Reviews

Chemical ligation as a method for the assembly of double-stranded nucleic acids: Modifications and local structure studies

N. G. Dolinnaya and Z. A. Shabarova*

Department of Chemistry, M. V. Lomonosov Moscow State University, Vorob'evy Gory, 119899 Moscow, Russian Federation. Fax: +7 (095) 939 3181. E-mail: zoeshab@biorg.chem.msu.su

A new procedure was developed as an alternative to the enzymatic assembly of natural and modified double-stranded DNAs using chemical reagent (chemical ligation). BrCN was suggested as an efficient coupling reagent, which induces superfast reactions in DNA duplexes. The physicochemical properties and the structure of new types of DNA duplexes, which are the substrates for chemical ligation, with breaks in phosphodiester chains, including concatemers, were studied. Chemical ligation was applied to prepare biologically active 17—200 base-pair double-stranded DNAs and DNA—RNA block-copolymers, to incorporate various modifications into DNA duplexes including pyrophosphate and phosphoramidate unnatural internucleotide bonds. The unique possibilities of this approach were demonstrated in the development of methods for circularization of oligodeoxyribonucleotides and assembly of branched DNAs. The structural-kinetic concept of chemical ligation was created and the relationship between the reactivity of interacting groups and sequence-dependent local conformation of the ligation site in B-DNA was established. The lesser efficiency of chemical ligation of RNA fragments in comparison to that of DNA analogs was demonstrated and rationalized. This approach was used as a sensitive monitor of a stable double helix formation and third-strand binding to a DNA duplex.

Key words: chemical ligation, modified DNA duplexes, concatemeric DNAs, sequence-dependent conformation of B-DNA, DNA ligase, circular DNAs, branched DNAs, RNA—DNA hybrid duplexes.

Introduction

The double helix of DNA is a unique example of a self-organized system. The complementary fragments of the nucleic acid are unequivocally assembled in aqueous solution into a double-helix structure, which is stabilized by specific interactions of heterocyclic bases. This fundamental property underlies the modern chemoenzymatic synthesis of genes. For covalent binding of oligonucleotide blocks, the enzyme DNA-ligase, which *in vivo* repairs single-strand breaks in DNA, is usually employed.

This review discusses the study of chemical reactions that mimic the action of DNA-ligase, "chemical ligation", aimed at the development of highly efficient methods of assembly of genetic structures. Synthetic DNAs have found extremely broad application in molecular biology, biotechnology, and medicine. They

Translated from *Izvestiya Akademii Nauk. Seriya Khimicheskaya,* No. 8, pp. 1889—1911, August, 1996.

1066-5285/96/4508-1787 $15.00 © 1996 Plenum Publishing Corporation

include artificial genes, which are producers of relatively unavailable proteins and enzymes, gene expression regulators, diagnostic reagents, highly specific pharmacological preparations, kits of unique substrates to study various aspects of protein-nucleic acid interactions, and conformational peculiarities of nucleic acids, *etc.*

The approach being developed allows one to obtain both natural and modified compounds, because the chemical activation of the reactive groups removes restrictions caused by the substrate specificity of the enzyme. Chemical ligation opens up wide possibilities in the design of nucleic acids with the anomalous tertiary structure, and allows one to scale up substantially the synthesis of the genetic material and to cut the cost of the process of DNA assembly by a factor of several hundreds. Evidently, the problem of automatizing the synthesis of genes may be solved exclusively on the chemical basis.

The idea of purely chemical synthesis of stretched nucleic acids has attracted the attention of scientists for a long time. As early as the middle of 1960s, the first, not very successful attempts to apply a chemical reagent for the condensation of hexathymidylates on a polyadenylate template were made.[1] At the same time, studies were begun on template-directed polymerization of ribonucleotides, which could mimic a prebiotic RNA synthesis.[2,3] However, the problem of developing chemical ligation as an alternative method for assembly of genetic material was set up for the first time at Moscow State University.[4] From the late 1960s, a systematic study of regularities of chemical reactions within nucleic acid duplexes was begun there. Later, the Novosibirsk group of bioorganic chemistry joined the studies in this area.[5] The development of chemical ligation proceeded in parallel with the advances in the techniques of oligodeoxyribonucleotides synthesis. The progress in this field and the successful automation of oligonucleotide synthesis[6] made heterogeneous DNA fragments available by the late 1970s.[7] Later, the logic of the study predetermined the use of chemical ligation not only as the synthetic method for various genetic structures, but also as a tool for the analysis of local conformation of nucleic acids, in particular, near the modified site. It should be noted that the studies of chemical transformations in the spatially arranged nucleic acids is one of a few branches of physicochemical biology originally developed in Russia. In recent years, the interest in these studies was stimulated by works on RNA selfsplicing, *i.e.*, the directed cleavage and formation of internucleotide bonds without involvement of an enzyme.

The aim of this study was the development of chemical methods of DNA assembly, which is the least studied step in the synthesis of genes. The required steps of this work included the choice of adequate model systems, which made it possible to characterize the general regularities of chemical ligation; the study of the properties and structure of complexes of complementary oligonucleotides, in which template condensation proceeds; and a search for the most efficient chemical reagents able to activate the phosphate group under conditions of the stability of the double helix. Evidently, the efficiency of chemical reactions in self-organized nucleic acids is connected with structural factors, *viz.*, the proximity and orientation of the reacting groups. Therefore, the study of possibilities of chemical ligation as a tool of studies of the local and general structures of multistranded nucleic acids including modified and anomalous ones, was a special problem. Finally, it seemed expedient to compare the possibilities of chemical and enzymatic ligation and to determine the requirements to be met by oligonucleotide complexes to serve as substrates of DNA-ligase and of reactions of chemical ligation.

1. Design and physicochemical study of model system for chemical ligation. General regularities of chemical reactions in DNA-duplexes

The method of chemical ligation was developed in two directions.

1. The design and study of physicochemical characteristics of oligonucleotide duplexes, the reaction substrates. The formulation of requirements for their thermodynamic and geometric parameters.

The specificity of the systems for chemical ligation is the presence of breaks in sugar-phosphate backbone, which destabilize substantially the double helix. Hence the need for analysis of extended heterogeneous sequences, which make possible unequivocal arrangement of oligomers in the template. The importance of using adequate model systems in the study of chemical ligation was clearly demonstrated[8] when the formation of 5′-5′-pyrophosphate bonds in the oligo d(A) · 2 poly(U) triple-stranded complex due to irregular orientation of the components to be linked was investigated.

2. The choice of the most efficient methods of activation of the phosphate residue, which would make it possible to perform the synthesis of the internucleotide bond under conditions when the double helix is stable (aqueous solutions, 0—10 °C, pH close to neutral). They may be arbitrarily classified in three groups:

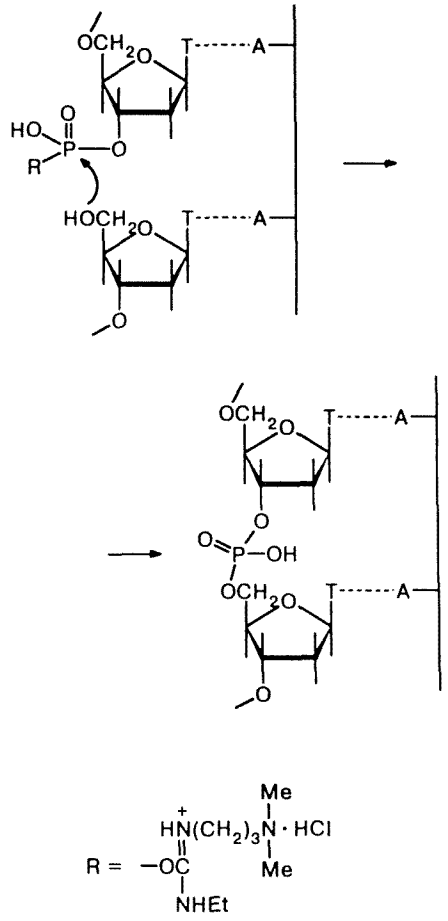
A. The methods based on the use of water-soluble reagents activating the phosphate group directly within the complementary complex.

B. The use of stable activated oligonucleotide derivatives (azolides, esters) capable of forming the double helix.

C. Activation of phosphate-group derivatives after complex formation using external factors, *viz.*, oxidation, irradiation, *etc.* (so-called "switch reagents").

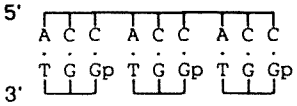
We believe that the first method is the most promising for practical development; it involves the use of convenient, water-soluble coupling reagents. The formation of the phosphodiester bond between oligonucleotides brought together on a complementary matrix by the action of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC), which has been widely used in this work, is presented in Scheme 1.

Scheme 1



1.1. DNA duplexes with single-strand breaks

The first model system we designed was represented by double-helix complexes of heterogeneous trinucleotides with complementary oligomers.[9,10] One of the duplexes studied is shown below (hereinafter, the symbol d is omitted).



Thermodynamic parameters of these oligomeric complexes were calculated using the "all-or-none" model, which is an adequate approximation for the systems studied by a series of criteria (a good correlation between the ΔH° and ΔS° values calculated from the data of CD and UV spectroscopy, the double helix growth constant of $S > 10$, *etc.*). In the framework of this model, we have derived the general expression for the equilibrium constant (K) of the complex formation of oligomers with different lengths, including the systems with "free" ends

$$K = \frac{f(2m)^n}{C_o^n(1-f)^{n+1}}$$

(duplexes with the total filling of the template), where f is the proportion of mononucleotide units involved in the complex formation; C_o is the total nucleotide concentration (per monomer); m is the number of nucleotide units in the shorter component of the complex; and n is the number of molecules of the oligomer interacting with the oligomeric template.

It was found that a trimer fills the template without gaps, forming quite a stable complex (melting temperature, T_m *ca.* 22 °C for a dodecanucleotide double helix). It was experimentally demonstrated for these systems that the presence of breaks in one of the chains does not affect substantially the thermodynamic parameters of the helix.[10] The similarity of the CD spectra of d(pGGT) · d(pACC)$_4$ and d(pA)$_3$ · d(pT)$_{12}$ complexes to the spectra of DNAs having the content of the G · C pairs 64 and 33 %, respectively, was observed (Fig. 1), which indicates that the internal geometric characteristics of the helix, which depend on the nucleotide composition, are retained despite the presence of single-strand breaks.[9]

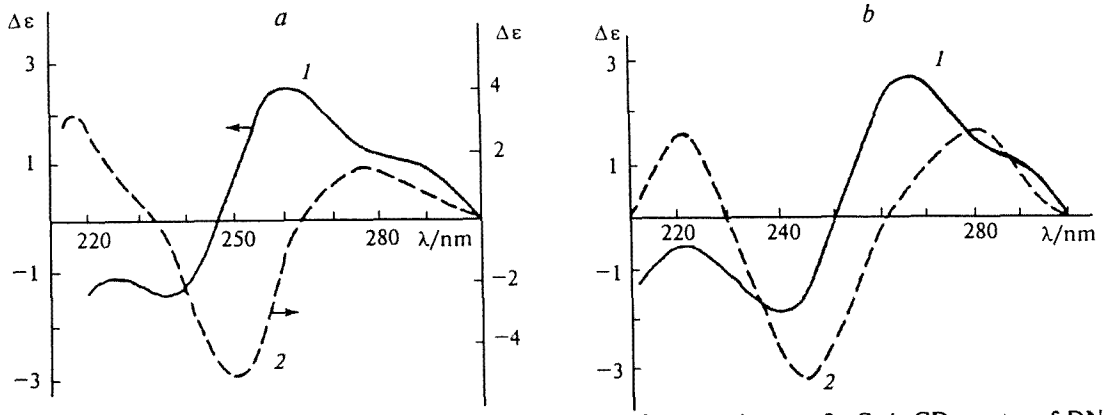


Fig. 1. *a*. CD spectra of (*1*) d(pGGT) · d(pACC)$_4$, and (*2*) d(pA)$_3$ · d(pT)$_{12}$ complexes at 2 °C. *b*. CD spectra of DNA from (*1*) *Pseudomonas aeruginosa* (64 % G · C pairs) and (*2*) from *Staphylococcus aureus* (33 % G · C pairs). Phosphate buffer, pH 7, 0.2 *M* NaCl, 0.075 *M* $MgCl_2$.

