by dithiothreitol, which deprotected the sulfhydryl groups. This enabled the reaction between the ribozyme chains under the action of atmospheric oxygen (Scheme 10). The presence of a cross-link between the chains was confirmed by limited alkaline hydrolysis. One cleavage of the ribozyme chain at the site between the modified residues gave a single product with the same molecular weight as that of the original ribozyme.

Ribozymes with covalently cross-linked chains obtained by the above-described procedure were used to study the tertiary structure of the ribozyme in relation to the catalytic mechanism of RNA cleavage. The parameters of the ribozyme tertiary structure obtained earlier by X-ray analysis (Scheme 11, **A**) and fluorescence resonance energy transfer (FRET) solution measurements (Scheme 11, **B**) were different. [Ribbon representation of the tertiary structures of the hammerhead ribozyme based on (**A**) X-ray crystallography and (**B**) FRET solution measurements. Ribozymes and substrates are coloured black and light gray, respectively. Residues that are close in space according to the X-ray structure and the FRET model (dark gray) are connected by solid lines and the broken lines represent the corresponding long distances in the other model.] At the same time, both putative structures had identical general geometry. The difference between them lay in the orientation of the two helical regions of RNA relative to each other (helices I and II, Scheme 11). In order to compare these two spatial models of the ribozyme and to choose

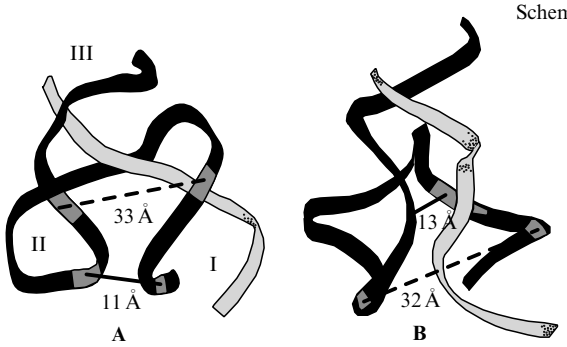
Scheme 10



42

the model corresponding to the genuine biologically active structure, synthetic ribozymes with covalently cross-linked chains were used.

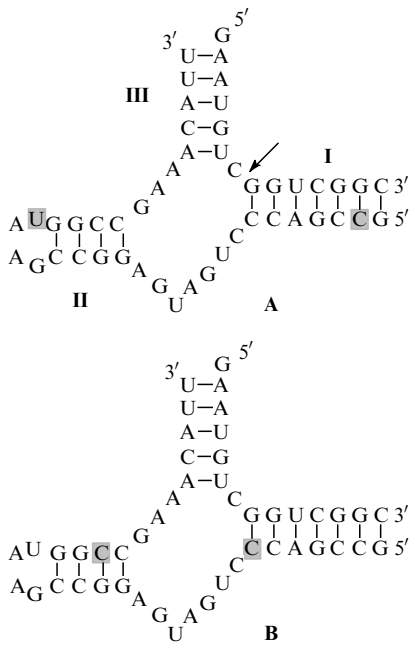
Scheme 11



The authors [17] have succeeded in stabilising the spatial structure of the ribozyme according to each of the putative models. After hybridisation of the modified ribozymes with a non-hydrolysable substrate analogue containing a deoxynucleoside in the reaction centre, the reaction mixtures were treated with oxygen for 30 h. This permitted a highly efficient reaction between the chains irrespective of the localisation of the modified units in the ribozymes (Scheme 12). Roman numerals indicate the number of the helices. An arrow indicates the site of cleavage. Residues contained a 2′-amino function and utilised for cross-linking are filled by dark gray colour. The yield of the ribozymes with covalently cross-linked chains after all purification procedures was 15%.

The cross-links between the chains stabilised the spatial structure of the ribozyme in accordance with the predetermined one, which made it possible to measure the catalytic activity of two synthetic ribozymes with different spatial structures. As expected, the catalytic activity of the ribozyme depends on the localisation of the cross-link between the chains. The cross-link between the adjacent regions of the helices corresponding to the X-ray model (Schemes 11, 12, **A**) had practically no effect on the catalytic