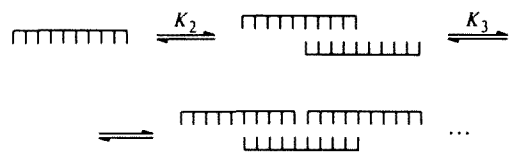
In the d(TGGp) · d(pACC)$_4$ duplex, the first successful experiment on chemical ligation in heterogeneous systems was performed.[11] The yields of hexa- and nonanucleotides by the action of *N*-cyclohexyl-*N'*-2-(4-methylmorpholinium)ethylcarbodiimide *p*-toluenesulfonate (CMEC) were 25 and 10 %, respectively.

1.2. Concatemeric duplexes with noncoinciding breaks in the complementary chains

An important step in the studies on the general regularities of chemical ligation was connected with the discovery and the description of properties of a new class of synthetic, DNA-like duplexes containing repeating oligonucleotide sequences (concatemers).[12] These complexes are formed from one or more oligonucleotides with specific primary structure, whose characteristic feature is the neighborhood of two palindromic sequences. Thus, the initiating duplex has "sticky" ends stimulating the subsequent growth of the double helix (Fig. 2).

It should be noted that the physicochemical properties of nucleotide complexes with "broken" sugar-phosphate backbone are practically unexamined, although structures of this kind are widely used in the assembly of artificial genes. This work used UV spectroscopy and CD to show that the concatemer formed upon interaction of one or two oligonucleotides is an intermolecular, double-helix possessing a geometry similar to the geometry of B-DNA.[13,14] The size of the self-associates is estimated by sedimentation. Thus, the double-helix concatemer formed from d(TGGCCAAGCTp) contains on the average 6—8 molecules of the decanucleotide (the sedimentation coefficient is 2.7±0.2 S which is characteristic of a DNA fragment with 32—35 base pairs (b.p.)).

To describe thermodynamically the process of concatemerization, we used a model in which the interaction of oligomers leads to the formation of a series of structures from a dimer to the longer polymers.[12]



The concentration of an oligomer as a constituent of a concatemer comprising i molecules is designated as C_i. Then, the concentration of free oligomers, C_c, and of those involved in the complex formation, C_h, may be expressed as:

$$C_c = C_1 + C_2 + \ldots + C_n,$$

$$C_h = C_2 + 2C_3 + 3C_4 + \ldots + (n-1)C_n.$$

Assuming that the equilibrium constants of each step of the complex formation, K_i (i = 2,... n), are independent of the length of concatemers and are equal to each other ($K_i = K$), the concentration C_i may be expressed as $C_i = K^{i-1}C_1^{\,i}$ and the expressions for C_c and C_h are as follows:

$$C_c = C_1\frac{1-(KC_1)^n}{1-KC_1};$$

$$C_h = KC_1^2\left[\frac{1-(KC_1)^{n-1}}{(1-KC_1)^2} - \frac{(n-1)(KC_1)^{n-1}}{1-KC_1}\right].$$

If $KC_1 < 1$ and n is high, from the equality of $C_h = C_c$ in the melting point, the expression for the equilibrium constant is obtained as:

$$K = \frac{2}{C_o},$$

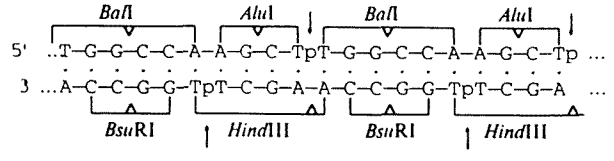
where C_o is the total concentration of oligonucleotides ($C_o = C_c + C_h$). The resulting expression coincides with that found for the isodesmic model.[15]

The equilibrium constant of n-mer formation , K^*, is the product of individual constants K_i:

$$K^* = K_2 \cdot K_3 \cdot \ldots K_n = K^{n-1} = \left(\frac{2}{C_o}\right)^{n-1}.$$

The values of enthalpy of formation of an averaged base pair upon concatemerization of d(TGCACATGp) and d(TGGCCAAGCTp) are in good agreement with the thermodynamic parameters of the helix—coil transition in DNAs with the same content of G · C pairs obtained by the direct calorimetric measurements.[12,14]

The peculiar features of the secondary structure of the concatemeric complexes are very suitable for the study of template-directed reactions. Using these systems, it was shown that under conditions where a duplex is stable, water-soluble CMEC and more reactive EDC induce effective polycondensation of the initial oligonucleotides, affording a mixture of polymers of different length (Fig. 3).[13,14,16,17] These double-helix polymers may contain sense tandem repeats, for example, recognition sites for the restriction endonucleases:



(the ligation sites are marked by arrows; the site of cleavage with the corresponding restriction endonucleases

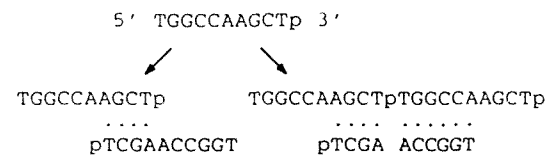


Fig. 2. Scheme of formation of concatemeric DNA-duplexes.

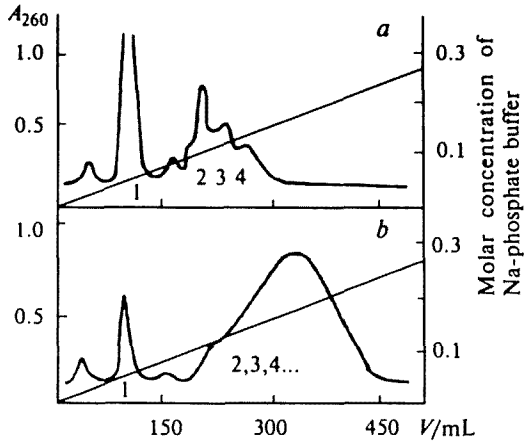


Fig. 3. Chromatographic separation of condensation products of d(TGGCCAAGCTp) obtained under the action of EDC after (*a*) 1 and (*b*) 4 days; 1, 2, 3, 4, *etc.*: degree of oligomerization.

is marked by a triangle). The polymeric duplexes obtained by chemical ligation are cleaved completely with the corresponding enzymes into decanucleotides.[14] This fact confirms the structure of the compounds formed and the absence of the products of modification of heterocyclic bases with carbodiimide. The use of concatemeric duplexes with unnatural pyrophosphate or phosphoramidate bonds opens up quite new possibilities for the study of functioning of various DNA-binding proteins and enzymes. These bonds are introduced easily into the "linked" concatemer by replacing the 5′-hydroxyl group by more reactive amino- or phosphate groups.

It was shown experimentally for the first time, using the concatemeric systems, that the efficiency of chemical ligation (for concatemers, it is directly proportional to the length of the polymers formed) depends on the stability of the initial oligonucleotide complex. This is illustrated in Table 1, which presents the concatemeric complexes of different types studied (the initial one or two oligonucleotides are included in boxes) given schematically and melting temperature and degrees of polymerization for the products of chemical ligation. In the case of complexes of equal stability, the electrophilicity of the activated phosphate and the nucleophilicity of the attacking groups affect the efficiency of the reaction. A series of nucleophilic groups arranged in order of the increase in the ligation efficiency, which has been obtained in the studies on concatemeric systems, is given below.

$$3'\text{—OH} < 5'\text{—OH} << \text{—}\overset{\overset{\displaystyle O}{\|}}{\underset{\underset{\displaystyle O^-}{|}}{P}}\text{—}O^- \leq \text{—}NH_2$$

The reaction rate increases 20—30-fold for strong nucleophiles, and the average chain length of pyrophosphate-containing polymers is 200—220 monomer units, whereas for concatemers with the phosphodiester bonds, it is equal to 60—70 units.[18] It was also shown for concatemeric systems that EDC used at concentrations suitable for chemical ligation decreases the melting temperature of concatemeric duplex only by several degrees, and the destabilizing effect decreases with an increase in the content of G · C pairs.[14] The 20—200 b.p. DNA-duplexes synthesized containing tandem repeating fragments of procaryotic and eucaryotic promoters and the recognition sites of restriction endonucleases, including

Table 1. Data on EDC-induced polycondensation of oligonucleotides

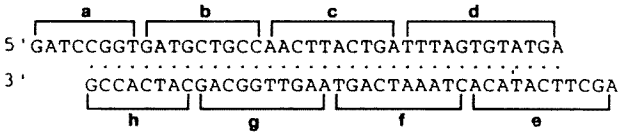
Structure of concatemeric DNA-duplex	T_m* /°C	Degree of conversion of oligonucleotide** (%)	Degree of polymerization, *n*
TGCATTATAAp	3—5	70	2—7
TGGCCAAGCTp	31	90	2—25
CCTGGAATTp	18	85	2—20
TTTCGACTTCGAp	38	97	10—30

Note. Similar oligonucleotides are desigated by similar rectangles.
* Accuracy of measurement was ±2 °C. ** Determination with accuracy of ±3 %.

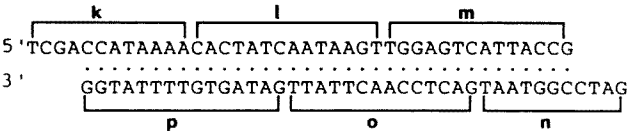
modified ones,[16,17] were used successfully for the study of different aspects of protein-nucleic acids interactions.[19]

1.3. Extended DNA-duplexes built of oligomers with different nucleotide sequences

Despite successful polycondensation of oligonucleotides in concatemeric complexes, the following problems remained unsolved. Is it possible that all the acts of the template "linking" in the chemical assembly of unique DNA fragments from oligonucleotides with different primary structures were equially efficient? What factors affect the formation of interoligomeric bonds involving the weak nucleophile, hydroxyl group? To solve these problems, the oligonucleotide fragments, which form 35 b.p. DNA-duplexes **I** and **II**, were synthesized; the "upper" strands of these duplexes correspond to the fragments of the terminal, inverted repeat of the IS1-element.[20]



Duplex **I**



Duplex **II**

Variation in the length of the overlapping blocks allows one to control the thermal stability of the double helices formed. It was found that, in contrast to the concatemeric systems, the length of 4—6 overlapping b.p. (duplex **I**) in a mixture of oligomers differing in composition, is insufficient to form complexes stable at the temperatures >0 °C (the "weak" sites result from the heterogeneity of the ends). To form stable double-helix structures (T_m *ca.* 30 °C), the overlap of 7—8 b.p. (duplex **II**) is necessary. These DNA-duplexes exhibit a monophasic helix—coil transition independent of the number of constituent components; the elongation of the double-helix core due to the further addition of oligonucleotide molecules does not affect significantly the thermal stability and hypochromicity of complex-formation. As can be expected, the EDC-induced ligation of the 5′-phosphorylated components of duplex **I** proceeds with low efficiency (the degree of transformation is <5 %). On the contrary, in mixtures composed of the components of duplex **II**, which is stable under reaction conditions, the yields of products of chemical ligation are significantly higher, they vary from 14 to 60 % depending on the nature of nucleotide units, which are involved in the new phosphodiester bond formation. The individual acts of chemical ligation were studied for three-component systems, *e.g.*, oligonucleotides **k** and **l** were linked in the presence of pentadecanucleotide **p**. To assemble each complete single-stranded fragment of duplex **II**, mixtures of nonphosphorylated terminal and "template" oligonucleotides with phosphorylated oligonucleotides of the chain to be assembled were composed (Fig. 4). The yields of the final products, trimers, reached 32 and 49 % for "lower" and "upper" chains of the double helix, respectively.[20]

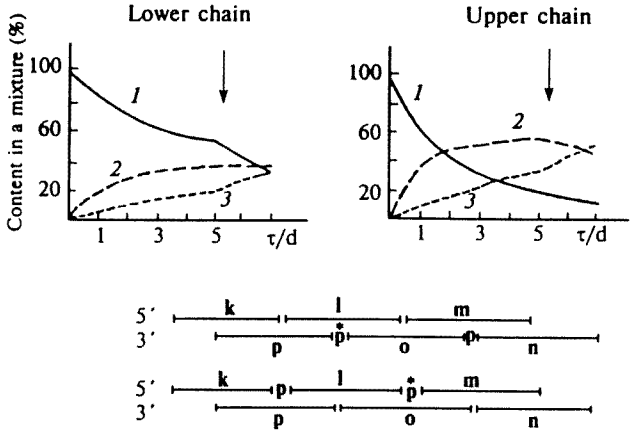


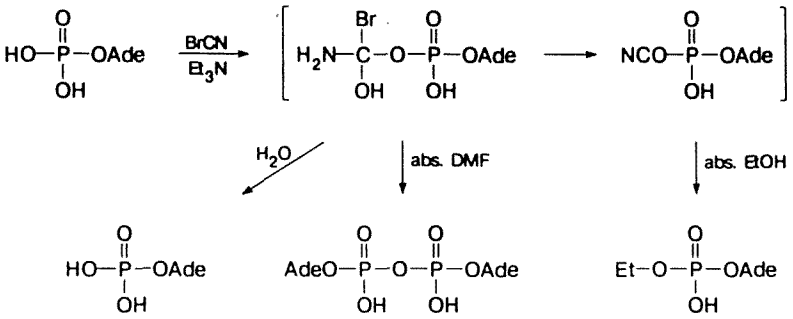
Fig. 4. EDC-induced assembly of the "lower" and "upper" chains of duplex **II**: (*1*) the starting ^{32}P-labelled oligomer, (*2*) the intermediate dimer, (*3*) the total fragment. Addition of a new portion of the reagent is marked by an arrow. A scheme of assembly of fragments is presented below.

The results necessitated a systematic study of the dependence of chemical ligation efficiency on the nature of nucleotide residues to be connected. This aimed at making each stage of chemical assembly highly efficient by performing rational disconnection of the target duplex into blocks. Evidently, the study of the individual acts of chemical ligation in order to reveal the stereochemical and kinetic regularities of the reaction would favor the creation of the optimal synthetic route for natural and modified DNAs of a specified composition. A search for more efficient coupling reagents capable of forming a new internucleotide bond at a rate close to that of enzymatic processes is also urgent.

2. BrCN as a new, efficient reagent for chemical ligation

The prerequisite for a reagent to be suitable for template-directed condensation of oligonucleotides is the formation of 2′,3′-cyclophosphate from ribonucleoside-3′-phosphate under the action of this reagent in an aqueous medium. Since BrCN satisfied this criterion, it has been suggested for use as the reagent for chemical ligation.[21] Optimization of the reaction conditions was carried out for a simple three-oligomer model system, which made possible the study of a single act of "linking" of two oligonucleotides on the complementary template in detail. Varying the composition, capacity, and pH of

Scheme 2



the buffer solution and a temperature regime, we succeeded in achieving a yield of product with the phosphodiester bond up to 100 %.[21,22] The main advantages of BrCN are the substantially higher reaction rate (1—3 min instead of several hours and even days for EDC) and the absence of side products resulting from modifications of nucleotides. Simultaneously and independently, Japanese scientists[23] have recommended using BrCN for condensation of oligo(pA) on a polyuridylate template. The reaction was carried out for 20—25 h at 25 °C; it occurred only in the presence of imidazole and ions of bivalent metals. The authors suggest that the actual coupling reagents in this reaction are *N,N'*-iminodiimidazole and *N*-cyanoimidazole. In our verson for chemical ligation using BrCN, the reaction occurs substantially faster and in the absence of imidazole. This indicates the absolutely different mechanism of activation of the nucleotide component. The control experiments for duplexes containing two phosphate or two hydroxyl groups in the ligation site demonstrated that the condensation proceeds only in the duplex with phosphomonoester groups, *i.e.*, it is these groups that are activated by BrCN. The ^{31}P NMR spectroscopy was used to study the mechanism of reaction of BrCN with deoxynucleotides under conditions of chemical ligation. Based on these experiments, the following scheme of transformations of initial dpA (Scheme 2) was suggested. In anhydrous DMF, the reaction of the activated, unstable intermediate with the strongest nucleophile in this system, *viz.*, the phosphate group of another molecule of dpA affords mononucleotide pyrophosphate and tripolyphosphate. This reaction is inhibited in anhydrous ethanol; ethyl ester of dpA is the major product in this case. In an aqueous medium, hydrolysis of the BrCN-activated phosphate derivative occurs extremely rapidly with regeneration of the initial nucleotide. Only in the double helix, the activated phosphate can react with the hydroxyl group of a proximate oligonucleotide to form a new internucleotide bond.

3. Comparison of coupling abilities and substrate specificities of a water-soluble carbodiimide and BrCN

A series of DNA-duplexes with one single-strand break, differing in the nature and orientation of the reacting groups, was designed in order to correctly compare the efficiency of the formation of natural and modified internucleotide bonds under the action of both coupling reagents (Table 2).[24,25] Duplexes **IIIa—p** are built according to the same general principle: the template and one of complementary oligomers remain unchanged, and the second oligomer differs only in the nucleotide residue adjacent to the break. The modifications in the ligation site included the replacement of a hydroxyl group by amino- or phosphate groups, the terminal dT residue by dA or dC (duplexes **IIIe,f** with mismatched pairs AA or AC), by rU or nucleotide residues with the inverted configuration at the C(2') or C(3') atoms, and the deletion of the nucleotide residue. The choice of the nucleotide sequence and the site of single-strand break is determined by the fact that the recognition site of restriction endonuclease *Eco*RII ($CC\overset{A}{\underset{T}{}}GG$) is included in the primary structure of the duplex, and the modifications are adjacent to the cleavable bond.

A necessary preliminary step was the study of thermal stability of modified duplexes containing single-strand breaks, so-called nick-containing duplexes. Only duplexes in which all nucleotide pairs are canonical melt as a single cooperative system. The presence of two phosphate groups, an amino group, and the carbohydrate residues with the inverted configurations at the C(2') or C(3') atoms at the nick does not affect the melting temperature of the helix of the considered type. The biphasic melting is characteristic of duplexes with other types of structural anomalies (Fig. 5, *a*); the first step corresponds to dissociation of the less tightly bound modified hexanucleotide, and the second step corresponds to melting of the complex formed by the undecanucleotide and the template. The absence of the nucleotide residue in the break point and, to a lesser extent, noncomplementary AC and AA pairs give the strongest destabilizing effect. The hybrid rU · dA pair disturbs the structure of the helix the least.[24,26] The order obtained coincides completely with that expected on the basis of the published data on the stability of intact oligomeric duplexes with structural anomalies.[27] The thermodynamic parameters of conformational transitions were calculated for all of the complexes studied, including two-phase melting systems.[26]

These studies demonstrated that at a temperature below 10 °C, all duplexes are thermodynamically stable and may serve as the models for the investigation of

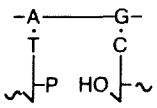
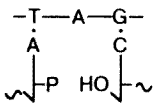
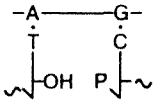
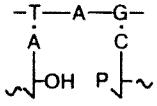
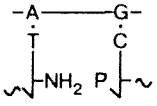
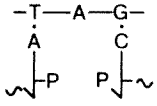
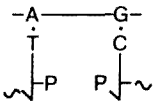
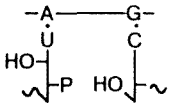
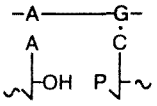
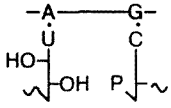
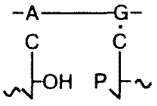
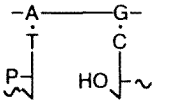
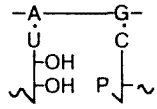
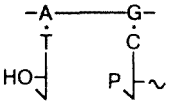
Table 2. Condensation of oligonucleotides in duplexes **IIIa—p**

3′ G–C–C–T–A–G–G–T–C–C–T–C–A–C

. . . . . .

5′ A–C–G–G–A–**X Y**–C–A–G–G–A–G–T–G–A–C

under the action of chemical reagents and T4—DNA-ligase

Duplex	Structure of the ligation site –X Y–	Yield of the condensation product* (%) EDC	BrCN	DNA-ligase	Duplex	Structure of the ligation site –X Y–	Yield of the condensation product* (%) EDC	BrCN	DNA-ligase
IIIa		92	95	—	**IIIh**		25	10	—
IIIb		80	35	95	**IIIk**		8	5	0
IIIc		97	86	0	**IIIl**		85	16	0
IIId		95	67	0	**IIIm**		75	73	—
IIIe		32	5	0	**IIIn**		23	10	0
IIIf		50	7	0	**IIIo**		18	26	—
IIIg		28	16	80	**IIIp**		12	4	5

* Determination with accuracy of ±2 %.