Scheme 12

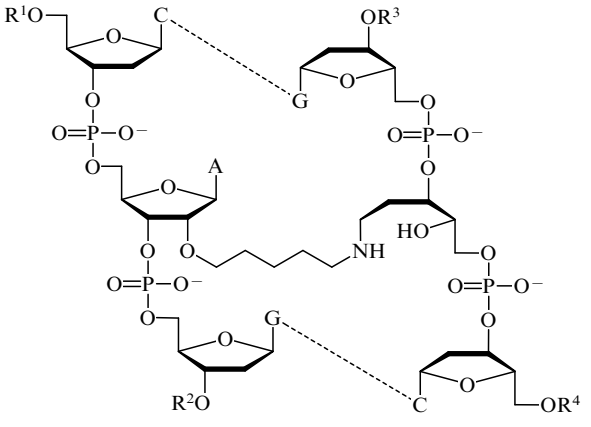


activity of the ribozyme. On the other hand, the cross-link between the adjacent nucleotides according to the other model (Schemes 11, 12, **B**) decreased the rate of the ribozyme-induced cleavage of the corresponding substrate by three orders of magnitude. It was thus concluded that the X-ray model corresponds to the real structure and that the previous methods that have previously been used in the study of the tertiary structure of RNA are inadequate. The results obtained by Eckstein *et al.*[17] have provided the basis for further studies on ribozymes.[127]

VIII. Synthesis of cross-linked nucleic acid duplexes using sugar moieties of nucleosides

In contrast to studies devoted to the modification of heterocyclic bases and phosphate groups of oligonucleotides, reports on the reactions between chains involving sugar residues are scarce.

A method for the synthesis of a duplex was proposed,[15] which entails the formation of a cross-link between the amino group attached through a spacer to position 2 of the sugar residue of adenosine, on the one hand, and the aldehyde group in a non-nucleoside insert of the complementary chain, on the other hand.



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

Synthesis of this type of compound has been patented.[16] In this case, the aldehyde group results from enzymic elimination of uracyl by uracyl-DNA-glycosylase from 2′-deoxyuridine incorporated into the oligonucleotide. The elimination of the heterocyclic base is accompanied by the sugar ring opening and the formation of the aldehyde group. Treatment of the latter with the primary amine gives a Schiff's base, which is reduced to give a stable secondary amine. However, enzymic treatment does not yet allow the preparative synthesis of these compounds, which restricts their application as therapeutic agents.

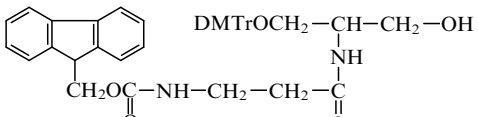
As mentioned above, the main drawback of many procedures employed in the synthesis of cross-linked DNA duplexes is their low specificity. It is the specific cross-linking of DNA chains that allows the development of the most convenient tools for the study of protein-nucleic acid interactions. Therefore, in recent years most studies have been aimed at the elaboration of methods of site-specific cross-linking of chains in synthetic DNA duplexes.

A series of publications [128–131] are devoted to elaboration of a new effective method for the synthesis of cross-linked DNA duplexes that involves sugar residues of nucleosides. This method required the use of oligonucleotides modified at the carbohydrate fragment and containing a primary aliphatic amino group or a carboxy group. The existing procedures for incorporation of modified units into oligonucleotides permit introducing them into any predetermined position of the oligonucleotide chain or incorporate several modified units into the same oligonucleotide by a standard protocol of solid-phase phosphoramidite oligonucleotide synthesis.

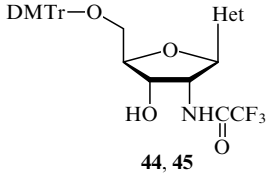
The reaction between the primary amino group and the carboxy groups selected for the cross-linking of the duplex chains is activated by water-soluble CDI and results in the formation of an amide bond between the chains.

In the early 90's, several methods for the synthesis of 2′-amino-2′-deoxynucleotides were developed.[120–126] High reactivity of the 2′-amino groups of 2′-amino-2′-deoxynucleosides in acylation reactions[120–122] has led to an idea of formation of a cross-link between the 2′-amino group in one chain and the carboxy group derived upon acylation of the 2′-amino-2′-deoxynucleoside of the complementary chain by a dicarboxylic acid anhydride. However, preparation of purine 2′-amino-2′-deoxynucleosides by transglycosylation (substitution of a heterocyclic base) of preformed 2′-amino-2′-deoxyuridine [123, 124] inevitably resulted in the formation of a complex mixture of products and extremely low yield of the target product.

Thus, a non-nucleoside insert was used instead of purine 2′-amino-2′-deoxynucleosides. On the one hand, this disturbs the hydrogen bonds in the pair in which one heterocylic base is missing. On the other hand, the use of this nucleoside substitution has a number of benefits. Compound **43** can be synthesised by a combination of simple high-yielding chemical reactions.[128, 129] The distance of three carbon atoms between the phosphate groups, which is characteristic of natural oligonucleotides, is preserved in the structure of the non-nucleoside insert. This compound is resistant to synthetic and post-synthetic treatments of the oligonucleotide.



43

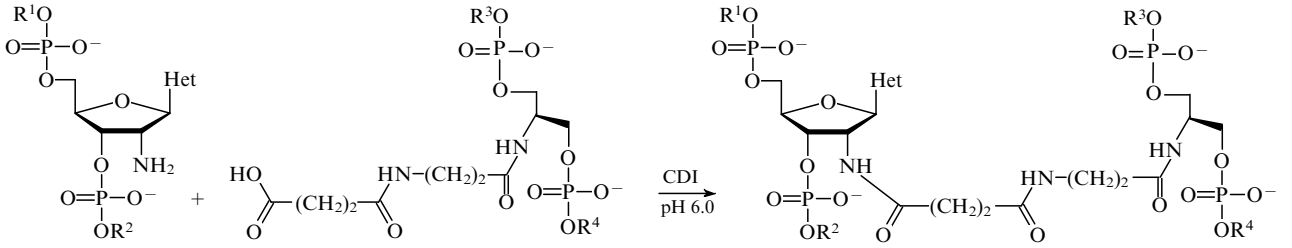


Het = U (**44**); C^{Bz} (**45**).

44, **45**

Synthesis of specifically protected pyrimidine 2′-amino-2′-deoxynucleosides **44** and **45** has been described.[120, 121] The reaction between the chains of a modified duplex is performed in the following steps. First, a carboxy group is introduced into one chain. This is achieved by acylating the amino group incorporated

Scheme 13



Het = U, C^{Bz}; $R^1 - R^4$ are fragments of oligonucleotide chains.

into the oligonucleotide with a dicarboxylic anhydride. Molecular simulation allowed one to select succinic anhydride as the acylating reagent. Then the reaction between the chains was conducted in a morpholinoethyl sulfonate buffer (pH 5.0 – 6.7) for 1 – 3 days at 0 – 20 °C using water-soluble CDI as a condensation reagent (Scheme 13).

Previously, it had been demonstrated [128, 129] that the reaction between the strands of modified DNA duplexes with a primary aliphatic amino group in one of them and a carboxy group in the other proceeds much more efficiently with the free amino group of 2′-amino-2′-deoxycytidine. This is explained by the fact that the pK_a value for this amino group (6.2) [121] is much lower than that for the amino group of aliphatic amines, so this group exists mainly in a deprotonated form under the reaction conditions.

The effects of some other factors, such as the mutual arrangement of modified units, their number, nucleotide environment, thermodynamic stability of the precursor duplex, *etc*., on the efficiency of the reaction between the chains, were studied.[130, 131] Under optimum conditions, the yield of the corresponding DNA duplexes with covalently cross-linked chains reached 85% for certain systems.

The yields of the target products differ considerably for a 20- and a 25-membered systems (26% and 85%, respectively). This is accounted for by the fact that the efficiency of the reaction increased with an increase in the melting point of the original modified duplex.

Depending on the nucleotide environment of the modified units, the yield of the target product varied from 50% to 85%.[130, 131] Substitution of 2′-amino-2′-deoxyuridine for 2′-amino-2′-deoxy-*arabino*-adenosine does not affect significantly the yield of the DNA duplex with covalently cross-linked chains.

The thermodynamic stability of DNA duplexes with incorporated modified units was in all cases lower than that of non-modified duplexes of the same primary structure. The decrease in the melting point depends on the length and the composition of the DNA duplex. In contrast, for duplexes with covalently cross-linked chains the melting point was higher than that for the corresponding modified and non-modified duplexes. The increase in the melting point was 4 – 8 °C relative to natural analogues, while that for the corresponding modified duplexes was 15 – 33 °C.[128, 129]