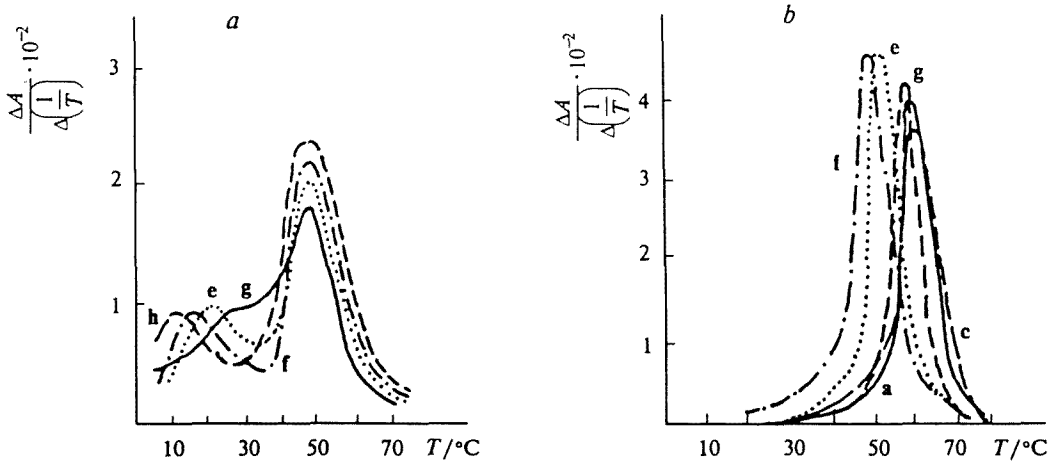


Fig. 5. Differential melting curves of duplexes **IIIa,c,e—h** containing (*a*) single-strand break and (*b*) ligated duplexes (the structures of duplexes see Table 2).

stereochemical aspects of chemical ligation. The yields of the ligation products under the action of BrCN, EDC, and T4—DNA-ligase are given in Table 2. The primary structures of the "ligated" oligonucleotides were confirmed by the Maxam—Gilbert sequencing; the nature of the pyrophosphate and phosphoramidate bonds formed was established by their selective cleavage with trifluoroacetic anhydride or 15 % acetic acid, respectively.[28—30] The nature of the bond formed by the 3′-terminal rU (**IIIg**) and the nucleotide residues with the inverted hydroxyl groups at the C(2′) or C(3′) atoms of furanose (**IIIm—p**) was determined using the alkaline and enzymatic hydrolyses of the corresponding oligomers.[31,32]

The formation of the internucleotide bond under the action of both reagents occurs more effectively, if the reacting phosphate group is located at the 3′-end of an oligomer (*cf.* **IIIa** and **IIIb**, and also **IIIh** and **IIIk** in Table 2).[26,33] This can be explained by the higher reactivity of the 5′-hydroxyl group and the fixed position of the 3′-phosphate group. On the other hand, the localization of the phosphate group at the 5′-end of the oligomer implies its considerable conformational freedom and, hence, the ease of hydrolysis of the activated adduct. Replacement of the hydroxyl group by more nucleophilic amino and phosphate groups (**IIIc,d**) intensifies strikingly the EDC-induced ligation, so that the reaction runs to completion after 6 h, not 6 days, as for duplexes **IIIa,b**; while the yields of products of chemical ligation are 95—97 %.[26,30] In the case of BrCN used as the coupling reagent, ligation of the same duplexes is not so efficient (the reaction was carried out for 3—5 min). This may be explained by the side reaction of BrCN with the terminal amino group of hexanucleotide (**IIIc**) or by the blocking of a nucleophilic attack by accumulation of active derivatives at both phosphate groups (**IIId**)[29]. The change in the orientation and the proximity of the reacting groups caused by the presence of noncomplementary pairs (**IIIe,f**), deletions (**IIIh—l**), and alterations in the structure (**IIIg**) and the configuration of the carbohydrate fragments (**IIIm—p**) significantly decreases the yield of the ligation product (see Table 2).[29,33,34] A definite correlation between the action of EDC and BrCN can be seen from the data shown, and the following conclusion can be drawn. Independent of the nature of the reagents, the yields of products of chemical ligation is mainly determined by the reactivity of the groups (when the nature of a nucleophile varies) or their mutual arrangement in the ligation cite (when the groups of the same chemical nature interact). Apparently, the accessibility of the activated derivative for the water molecules is essential. As can be seen from Table 2, the decrease in the yield caused by dislocation in the structure of the ligation site is more substantial when BrCN is used as the coupling reagent. The data demonstrate that both reagents possess their own advantages and disadvantages. Thus, for the synthesis of an unnatural internucleotide bond, the use of EDC is more appropriate. In the assembly of extended DNA of the natural structure, BrCN should be chosen, and 3′-phosphorylated oligonucleotides should be used as the substrates.

With a unique set of monomodified double-helix DNA at our disposal, we studied their physicochemical properties.[26] The melting profiles of some "ligated" duplexes are given in Fig. 5, *b*. As expected, they are substantially more stable than their analogs, which contain a single-strand break, and their melting curves are monophasic. Specific modifications decrease the melting temperature of the "ligated" duplexes, and the degree of destabilization is in accord with those of the corresponding precursors. The calculated values of ΔH° of intact duplexes practically coincide with the sum of the ΔH° values of individual transitions of nick-containing duplexes. The CD spectra of monomodified double helices are close to each other and possess the features characteristic of DNAs in the B-form, and the presence of a break practically does not affect the shape of the spectrum. In other words, the conformational changes in a double helix caused by point modifications of the type studied are of a local character. The data obtained for the structures and the energetics of imperfect DNAs are of special importance. The growing demand for modified DNA fragments that can be used as tools for the studies of mutagenesis, reparation, and biosynthesis of nucleic acids, as well as analogs of substrates or inhibitors in the studies of proteins of nucleic acid metabolism, is stimulating the development of efficient methods for their synthesis. In this aspect, chemical ligation can be regarded as a method for the modification of a polynucleotide chain at the specific site for the assembly of DNA-duplexes.

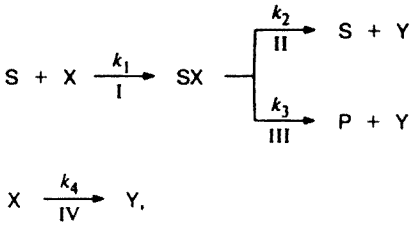
4. Structural-kinetic aspect of chemical ligation

4.1. Kinetic description of the reaction

The results obtained for the modified duplexes with a single break show the definite freezing of the reacting groups within the double helix, while chemical ligation can discriminate deviations of their conformations from the canonical ones. In subsequent studies, an attempt was made to reveal the interrelations between the kinetic parameters of chemical ligation and the structure of the reaction center. The preliminary theoretical description of the kinetics of condensation of two oligonucleotides on a complementary template was made, and the possibility of determination of rate constants of separate reaction stages and the correctness of the approximations accepted were analyzed.[32,35] The kinetic scheme[36] proposed for the description of the CMEC-induced ligation in triple-stranded, oligomer-polymeric complexes was taken as the base. The DNA-duplexes with a single break that we considered present a much more appropriate model, which permits quantitative interpretation of the experimental data (Scheme 3).

A set of equations related to Scheme 3 cannot be solved completely without several approximations. It is interesting to consider two extreme cases: (*a*) stages I

Scheme 3



where S is a double-stranded complex with a break in one of the chains, P is the condensation product (the "ligated" complex), X is an activating agent, Y is a product of hydration of the activating agent, and SX is an activated adduct.

and IV, *i.e.*, the reactions with nucleophiles, are rate-determining; (*b*) stages II and III (the decomposition of the activated adduct) are rate-determining. It is known that the rate-determining stage of the reaction of carbodiimides with phosphates is the interaction of protonated form of the carbodiimide and the dianion of the nucleotide, which corresponds to approximation (*a*). In this case, the concentration of SX may be considered quasistationary and the solution of the kinetic problem leads to the following equation:

$$\ln\frac{[S]_0}{[S]} = \frac{k_1[X]_0}{k_4}\cdot\frac{k_3}{k_2+k_3}\cdot(1-e^{-k_4t})$$

The rate of accumulation of the ligation product depends on the ratio of constants k_1 and k_4 (indicating the proportion of the reagent consumed to activate the phosphate group), and on the ratio of k_2 and k_3 (indicating the proportion of the activated adduct SX decomposed along the productive way). Let us consider a possibility to determine the rate constants for the separate stages of the reaction. The value of k_4 can be found in an independent experiment. The value of k_1 can be determined using model systems, *e.g.*, a nucleotide or a phosphorylated oligonucleotide, whose phosphate group does not undergo further transformations. In the framework of the accepted approximations, the determination of the absolute rate constants for the decomposition of the activated adduct, k_2 and k_3, is impossible. However, from the slope of the straight line in the $\ln([S]_0 - [P]/[S]_0)$ *vs.* $-(1/k_4)\cdot[X]_0(e^{-k_4t} - 1)$ coordinates, one can find the effective rate constant of chemical ligation $k_3k_1/(k_2+k_3)$. From the known k_1, it is easy to calculate the $k_3/(k_2+k_3)$ ratio characterizing the proportion of the activated adduct, which decomposes, forming the product of chemical ligation. The general pattern of the changes in the concentration of the reacting compounds in the reaction mixture, when the stage of formation of the activated adduct is rate-determining, is shown in Fig. 6, *a* using arbitrary units.

The second approximation, when the stage of decomposition of the activated adduct is rate-determining, *i.e.*, $(k_2 + k_3) \ll k_1[S] \ll k_4$ has also been considered. This can be realized, if the activating reagent is very reactive and rapidly reacts with nucleophiles. One can assume that it is the case for BrCN, whose half life in an aqueous medium is *ca.* 1 min. The change in the concentration of compounds in the reaction mixture under these conditions is presented in Fig. 6, *b* (for more convenience, the Scheme is divided into several stages). At the initial stages, the change in the concentration of the activating reagent depends mainly on the rate of noncatalytic hydrolysis; at the final step, the concentration of the reagent is practically equal to zero. The changes in the concentration of the activated derivative [SX] and the nick-containing duplex [S] are more complex. At the initial period, $d[S]/dt = k_1[S][X]$. At the second step, [S] becomes close to zero, and [SX] is practically equal to $[S]_0$. The third step, *i.e.*, the decomposition of the activated adduct along two pathways, is the most interesting. After transformation, the following equation is obtained: $\ln([SX]/[S]_0) = (k_2 + k_3)t$. This equation allows one to calculate $(k_2 + k_3)$ from the tangent of the slope of the straight line in the coordinates ln[SX] — *t*. It is interesting to note that this equation is also suitable when the activation of the phosphate group is carried out outside the complex.

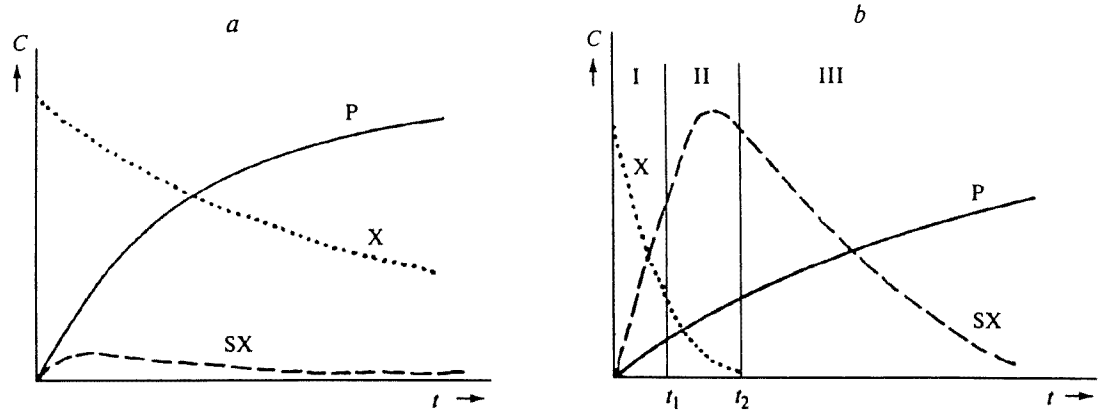
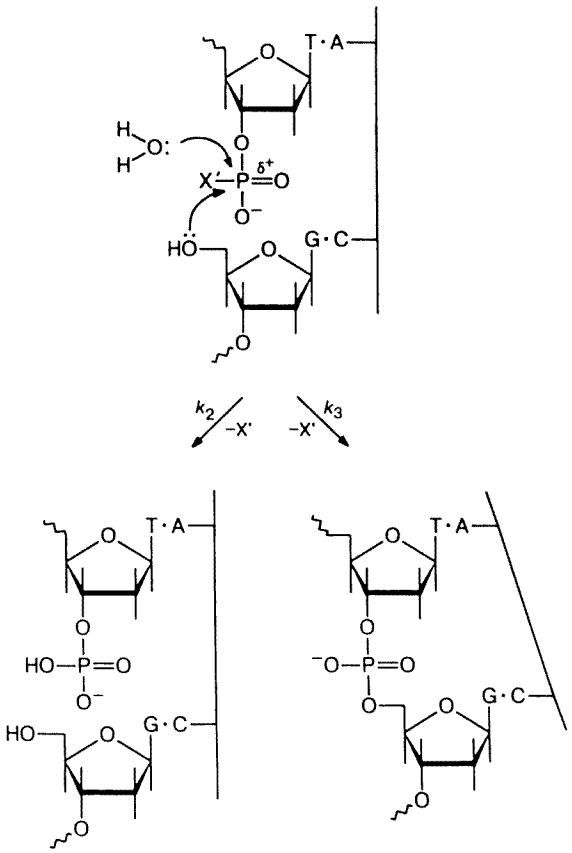


Fig. 6. Changes in the concentration of reactants during chemical ligation. The rate-determining stage: (*a*) formation of activated adduct SX; (*b*) its decomposition. Designations are presented in Scheme 3.