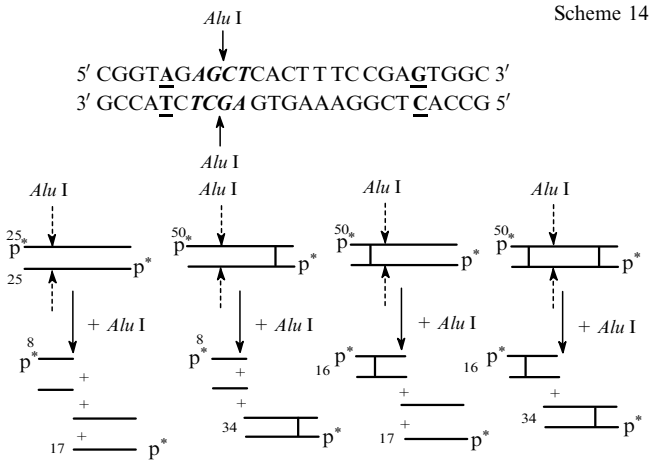
Duplexes with covalently cross-linked chains are much more resistant to the action of a mixture of snake venom phosphodiesterase and alkaline phosphatase than those devoid of interchain cross-links. Under conditions where natural and modified duplexes are completely hydrolysed to give nucleosides, duplexes with covalently cross-linked chains maintain their structure by no less than 70%. Such a low level of hydrolysis under these conditions is due to the presence of a cross-link between the chains.

Comparative kinetics of enzymic hydrolysis of a 25-membered duplex with covalently cross-linked chains and of a non-modified duplex of the same primary structure by snake venom phosphodiesterase was carried out.[132] Under selected conditions, enzymic hydrolysis of single-stranded oligonucleotides and a non-modified duplex occurred at a high rate. Single-stranded oligonucleotides were completely hydrolysed to give nucleotides in 5 min, whereas the duplex with covalently cross-linked chains was not completely hydrolysed under identical conditions even in 3 h. Relatively easy splitting of two or three units was observed together with accumulation of products of partial hydrolysis with chains shortened by no more than 3 – 8 units. After scission of eight units, the reaction is appreciably retarded. Thus, the presence of a cross-link between the chains causes strong inhibition of the effect of snake venom phosphodiesterase.

In order to prove the ability of compounds obtained to serve as substrates of enzymes binding with unique nucleotide sequences of DNA and to confirm their structure, DNA duplexes with covalently cross-linked chains were site-specifically cleaved with restriction endonuclease *Alu* I.[130, 131] The primary structure of covalently cross-linked DNA duplexes corresponded to the 25-membered non-modified DNA duplex; its nucleotide sequence is shown in Scheme 14. The recognition site of restriction endonuclease *Alu* I is given in italics; the sites of incorporation of the corresponding modified units **43** and **44** are underlined.

The position of cross-links between the chains relative to the recognition site of restriction endonuclease *Alu* I in DNA duplexes determines the composition of the hydrolysate. The formation of oligonucleotides of different lengths resulting from hydrolysis of DNA duplexes with covalently cross-linked chains and the non-modified duplex is illustrated by Scheme 14.

Scheme 14



p* is ^{32}P-phosphate.

The composition of reaction mixtures for all DNA duplexes is in agreement with the theoretical Scheme 14. Thus, it has been shown that the presence of cross-links between the chains in synthetic DNA duplexes does not prevent their cleavage with restriction endonuclease *Alu* I. The formation of a series of products of a definite length confirms the presence and position of cross-links between the chains in the synthetic DNA duplexes.

The experimental results [128 – 131] suggest that DNA duplexes with covalently cross-linked chains can be used as efficient ‘traps’ for proteins binding with nucleic acids.

IX. Conclusion

The data presented in this review testify to the existence of efficient methods for preparing DNA duplexes with covalently cross-linked chains. Cross-linking of DNA chains can be achieved either nonspecifically (statistically) or at predetermined positions of the double helix.

Owing to existing methods, the synthesis of these compounds is no longer a very complicated task. DNA duplexes with covalently cross-linked chains have become a convenient tool for the study of protein – nucleic acid interactions. Convincing evidence for the significant potential of this approach can be derived from the investigations cited in the review.

References

1. V V Vlasov, M A Grachev, O I Lavrik *Afinnaya Modifikatsiya Biopolimerov* (Affine Modification of Biopolymers) (Ed. D G Knorre) (Novosibirsk: Nauka, 1983)
2. D G Knorre, V F Zarytova, A G Badashkeeva, O S Fedorova *Biotekhnologiya (Itogi Nauki i Tekhniki)* [Biotechnology (Advances in Science and Engineering Ser.)] (Ed. Z A Shabarova) (Moscow: Izd. VINITI, 1991) Vol. 37
3. N I Grineva *Vestn. Akad. Med. Nauk SSSR* **2** 83 (1981)
4. P B Dervan *Science* **232** 464 (1986)
5. L H Hurley *J. Med. Chem.* **32** 2027 (1989)
6. Z A Shabarova *Biochimie* **70** 1323 (1988)
7. M Takasugi, A Guendouz, M Chassignol, J L Decout, J Lhomme, N T Thuong, C Helene *Proc. Natl. Acad. Sci. USA* **88** 5602 (1991)
8. J Goodchild *Bioconjugate Chem.* **1** 165 (1990)
9. B S Sproat, A I Lamond, B Beijer, P Neuner, U Ryder *Nucl. Acids Res.* **17** 3373 (1989)
10. I A Kozlov, M G Ivanovskaya, E A Kubareva, E M Volkov, Z A Shabarova *Mol. Biol.* **30** 503 (1996) [a]
11. J Piette, H B Gamper, A Van de Vorst, J E Hearst *Nucl. Acids Res.* **16** 9961 (1988)
12. B L Lee, K R Blake, P S Miller *Nucl. Acids Res.* **16** 10681 (1988)
13. O N Koroleva, E M Volkov, T S Oretskaya, Z A Shabarova *Bioorg. Khim.* **20** 40 (1994) [b]
14. D A Erlanson, L Chen, G L Verdine *J. Am. Chem. Soc.* **115** 12583 (1993)
15. M Manoharan, L K Johnson, P D Cook, in *International Conference of Nucleic Acids Medical Applications (Poster Abstract) Mexico, 1993* p. 1
16. Appl. WO PCT 9 318 052; *Chem. Abstr.* **121** 180 122 (1994)
17. S Th Sigurdsson, T Tuschl, F Eckstein *RNA* **1** 575 (1995)
18. R B Meyer, J C Tabone, G D Hurst, T M Smith, H Gamper *J. Am. Chem. Soc.* **111** 8517 (1989)
19. A M Belikova, N I Grineva *Izv. Sib.Otd. Akad. Nauk SSSR. Ser. Khim. Nauk* **3** 79 (1966)
20. N I Grineva, G G Karpova, S D Myzina, I Fodor, A A Baev *Dokl. Akad. Nauk SSSR* **223** 1477 (1975) [c]
21. N I Grineva, G G Karpova *FEBS Lett.* **32** 351 (1973)
22. N I Grineva, G G Karpova, L M Kuznetsova, T V Venkstern, A A Bayev *Nucl. Acids Res.* **5** 1609 (1977)
23. N I Grineva, G G Karpova *Mol. Biol.* **8** 832 (1974)
24. L Z Benimetskaya, L M Gerasimova, N I Grineva, G G Karpova *Mol. Biol.* **12** 988 (1978) [a]
25. L Z Benimetskaya, L M Gerasimova, G G Karpova, A I Krutilina, N I Grineva *Biokhimiya* **44** 803 (1979) [d]
26. V G Budker, N D Kobets, I E Kollekchonok, G G Karpova, N I Grineva *Mol. Biol.* **14** 400 (1980) [a]
27. G G Karpova, N P Pichko, T A Chimitova, N I Grineva *Mol. Biol.* **13** 1012 (1979) [a]
28. G M Dymshits, G V Rumyantseva, L A Frumgarts, A D Gruzdev, G A Zainiev, I E Shilova, G G Karpova, V I Yamkova, N I Grineva *Mol. Biol.* **15** 66 (1981) [a]
29. J Summerton, P A Bartlett *J. Mol. Biol.* **122** 145 (1978)
30. M A Grachev, A A Mustaev, S I Oshevski *Nucl. Acids Res.* **8** 3413 (1980)
31. V V Vlasov, A A Gall', A A Godovikov, V F Zarytova, I V Kutyavin, I P Motovilova, G V Shishkin *Biokhimiya* **48** 1244 (1983) [d]
32. N K Kochetkov, E I Budovskii, E D Sverdlov, N A Simukova, M F Turchinskii, V N Shibaev *Organicheskaya Khimiya Nukleinovykh Kislot* (Organic Chemistry of Nucleic Acids) (Moscow: Khimiya, 1970)
33. Z A Shabarova, A A Bogdanov *Khimiya Nukleinovykh Kislot i Ikh Komponentov* (The Chemistry of Nucleic Acids and Their Components) (Moscow: Khimiya, 1978)