As can be seen from these data, the yield of the condensation product only slightly depends on the initial concentration of the activating reagent and depends mainly on the properties of the duplex, *i.e.*, on the ratio of the constants $k_3/(k_2 + k_3)$. Let us analyze the factors affecting this ratio. In the decomposition of SX (Scheme 4), the activated phosphate group can take part in two competitive reactions: with the hydroxyl group of the neighboring oligonucleotide and with water molecules. Evidently, the k_2 and k_3 values depend on the value of the partial positive charge on the phosphorus atom and the nucleophilicity of the reacting groups. When different coupling reagents are used, the absolute values of constants would change, but if the reactivity of the attacking groups is retained and no substantial distortions in the structure of the "linking" site occur, one can expect that the ratio of constants $k_3/(k_2 + k_3)$ for a given complex would be the same.

Scheme 4



4.2. Relationship between the kinetic parameters of chemical ligation in modified DNA-duplexes and the structure of the reaction site

A series of eight water-soluble carbodiimides (CDI) R^1NCNR^2, which differ in the nature of substituents R^1 and R^2, was used to demonstrate that the $k_3/(k_2 + k_3)$ ratio is independent of the structure of a coupling reagent.[37] It is known that EDC can undergo tautomeric transformation into the cyclic form in aqueous solutions. Some structural changes were introduced, *viz.*, a decrease in the ring size as a result of a decrease in the number of methylene units in the substituent R^2, blocking of the possibility of cyclization as a result of the replacement of the tertiary amine by the quaternary ammonium salt residue, *etc.*, which made it possible to reveal the role of both tautomeric forms in chemical ligation. Duplexes **IVb** (Fig. 7) (the formation of the phosphodiester bond) and analogous duplex **IVb'** (not shown), in which two phosphate groups are in contact (the formation of the pyrophosphate bond) were used as the substrates. The following conclusions were made from the data obtained.[32,35]

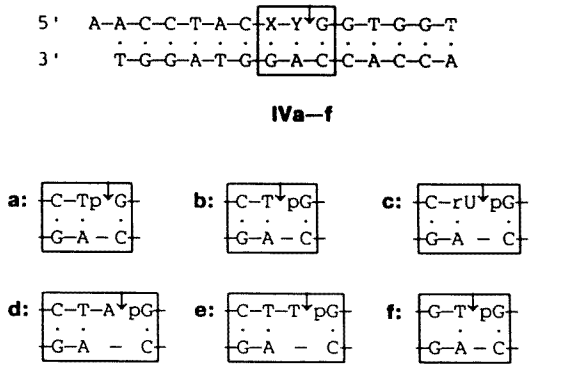


Fig. 7. Duplex **IV** (the structure of the reaction site is given in boxes).

1. A linear form of CDI is involved in reactions of chemical ligation.

2. A pronounced dependence of the kinetic parameters of the reactions, *i.e.*, rate constants for hydration and the effective reaction constants for chemical ligation on the structure of CDI, is observed.

3. The $k_3/(k_2 + k_3)$ value for the same duplex, but for different CDI, are similar and are equal to *ca.* 0.004 for **IVb** and *ca.* 0.13 for **IVb'**, *i.e.*, the ratio of constants characterizing the decomposition of the activated adduct in the productive and nonproductive pathways for a given duplex depends only slightly on the nature of the activating reagent.

4. Only a negligible part of CDI is consumed for the activation of the phosphomonoester group (*ca.* 7 % for EDC). In the course of chemical ligation with the formation of the pyrophosphate bond, *ca.* 15 % of SX decomposes in a productive pathway; when phosphodiester bond is formed, this percentage decreases to 0.4 %. Thus, the nucleophilicity of the hydroxyl and phosphomonoester group attacking the activated phosphate differ *ca.* 30-fold.

The kinetic aspects of chemical ligation under the action of EDC in duplexes **IIIa,b,g,m—p** (see Table 2), and **IVa—f** (see Fig. 7) with the variable reaction site

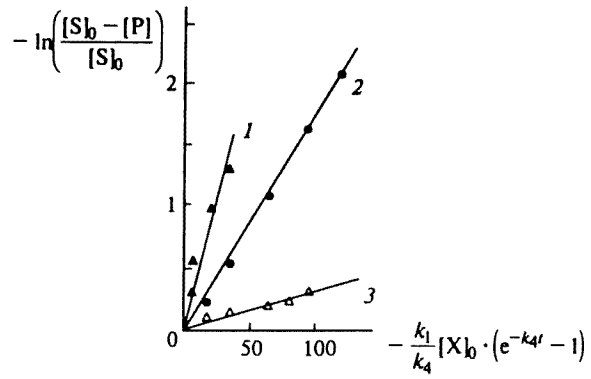


Fig. 8. Linear anamorphisms of curves of accumulation of the chemical ligation products in duplexes (*1*) **IIIa**, (*2*) **IIIb**, and (*3*) **IIIg**.

were then studied. Here, modifications of those types were used, which altered the relative position of groups with retention of their chemical nature.

In two series of duplexes **III** and **IV**, which differ in the shape of a helix (B- and A-like) and the nature of linkage, the position of the phosphate group varied; the terminal dT was replaced by rU (**IIIg**) or by the nucleotide with reversed configurations of the C(2′) or C(3′) atoms of pentofuranose, this unit could be both the acceptor and the donor of the phosphate group (**IIIm—p**). In **IVf**, the GC pair one base step away from the break was replaced by the noncomplementary GG pair; duplexes **IVd,e** contain an "extra" purine or pyrimidine monomer.[31,32] It was found that the EDC-induced reaction occcurs in all the systems studied, but with different effectivness. Linear anamorphisms of some kinetic curves are presented in Fig. 8.

From the slope of the straight line in similar plots, the ratios of constants $k_3/(k_2 + k_3)$ were calculated (Table 3). Evidently, this parameter changes strikingly for different duplexes. The analysis of the results for a series of systems and the published data taken together made it possible to obtain nontrivial information on the effect of different modifications on the local structure of the double helix. Thus, the $k_3/(k_2 + k_3)$ values are practically equal for **IVb** and **IVf**, *i.e.*, the distortion of the structure caused by the GG pair does not affect the orientation of the reacting groups of the adjacent unit and, hence, it is of a local character. It is interesting that the yields of the ligation products in **IVb** and **IVf** under the action of BrCN are also the same. An extra nucleotide residue decreases the efficiency and the rate of chemical ligation due to a change in the distance between the reacting groups and the distortion of the helix structure.

It is important and nontrivial that the parameters of ligation are different for unmodified duplexes of two series (see Table 3). This can be explained by the different nature of the residues to be bound, dT and dC in **IIIb** and dT and dG in **IVb**, and also by different secondary structures of the double helixes. CD spectra and the published data suggest that the elements of the A form are present in the structure of duplexes **IV**.[31] Is this phenomenon connected with the decrease in the efficiency of ligation in **IVb**? Lower (by 50 %) accessibility of the oxygen atoms of the phosphate group for the solvent (and the coupling reagent) in the A form in comparison to the B form affects substantially[38] the k_1 value. The shape of the helix should also influence the ratio of the constants $k_3/(k_2 + k_3)$. Thus, in duplexes **IIIg** and **IVc**, in which dT at the 3′-end is replaced by rU, the efficiency of ligation is substantially decreased in comparison to duplexes **IIIb** and **IVb** containing only deoxyribonucleotide residues. One can assume, that this is related to the change in conformation of the 3′-terminal unit. It is known that the C(2′)-*endo*(*S*)-conformation of furanose rings with the axial orientation of the 3′-hydroxyl group (Fig. 9) is characteristic of B-DNA. At the same time, in the hybrid RNA—DNA-duplex, ribonucleotide residues possess a conformation close to C(3′)-*endo*(*N*) which is characteristic of the A form of the helix, due to their structural conservative nature.[39] The mutual arrangement of the phosphate and hydroxyl groups should change substantially to become unfavorable for the formation of a new phosphodiester

Table 3. Kinetic parameters of the reaction of chemical ligation in duplexes **III** (see Table 2) and **IV** (see Fig. 7)

Duplex	Reaction site	$\frac{k_3}{k_2+k_3}\cdot 10^3$	Duplex	Reaction site	$\frac{k_3}{k_2+k_3}\cdot 10^3$
IIIa	-dTp↓dC-	25.8±8.0	**IVb**	-dT↓pdG	4.75±0.28
IIIb	-dT↓pdC-	18.5±0.9	**IVc**	-rU↓pdG-	1.75±0.54
IIIg	-rU↓pdC-	2.93±0.45*	**IVd**	-dTA↓pdG-	2.35±0.80
IIIm	-aUp↓dC-**	8.2±2.0	**IVe**	-dTT↓pdG-	2.15±0.48
IIIn	-aU↓pdC-	1.39±0.48	**IVf**	-dGT↓pdG-	4.12±0.41
IIIo	-dxTp↓dC***	1.02±0.38			
IIIp	-dxT↓pdC-	0.72±0.22			

* k_3 is an effective rate constant of formation 3′—5′- and 2′—5′-phosphodiester bonds.
** aU is 1-β-D-arabinofuranosyluracyl.
*** dxT is 1-(2-deoxy-β-D-*threo*-pentofuranosyl)thymine.

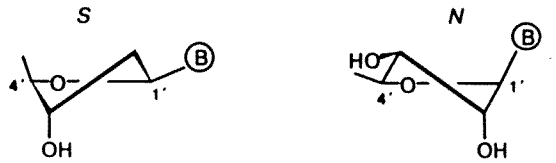


Fig. 9. Schematic presentation of typical conformations of deoxyribose and ribose in the double helix of nucleic acid.