34. K W Kohn *Molecular Aspects of Anticancer Drug Action* (Eds S Neidle, M J Waring) (London: Macmillan, 1983) p. 315
35. J P Bachellerie, J F Thompson, M R Wegnez, J E Hearst *Nucl. Acids Res.* **9** 2207 (1981)
36. S Peckler, B Graves, D Kanne, H Rapoport, J E Hearst, S H Kim *J. Biol. Chem.* **162** 157 (1982)
37. U Pieles, B S Sproat, P Neuner, F Cramer *Nucl. Acids Res.* **17** 8967 (1989)
38. U Pieles, U English *Nucl. Acids Res.* **17** 285 (1989)
39. P S Song, K J Tapley *Photochem. Photobiol.* **29** 1177 (1979)
40. J E Hearst *Annu. Rev. Biophys. Bioeng.* **10** 69 (1981)
41. B H Johnston, M A Johnson, C B Moore, J E Hearst *Science* **197** 906 (1977)
42. Y Shi, J E Hearst *Biochemistry* **26** 3786 (1987)
43. Y Shi, J E Hearst *Biochemistry* **26** 3792 (1987)
44. Y Shi, J E Hearst *Biochemistry* **25** 5895 (1986)
45. D Kanne, K Straub, J E Hearst, H Rapoport *J. Am. Chem. Soc.* **104** 6754 (1982)
46. G D Cimino, Y Shi, J E Hearst *Biochemistry* **25** 3013 (1986)
47. J F Thompson, J E Hearst *Cell* **32** 1355 (1983)
48. G D Cimino, H B Gamper, S T Isaacs, J E Hearst *Ann. Rev. Biochem.* **54** 1151 (1985)
49. J W Tessman, S T Issacs, J E Hearst *Biochemistry* **24** 1669 (1985)
50. E Garret-Wheeler, R E Lochard, A Kumar *Nucl. Acids Res.* **12** 3405 (1984)
51. C Shen, J E Hearst *Proc. Natl. Acad. Sci. USA* **73** 2649 (1976)
52. J Piette, J E Hearst *Proc. Natl. Acad. Sci. USA* **80** 5540 (1983)
53. Y Shi, H Gamper, J E Hearst *Nucl. Acids Res.* **15** 6843 (1987)
54. Y Shi, H Gamper, J E Hearst *J. Biol. Chem.* **263** 527 (1988)
55. Y Shi, J E Hearst *J. Mol. Biol.* **199** 277 (1988)
56. B Van Houten, H Gamper, J E Hearst *Proc. Natl. Acad. Sci. USA* **83** 8077 (1986)
57. B Van Houten, H Gamper, J E Hearst, A Sancar *J. Biol. Chem.* **261** 14135 (1986)
58. R S Cole *Proc. Natl. Acad. Sci. USA* **70** 1064 (1973)
59. J M Kean, A Murakami, K R Blake, C D Muracami, P S Miller *Biochemistry* **27** 9113 (1988)
60. B L Lee, A Murakami, K R Blake, S B Lin, P S Miller *Biochemistry* **27** 3197 (1988)
61. E Sage, E Moustacchi *Biochemistry* **26** 3307 (1987)
62. T Seiki, K Nozu, S Kondo *Photochem. Photobiol.* **27** 19 (1978)
63. J Piette, D Decuyper-Debergh, H Gamper *Proc. Natl. Acad. Sci. USA* **82** 7355 (1985)
64. M Lee, M C Roldan, M K Haskell, S R McAdam, J A Hartley *J. Med. Chem.* **37** 1208 (1994)
65. J Woo, S Th Sigurdsson, P B Hopkins *J. Am. Chem. Soc.* **115** 3407 (1993)
66. D S Bose, A S Thompson, J S Ching, J A Hartley, M D Berardini, T C Jenkins, S Neidle, L H Hurley, D E Thurston *J. Am. Chem. Soc.* **114** 4939 (1992)
67. D S Bose, A S Thompson, M Smellie, M D Berardini, J A Hartley, T C Jenkins, S Neidle, D E Thurston *J. Chem. Soc., Chem. Comm.* 1518 (1992)
68. S Th Sigurdsson, S M Rink, P B Hopkins *J. Am. Chem. Soc.* **115** 12633 (1993)
69. S Th Sigurdsson, P B Hopkins *Tetrahedron* **50** 12065 (1994)
70. J A Hartley *Biochemistry* **30** 11719 (1991)
71. J A Hartley, M D Berardini, R L Souhami *Anal. Biochem.* **193** 131 (1991)
72. S C Alley, K A Brameld, P B Hopkins *J. Am. Chem. Soc.* **116** 2734 (1994)
73. M Yuki, V Grukhin, C S Lee, I S Haworth *Arch. Biochem. Biophys.* **325** 39 (1996)