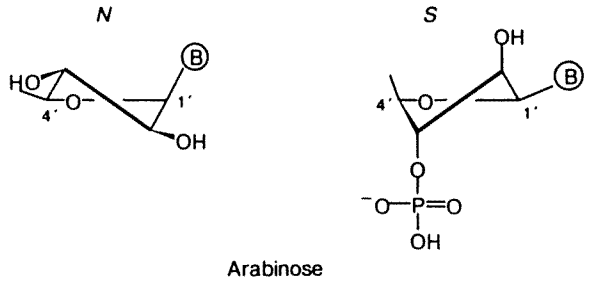


Fig. 10. Schematic presentation of conformations of arabinose in arabinonucleotides.

bond. Thus, one can expect that chemical ligation in the nucleotide duplex having A form of the helix would be much less efficient than in B-DNA. This prediction was confirmed in some experiments with synthetic RNA fragments (see below). The structural nonequivalence of duplexes **III** and **IV** also affects the different percentage of 3′—5′- and 2′—5′-bonded isomers: 87 % of natural bond in the "ligation" site of duplex **IIIg** and 35 % in duplex **IVc**. It should be noted that the same content of the natural isomer (86 %) was found in the product of BrCN-induced ligation in **IIIg**. Evidently, the activated phosphate is located at different positions in relation to the 3′- and 2′-hydroxyl groups in two series of duplexes.[32]

Let us focus our attention on the data obtained for arabinosyl-containing systems.[31,32] The value of $k_3/(k_2 + k_3)$ for the duplex with the aU unit at the 3′-end, which accepts the phosphate group (**IIIn**), is one order of magnitude smaller than that for the nonmodified duplex (see Table 3). At the same time, while the arabino-derivative is the donor of the phosphate group (duplex **IIIm** containing the aUp fragment), the value of $k_3/(k_2 + k_3)$ is comparable to that of nonmodified duplex **IIIa**. It is known that, in contrast to natural monomers, the arabinonucleotides show a clear preference for one or the other furanose conformation depending on the phosphate group position: *N*-type for paA, and the *S*-type for aAp. This regularity holds for oligomers: dinucleosidephosphate aApaA is a mixed *S*—*N*-conformer.[40] Based on the results of chemical ligation and the published data, one can postulate the *N*-type conformation for the paU unit of duplex **IIIn** (Fig. 10). When arabino-derivative is the donor of the phosphate group (duplex **IIIm**), the conformation of furanose changes from *N* to *S*, which determines the axial position of the reacting groups, which is optimal for the interaction (see Fig. 10). This conclusion does not contradict the data on the cleavage with restriction endonuclease *Eco*RI of DNA-duplex containing the aA unit in the site to be cleaved, for which the authors postulate the *S*-conformation.[41]

Other kinetic regularities are observed for duplexes **IIIo,p** containing dxT-residues.[31,32] The efficiency of chemical ligation changes slightly with the variation of the position of phosphate in the ligation site and remains substantially lower than that for nonmodified duplexes (see Table 3). This fact points to the retention of the same conformation of the 3′-terminal unit (*S*) unfavorable for chemical ligation, or to the steric hindrances due to the inversion of the 3′-hydroxyl group (phosphate).

The use of chemical ligation for the determination of the local structure of nucleic acids is based on the fact that different conformational changes in the point of break affect the efficiency of the reaction to a varying degree, but do not impede it completely. This approach complements the traditional physicochemical methods and the conformational analysis and makes it possible to estimate more reliably the conformations of residues at the single-strand break and even inside the intact double helix.

5. Chemical assembly of RNA—DNA hybrid complexes

The possibility of using EDC and BrCN for the condensation of two RNA fragments or RNA and DNA fragments in a complementary template was studied for a series of hybrid RNA—DNA duplexes (Table 4). The oligonucleotides taken for condensation differed in the number of ribonucleotide residues and the position of phosphate groups.[29,42]

The yield of the reaction product in duplex **IIIg** containing rU as the phosphate acceptor is more 50 % lower than that in the corresponding DNA—DNA duplex.[26] As was mentioned above, this can be explained by the difference in conformations between the ribo- and deoxyribonucleotide residues in the double helix. The modification of the hydroxyl groups in the 3′-terminal uridine can be an additional reason for the decreased efficiency of ligation in duplex **IIIg** under the action of BrCN. The possibility of this side reaction was confirmed by a control experiment. Localization of the phosphate group at the 3′-end of the ribose fragment (**IIIr** in Table 4) also does not intensify the chemical ligation. Activation of the 3′-phosphate in the ribonucleotide induces two competitive reactions (Scheme 5), *viz.*, intermolecular coupling of the oligomeric components (route *a*) and intramolecular formation of 2′-3′-cyclophosphate (route *b*). The latter is not a sufficiently reactive intermediate for chemical ligation; nevertheless, it can react with the 5′-terminal group of the adjacent oligomer, affording a mixture of the 2′—5′- and 3′—5′-linked isomers. Introduction of the second phosphate group (duplex **IIIs**) slightly increases

Table 4. Yield of products of chemical ligation in hybrid RNA—DNA duplexes (%)

Duplex	Composition of duplex*	Coupling reagent**	
		BrCN	EDC
IIIb	5′ ACGGAT↓*pCCAGGAGTGAC 3′ GCCTA-- GGTCCTCAC	35 (67)	75
IIIg	ACGGA(U)↓*pCCAGGAGTGAC GCCT A -- GGTCCTCAC	16	35
IIIr	*pACGGA(Up)↓CCAGGAGTGAC GCCT A - GGTCCTCAC	<5	<5
IIIs	*pACGGA(Up)↓pCCAGGAGTGAC GCCT A -- GGTCCTCAC	8	42
V	AATGG↓(*pAAAACCCAUG) TTTACC -- TTTTGGGTACC	7	10
VI	*pAATGGp↓(AAAACCCAUG) TTTACC - TTTTGGGTACC	40 (74)	90
VII	(AAUGG)↓(*pAAAACCCAUG) TTTACC --- TTTTGGGTACC	4	16

* Symbol d is omitted; ribonucleotide units (blocks) are given in parentheses; the position of the radioactive phosphate is marked by an asterisk.

** The reaction times are 3 min for BrCN and 6 days for EDC; the yield of product after three additions of BrCN is given in parentheses.

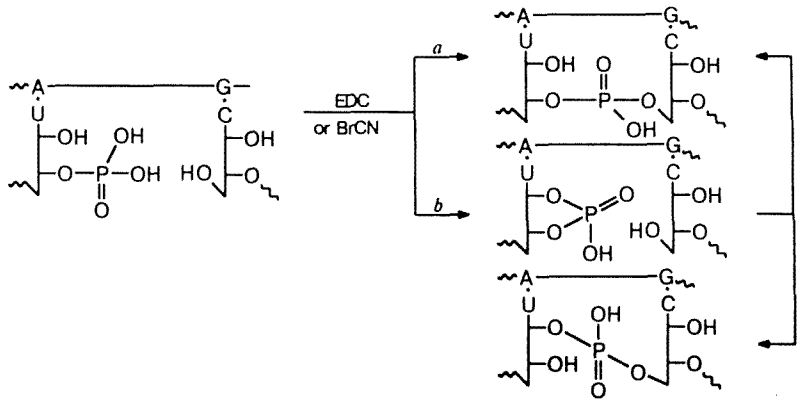
the yield of the product with the pyrophosphate bond (to 42 %), although it is quite far from 100 %, which is easily achieved in chemical ligation of the corresponding DNA blocks.

As expected, in duplexes **V** and **VI**, in which ribo- and deoxyribonucleotide fragments are assembled on a DNA template, the chemical ligation proceeds most efficiently for the combination of a 3′-phosphorylated DNA block and the oligoribonucleotide carrying the 5′-hydroxyl group (**VI**). However, the total yield of products of chemical ligation in these hybrids is lower than in their DNA analogs. A helix bend at the border of two conformations of the B form for the DNA—DNA part of the duplex and the A form for the RNA—DNA hybrid probably slighly decreases the efficiency of block-condensation. Triple treatment with BrCN increases the yields of the reaction products substantially (see Table 4).

The template directed coupling of RNA fragments is rather ineffective (**VII** in Table 4). This result is in accord with the above data and is explained by poor acceptor ability of the ribonucleotide residues for the activated phosphate group. In the double helix, the ribonucleotide units adopt a conformation unfavorable for chemical reactions. The effect of the secondary structure of duplexes on the efficiency of coupling of RNA fragments also cannot be excluded (see above). It is evident that the localization of the phosphate group at the 3′-end of an oligoribonucleotide would result in the side reaction (formation of 2′,3′-cyclophosphate) accompanying the coupling of oligomers, *i.e.*, in fact, the coupling is blocked as in duplex **IIIr**. The analysis of the ligation product in duplex **VII** (AAUGGAAACCCAUG) using RNAase T_2 demonstrated that the oligomer formed under the action of both reagents contained the natural 3′—5′-phosphodiester bond.[29,42]

Thus, the fundamental differences between the types of double helices and the chemical nature of RNA and DNA are the reasons for lower efficiency of coupling of RNA fragments using chemical reagents as compared to coupling of the DNA analogs. It is possible that the detailed study of the influence of ions of bivalent metals on the direction of the reaction will enhance the efficiency of this process. For example, the ability of Zn^{2+} and Pb^{2+} to increase the yield of products of template condensation of ribonucleotides and to result in the preferable formation of one of isomers (2′—5′- or 3′—5′-linked) has been shown.[43—45] Chemical ligation was rather efficient in the synthesis of block-copolymers

Scheme 5



of DNA and RNA. It should be taken into account that the high yields of the chemical ligation products can be achieved only when the 3′-phosphorylated oligodeoxyribonucleotide is the 5′-terminal fragment of the block-copolymer.

6. Chemical ligation as a model of an enzymatic reaction

A comparative study of chemical and enzymatic ligations, their general regularities and differences seems fruitful for deeper insight into kinetic and structural peculiarities of the reactions in complexes of nucleic acids. The template-directed reactions can be regarded as a relatively simple chemical model of the enzymatic process, which employs the principles valid for the active centers of enzymes. In fact, the phosphate and hydroxyl groups are brought together at a distance of a covalent bond in a nicked duplex due to weak, multipoint interactions, as in the sorption of a substrate on an enzyme. Here, the relative forward motion of the groups is impeded and their mutual orientation is fixed to some extent. In addition, as in the enzymatic systems, the micromedium and solvation conditions in the complementary complex differ substantially from those in solution. In the complementary complex, the free energy of complex formation is spent on lowering the barrier to free activation energy of the subsequent chemical reaction. An increase in the reactivity of the reacting groups due to interaction with the functional groups of the protein by the mechanisms of general acidic and general basic catalysis is one of the factors providing acceleration of the enzymatic process by 3—4 orders of magnitude. In chemical ligation, the enhancement of the reactivity of the reacting groups can be achieved only due to external changes: an increase in the electrophilicity of phosphomonoester by using more reactive coupling reagents or an increase in the nucleophilicity of the groups interacting with the activated phosphate. It has been demonstrated already that both factors substantially change the efficiency of chemical ligation. In addition, the rate of enzymatic processes strongly depends on the proximity and the mutual orientation of the reacting groups of the substrate in the active center. Likewise, in the systems simulating the enzymatic act, the change in the conformational situation in the reaction site due to introduced defects retards the condensation.

Nevertheless, the possibilities of chemical and enzymatic ligations and the demands made on DNA duplexes — substrates, differ strongly.[26,46]

The main requirement for the successful chemical ligation is the presence of a thermodynamically stable, double-stranded complex. The 3′-phosphate and 5′-hydroxyl group is the best combination of the reacting groups in the initial complex to form the phosphodiester bonds. The chemical method is sensitive to the nucleotide sequence near the site to be repaired. Using this method, it is possible to obtain quite efficiently the internucleotide links with a wide range of modifications of the sugar-phosphate backbone, including those with unnatural bonds.[26,29] The rates of their formation depend strongly on the nucleophilicity of the group attacking the activated phosphate, and in the case of the same chemical nature of the interacting groups, on their proximity and the character of mutual orientation, *i.e.*, the local conformation of the "linking" site. The rate of formation of phosphodiester bond can be substantially higher (BrCN) or lower (EDC) than the rate of the enzymatic reaction.

The thermostability of the double helix is not so substantial for T4 DNA ligase. The enzyme is active at the temperatures both above and below the melting temperature of the duplex.[20] The efficiency of ligation depends on the physical dimensions of the duplex; the smallest substrate should have 9 b.p. The enzyme is not very sensitive to the nucleotide sequence and catalyzes exclusively the formation of the natural internucleotide bond. Of the whole series of modified duplexes (see Table 2), the condensation was efficient (80—95 %) only in duplex **IIIg** containing the 3′-terminal rU-link at the site of ligation. In some duplexes containing $-T_{NH_2}$ (**IIIc**), —Tp (**IIId**), or dxT (**IIIp**) as the 3′-terminal unit, accumulation of 5′-adenylated oligomers is observed. In the case of duplexes **IIIe,f,k,n** and **IVf**, both reaction products and intermediate structures are absent. Apparently, the rejection of these substrates occurs already at the stages of formation of the enzyme-substrate complex. Based on the results of ligation in duplexes with mismatched pairs of bases, it is possible to conclude that the region of the double helix, whose secondary structure is important for the enzyme function, is ≥2 b.p. in the component accepting phosphate and <3 b.p. in the donor component. T4 DNA ligase was shown to catalyze the nick-sealing in duplexes **IVd,e** with "extra" bases at the ligation site (the product yield is *ca.* 50 % lower than that in the nonmodified duplex). The stage of formation of the phosphodiester bond follows the stage of elimination of the "extra" link, *i.e.*, the enzyme itself can repair some defects of the double helix.[46]

7. The effect of sequence-dependent structural variations of B-DNA on the efficiency of chemical ligation

It has been shown previously for limited experimental material (assembly of 35 b.p. duplex) that the method of chemical ligation is sensitive to the nature of the nucleotide residues to be linked. The phenomenon of dependence of the efficiency of the chemical template reactions on the nucleotide sequence (sequence-dependence) has both theoretical and practical importance, and the understanding of these regularities is essential to reveal the principles of rational disconnection of the target double-stranded DNA into blocks to be chemically assembled. Therefore, we performed a comparative study of chemical ligation for the majority of combinations of nucleotide contacts in the double DNA (14 of

16 possible).[47,48] For this purpose, three-component complexes consisting of the template oligomer and two complementary oligonucleotides were used, and the nucleotide sequences flanking the site of ligation were different. Parallel runs of experiments used both coupling reagents, EDC and BrCN. In order to make the model systems closer to the actual situation, *viz.*, the assembly of extended DNA, the intact gene containing 183 b.p. was disconnected into fragments 1—18 so that different nucleotide residues were in contact in the linking points of oligomers. A series of contacts was examined repeatedly using more prolonged templates (blocks **a**—**h**) (Fig. 11). In all sites of ligation, 3′-phosphate and 5′-hydroxyl groups were brought together. The results of chemical ligation are presented in Fig. 12, *a*. It can be seen that the yields of the reaction products vary from 17 to 94 % depending on the nature of nucleotides involved in the new phosphodiester bond formation. In some cases, the elongation of a template decreases the product yield, probably due to the presence of intramolecular structures in the components of the complex. It should be noted that both reagents show the same sequence-specific trend despite the differences in reaction rates. Taken as a whole these data suggest that the variation in the yields of products of chemical ligation can be explained by different orientation of the reacting groups, which, in turn, is related to the sequence-dependent modulations of the local structure of B-DNA. The presence of these modulations was found by X-ray analysis of monocrystals of DNA duplexes. Several hypotheses on the reasons for conformational polymorphism of DNA were proposed. One of them[49] explains the variability of the structure by local conformational tunings moderating energetically unfavorable collisions of the purine bases of the opposite chains in the minor groove (the most unfavorable collisions were observed for 5′-PyrPu-3′ sequences). According to another hypothesis, the sequence-dependent variations in the structure of B-DNA result directly from stacking interactions of the adjacent pairs of bases.[50] It seems at first glance that the presence of a single-strand break would result in serious violations of the fine structural modulations in DNA increasing the degrees of freedom of covalently nonbonded fragments of carbohydrate-phosphate backbone. However, recent studies by X-ray analysis and high-resolution NMR of break-containing and intact DNA duplexes demonstrated that a single-strand break results only in slight distortions of the structure, and the intrinsic, sequence-dependent properties are displayed in defective helices as clearly as in perfect helices.[51,52] Since ^{31}P NMR spectroscopy is a powerful tool for the study of conformational heterogeneity of the sugar-phosphate backbone, we carried out a comparative analysis of the ^{31}P NMR data[53] and the results of chemical ligation for the identical nucleotide contacts in the double helix (Fig. 12, *b*). A general trend of a change in the efficiency of ligation and in chemical shifts of signals of individual phosphates in oligonucleotide duplexes related to the torsion angles of phosphodiester bonds and local geometric parameters of the helix is clearly demonstrated: the downfield shift of the ^{31}P signal and the decrease in efficiency of ligation for the 5′-PyrPu-3′ sequence; the upfield shift and the parallel increase in the yield of the ligation product for the dinucleotide combinations: 5′-GpT-3′, 5′-TpT-3′, and 5′-CpT-3′. A plot of the positional dependence of

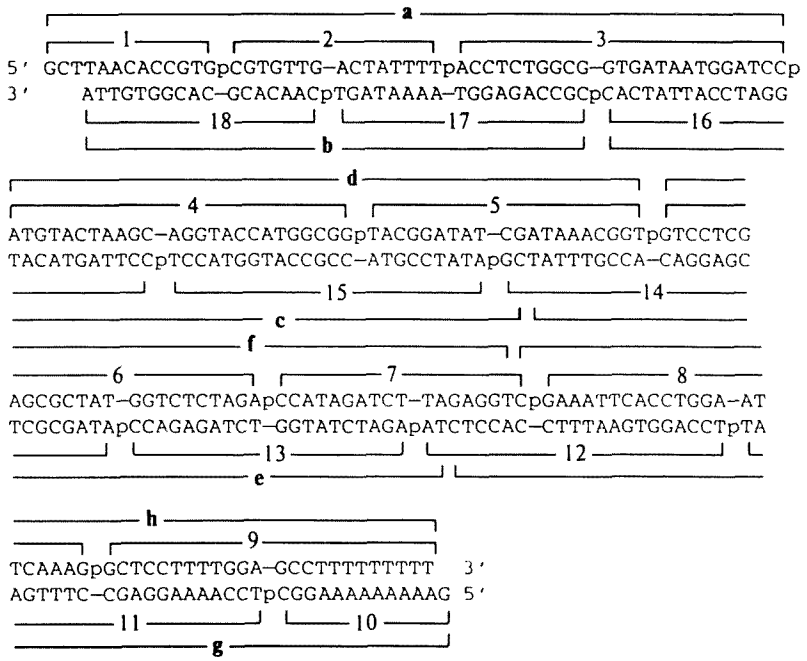


Fig. 11. Disconnection of 183 b.p. DNA duplex into blocks. Short fragments are designated by numbers, larger fragments are designated by letters. The site of chemical ligation is marked by symbol p between the fragments.

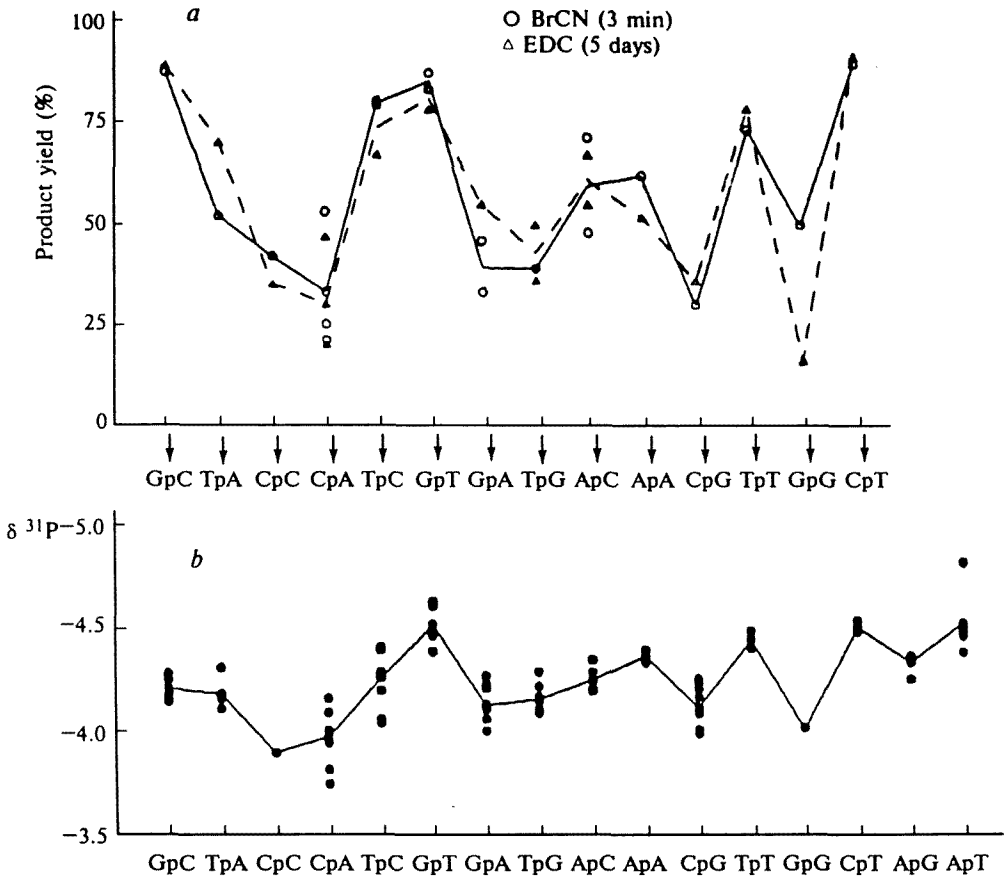


Fig. 12. *a.* Yield of products of chemical ligation depending on the nature of nucleotide residues involved in the formation of a new phosphodiester bond; the direction of the chain is 5′→3′. Averaged values of yields for different duplexes with equal ligation sites are connected by straight lines. *b.* Dependence of ^{31}P chemical shifts on the nature of nucleotide residues around the definite phosphodiester bond in B-DNA. Averaged values of chemical shifts for seven oligonucleotide duplexes are connected by straight lines.[53]

the chemical shift of the ^{31}P signal contains the data for 5′-ApT-3′ and 5′-ApG-3′, which are absent in Fig. 12, *a*. If they are taken into account, it is possible to predict the yields of products of chemical ligation for the ApT contact (80—90 %) and for the ApG contact (40—50 %). Additional experiments[54] revealed good agreement with these predictions. The correlation between some other parameters used for comparison was also observed in other systems. For example, it is known that the B → A transition is accompanied by a decrease of the local helical twist from 36.0 to 32.7°. This is manifested in the cooperative downfield shift of ^{31}P signals by 0.2—0.3 ppm[55] and a decrease in yields of products of chemical ligation by 50—67 % upon replacement of oligodeoxyribonucleotides by RNA fragments (see above). The direct comparison of the efficiency of ligation of different combinations of nucleotide residues with the experimental (X-ray analysis) or computed sequence dependence of the local helical twist in d(CGCGAATTCGCG)[50] also demonstrates the parallelism in variations of these parameters.[48]