74. S P Mayalarp, R H Hargreaves, J Butler, C C O'Hare, J A Hartley *J. Med. Chem.* **39** 531 (1996)
75. R J Knox, F Friedlos, P J Biggs, W D Flitter, M Gaskell, P Goddard, L Davies, M Jarman *Biochem. Pharmacol.* **46** 797 (1993)
76. A Skladanowski, J Konopa *Biochem. Pharmacol.* **47** 2279 (1994)
77. A Van Rosmalen, C Cullinane, S M Cutts, D R Phillips *Nucl. Acids Res.* **23** 42 (1995)
78. J D Farmer, G R Gustafson, A Conti, M B Zimmt, J W Suggs *Nucl. Acids Res.* **19** 899 (1991)
79. J D Farmer, S M Rudnicki, J W Suggs *Tetrahedron Lett.* **29** 5105 (1988)
80. P B Hopkins, T J Millard, J Woo, M F Weidner, J J Kirchner, S T Sigurdsson, S Raucher *Tetrahedron* **47** 2475 (1991)
81. J T Millard, M F Weidner, S Raucher, P B Hopkins *J. Am. Chem. Soc.* **112** 3637 (1990)
82. M Tomasz, R Lipman, D Chowdary, J Pawlak, G L Verdine, K Nakanishi *Science* **235** 1204 (1987)
83. M F Weidner, S T Sigurdsson, P B Hopkins *Biochemistry* **29** 9225 (1990)
84. H Huang, T K Pratum, P B Hopkins *J. Am. Chem. Soc.* **116** 2703 (1994)
85. H M Coley, N Brooks, D H Phillips, A Hewer, T C Jenkins, M Jarman, I R Judson *Biochem. Pharmacol.* **49** 1203 (1995)
86. M Walicka, L Szmigiero, E Ciesielska, I Gradzka *Biochem. Pharmacol.* **46** 615 (1993)
87. M G Remias, C S Lee, I S Haworth *J. Biomol. Struct. Dyn.* **12** 911 (1995)
88. I V Kutyavin, H B Gamper, A A Gall, R B Meyer *J. Am. Chem. Soc.* **115** 9303 (1993)
89. Q Dong, N Bullock, F Ali-Osman, O M Colvin, D D Bigner, H S Friedman *Cancer Chemother. Pharmacol.* **37** 242 (1996)
90. A Masta, P J Gray, D R Phillips *Nucl. Acids Res.* **22** 3880 (1994)
91. S M Rink, M S Solomon, M J Taylor, S B Rajur, L W McLaughlin, P B Hopkins *J. Am. Chem. Soc.* **115** 2551 (1993)
92. F Ali-Osman, A Rairkar, P Young *Cancer Biochem. Biophys.* **14** 231 (1995)
93. Y Fujiwara, M Nakamura, S Yokoo *Br. J. Cancer* **67** 1285 (1993)
94. G N Rudd, J A Hartley, R L Souhami *Cancer Chemother. Pharmacol.* **35** 323 (1995)
95. M Nakamura, M Yamamoto *Exp. Eye Res.* **59** 53 (1994)
96. I E Kanevskii, S A Kuznetsova, E M Volkov, Z A Shabarova *Mol. Biol.* **30** 1144 (1996) [a]
97. T R Webb, M D Matteucci *Nucl. Acids Res.* **14** 7661 (1986)
98. T R Webb, M D Matteucci *J. Am. Chem. Soc.* **108** 2764 (1986)
99. J F Milligan, M D Matteucci, J C Martin *J. Med. Chem.* **36** 1923 (1993)
100. J-P Shaw, J F Milligan, S H Krawczyk, M D Matteucci *J. Am. Chem. Soc.* **113** 7765 (1991)
101. D Praseuth, L Perrouault, T Le Doan, M Chassignol, N Thuong, C Helene *Proc. Natl. Acad. Sci. USA* **85** 1349 (1988)
102. S H Krawczyk, J F Milligan, S Wadwani, C Moulds, B C Froehler, M D Matteucci *Proc. Natl. Acad. Sci. USA* **89** 3761 (1992)
103. S L Young, S H Krawczyk, M D Matteucci, J J Toole *Proc. Natl. Acad. Sci. USA* **88** 10023 (1991)
104. O S Fedorova, D G Knorre, L M Podust, V F Zarytova *FEBS Lett.* **228** 273 (1988)
105. S A Strobel, P B Dervan *Science* **249** 73 (1990)
106. T Le Doan, L Perrouault, D Praseuth, H Habhoub, J L Decout, N T Thuong, J Lhomme, C Helene *Nucl. Acids Res.* **15** 7749 (1987)
107. H E Moser, P B Dervan *Science* **238** 645 (1987)
108. M Poidot-Forget, M Chassignol, M Takasugi, N T Thuong, C Helene *Gene* **72** 361 (1988)
109. T J Povsic, P B Dervan *J. Am. Chem. Soc.* **112** 9428 (1990)
110. M D Matteucci, T R Webb *Tetrahedron* **28** 2469 (1987)
111. M B Gottikh, O A Fedorova, M V Baud-Demattei, S Giorgi-Renault, J R Bertrand, Z A Shabarova, C Malvy *J. Am. Chem. Soc.* **118** 2126 (1996)
112. R Cosstick, M E Douglas *J. Chem. Soc., Perk. Trans. 1* 1035 (1991)
113. A E Ferentz, G L Verdine *J. Am. Chem. Soc.* **113** 4000 (1991)
114. A M McMillan, G L Verdine *Tetrahedron* **47** 2603 (1991)
115. C M Harris, L Zhou, E A Strand, T M Harris *J. Am. Chem. Soc.* **113** 4328 (1991)
116. H Lee, M Hinz, J J Stezowski, R G Harvey *Tetrahedron Lett.* **31** 6773 (1990)
117. A E Ferentz, T A Keating, G L Verdine *J. Am. Chem. Soc.* **115** 9006 (1993)
118. A E Ferentz, J Wiorkiewicz-Kuczera, M Karplus, G L Verdine *J. Am. Chem. Soc.* **115** 7569 (1993)
119. R J Cain, E R P Zuiderweg, G D Glick *Nucl. Acids Res.* **23** 2153 (1995)
120. L G Kuznetsova, E M Volkov, E A Romanova, V N Tashlitskii, T S Oretskaya, Z A Shabarova *Bioorg. Khim.* **17** 1289 (1991) [b]
121. L G Kuznetsova, E A Romanova, E M Volkov, V N Tashlitskii, T S Oretskaya, N F Krynetskaya, Z A Shabarova *Bioorg. Khim.* **19** 455 (1993) [b]
122. H Aurup, T Tuschl, F Benseler, J Ludwig, F Eckstein *Nucl. Acids Res.* **22** 20 (1994)
123. F Benseler, D M Williams, F Eckstein *Nucleosides Nucleotides* **11** 1333 (1992)
124. M Imazawa, F Eckstein *J. Org. Chem.* **44** 2039 (1979)
125. H Aurup, D M Williams, F Eckstein *Biochemistry* **31** 9636 (1992)
126. O Heidenreich, F Benseler, A Fahrenholz, F Eckstein *J. Biol. Chem.* **269** 2131 (1994)
127. S Th Sigurdsson, F Eckstein *Trends Biotechnol.* **13** 286 (1995)
128. S I Antsypovich, E M Volkov, T S Oretskaya, E A Romanova, V N Tashlitskii, M Blyumenfel'd, Z A Shabarova *Bioorg. Khim.* **21** 774 (1995) [b]
129. S I Antsypovich, T S Oretskaya, E M Volkov, E A Romanova, V N Tashlitsky, M Blumenfeld, Z A Shabarova *Nucleosides Nucleotides* **15** 923 (1996)
130. S I Antsypovich, T S Oretskaya, E A Romanova, E M Volkov, V N Tashlitskii, M Vasser, Z A Shabarova *Bioorg. Khim.* **22** 264 (1996) [b]
131. S I Antsypovich, T S Oretskaya, E A Romanova, E M Volkov, V N Tashlitsky, M Vasser, Z A Shabarova *FEBS Lett.* **378** 224 (1996)

[a] — *Mol. Biol. (Engl. Transl.)*
[b] — *Russ. J. Bioorg. Chem. (Engl. Transl.)*
[c] — *Docl. Chem. Technol., Dokl. Chem. (Engl. Transl.)*
[d] — *Biochemistry (Engl. Transl.)*