The results indicate direct correlation between the reactivity and mutual orientation of the reacting groups determined by the general and local structures of the double helix. Chemical ligation was found to be sensitive even to relatively weak sequence-dependent modulations of the conformation of nucleic acids. This conclusion is substantially confirmed by the fact that the revealed correlation is practically independent of the nature of the coupling reagent. These results support the assumption that single-strand breaks barely disturb the local structure of the site and the "ruptured ends" are frozen in conformations determined by the sequence-dependent properties of DNA. In the first approximation, the reactivity of two nucleotide residues can be predicted on the basis of the model of "nearest neighbors", which describes helical parameters in terms of the structure of complementary dinucleotides, neglecting long-range effects. In fact, different nucleotide sequences flanking dinucleotide contacts do not violate the above regularities. These findings open a way to chemical probing of the fine structure of unusual forms of nucleic acids: triplexes, tetraplexes, parallel duplexes, *etc.*, and also of peculiarities of local conformations of the sugar-phosphate backbone of natural and modified DNA (RNA) helices. Finally, practical information on a series of dinucleotide contacts in order of an increase in efficiency of ligation makes it possible to carry out an

optimal disconnection of a target DNA-duplex into blocks. On the basis of the resulting regularities, a 183-b.p. gene has been broken down into eight fragments and assembled successfully using BrCN.[56]

8. Probing of structure of triple helices using chemical ligation

Taking into account the high sensitivity of chemical ligation to the orientation of the reacting groups fixed by the secondary (tertiary) structure of the DNA-helix, we attempted to use this method to monitor the changes in the carbohydrate-phosphate backbone when unusual forms of nucleic acids are formed. Our attention was attracted by triple-stranded PyrPuPyr DNAs, the interest in which is connected with their presumed role in the regulation of the eukaryotic genome (H-form of DNA). In addition, the study of triple helices is of interest with regard to the design of oligonucleotide ligands, which recognize double-stranded targets. These ligands can serve as artificial repressors of genes and as DNA-cleaving agents. It is supposed that the triple helices exist in A form, while purine and one of the pyrimidine strands are bound by Watson-Crick pairs and the second pyrimidine strand lying in the major groove of the double helix parallel to the purine strand is bound with the latter by Hoogsteen pairs. Chemical approaches, along with physical methods, are used rather widely to study the molecular structure of triplexes. They involve the stability of triple-stranded sites to chemical modifications and specific cleavage of the double helix. It has been shown recently that the double-stranded DNA can serve as the template for condensation of homopyrimidine fragments of the third strand.[57] We have synthesized a

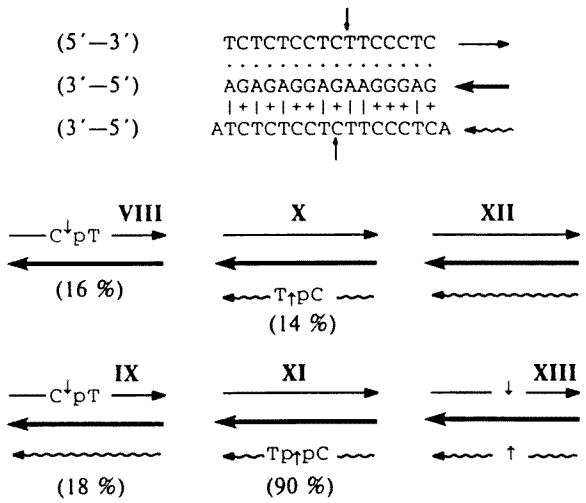


Fig. 13. Triple-stranded PyrPuPyr complexes. Vertical arrows mark the points of connections of oligomeric blocks. Pyrimidine (⟶) and purine (⟵) oligomers are bound by Watson-Crick pairs; oligomer ⟵ is the third chain bound by Hoogsteen pairs. The triple helices investigated are given below. The nucleotide residues in the reaction site are marked. The yield of chemical ligation product is given in parentheses.

series of DNA duplexes and triplexes including complexes containing nicks (Fig. 13); this series made it possible to follow the changes in the sugar-phosphate backbone upon conversion of DNA duplex into triplex and binding of the third chain.[58]

Since the two homopyrimidine strands of the triplex are antiparallel, they cannot replace each other in triple helices with an irregular sequences. This allows one to follow independently the EDC-induced ligation in both pyrimidine strands and to compare the reactivity and, hence, the mutual disposition of the interacting phosphate and hydroxyl groups in the triplex and in the initial duplex. The choice of the nucleotide sequence of oligomers was determined by the binding code of the third chain (protonated C recognizes the G · C pair, and T recognizes the A · T pair). A preliminary study of CD spectra and the thermal stability of the designed complexes confirmed the formation of the triple helix in the studied systems. All of the triplex-forming mixtures exhibit biphasic melting (Fig. 14); the first step corresponds to dissociation of the component bound by weaker Hoogsteen pairs. The break in the third chain substantially decreases its dissociation temperature (*cf.* Fig. 14, *a*, *b*).

The efficiency of condensation of pyrimidine oligonucleotides bound by Watson-Crick pairs is almost the same in duplex **VIII** and triplex **IX** (see Fig. 13). This result shows that there are no substantial structural rearrangements in the carbohydrate-phosphate backbone upon the addition of the third chain, because, as mentioned above, the difference in conformations of the carbohydrate part, *N* (A form) and *S* (B form), significantly affects the efficiency of chemical ligation (the yields of ligation products are substantially lower for the A family). The low yield of ligation (16—18 %) possibly indicates that the original duplex is in the A-like conformation adapted for binding of the third chain. According to results of ligation in triplexes **X** and **XI** (see Fig. 13) one can conclude that the reacting groups are brought together due to binding of the "broken" third chain with the initial duplex. It is interesting that the yields of ligation products in different pyrimidine strands of the same triplex are equal (*cf.* systems **IX** and **X** in Fig. 13); it should be noted that in both cases identical nucleotide residues are "linked". These data may be interpreted as follows. In both pyrimidine strands of the triple-stranded complex, the positions of the reacting groups are similar despite the difference in their hydrogen binding to the central purine strand. These results are in accord with the data on X-ray diffraction and NMR-spectroscopy as regards the conformation of the furanose residues in the triple helix.[59]

9. Chemical ligation in branched DNAs

In practice, chemical ligation seems to be applied for designing structures of unusual topology, because in this case, the enzymatic approach would be inappropriate due to steric hindrances or non-canonical geometry of

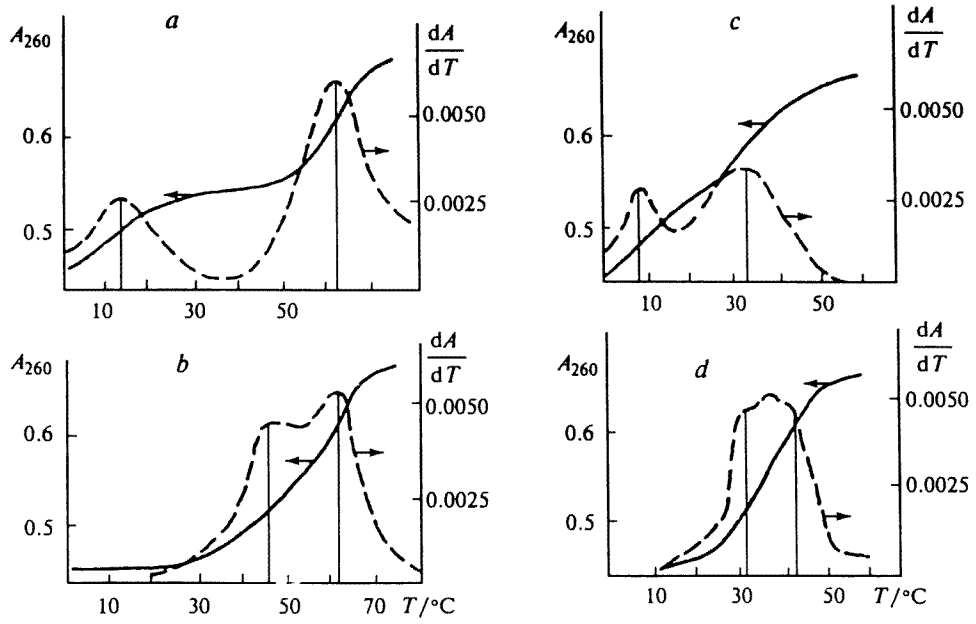
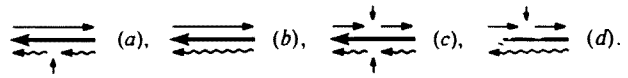


Fig. 14. Melting profiles of oligomeric DNA triplexes:



The curves in differential form are plotted as dashed lines.

the substrates. In this section, the data on designing extended, branched oligonucleotides (BO) using the chemical ligation techique are presented. The interest in these compounds is associated with the possibility of a multiple increase in the sensitivity of determination of nucleic acids using hybridization probes due to amplification of a signal. The BO should contain a fragment of the unique target (interacting with the oligonucleotide sequence DNA) covalently linked with several copies of another sequence hybridized with enzyme-labelled probes. Schematically, the structure of BO can be represented as a comb (⊣⊓⊓). The current level of oligonucleotide synthesis makes it possible to obtain BO using a DNA-synthesizer with only short "teeth" containing no more than six nucleotide residues.[60] The problem was to add a 24 b.p. linear oligonucleotide to each of five "teeth" of the original BO. In the development of the strategy of assembly and designing of the ligation site, all the factors that could intensify the template reaction were taken into account: the existence of a stable complex, which enables the reacting groups on the template to draw closer together, localization of the phosphate group at the 3′-end of the oligonucleotide, the optimal combination of interacting nucleotide residues, introduction of groups of higher nucleophilicity to serve as acceptors of the activated phosphate, and the use of different activating reagents taking into account their advantages and drawbacks. The elongation of short sequences of BO is illustrated in Fig. 15. To stabilize the complementary complex, chimeric, branched compound (**6**), which represents a template (**5**) covalently linked with the incoming oligomer (**4**), was also used (see Fig. 15).[61]

At the first stage, this complex system was replaced by a simple model, *viz.*, a three-component complex consisting of oligomers **4** and **5** and a 37-membered, linear oligonucleotide, which mimics one branch of structures **1**—**3**. The experiments demonstrated that 24- and 37-mers can be coupled very efficienly on a 12-membered template yielding 80—85 % of a product with the natural bond and >95 % of products with the pyrophosphate or phosphoramidate bonds. BrCN was used to form the phosphodiester bond, and the modified bonds were formed under the action of EDC. The conditions found were used for addition of 24-mers to each of five branches of the initial BO. Prior to the addition of the chemical reagent, the reaction mixtures were "annealed" for 2—3-h and then analyzed by capillary electrophoresis in 8 % denaturing polyacrylamide gel. This procedure allowed us to obtain quantitative data on distribution of substituted (from mono- to pentasubstituted) derivatives. The most favorable distribution of products of chemical ligation was achieved in pyrophosphate-linked BO (Fig. 16, *a*): 17 % of tri-, 43 % of tetra-, and 39 % of pentasubstituted compounds. For the phosphoramidate-linked BO, the product ratios were 31, 39, and 19 %, respectively, and the percentage of disubstituted derivative was 9 %. The least efficient condensation occurred in the case of attack of the activated phosphate with the 5′-hydroxyl group: the disubstituted compound

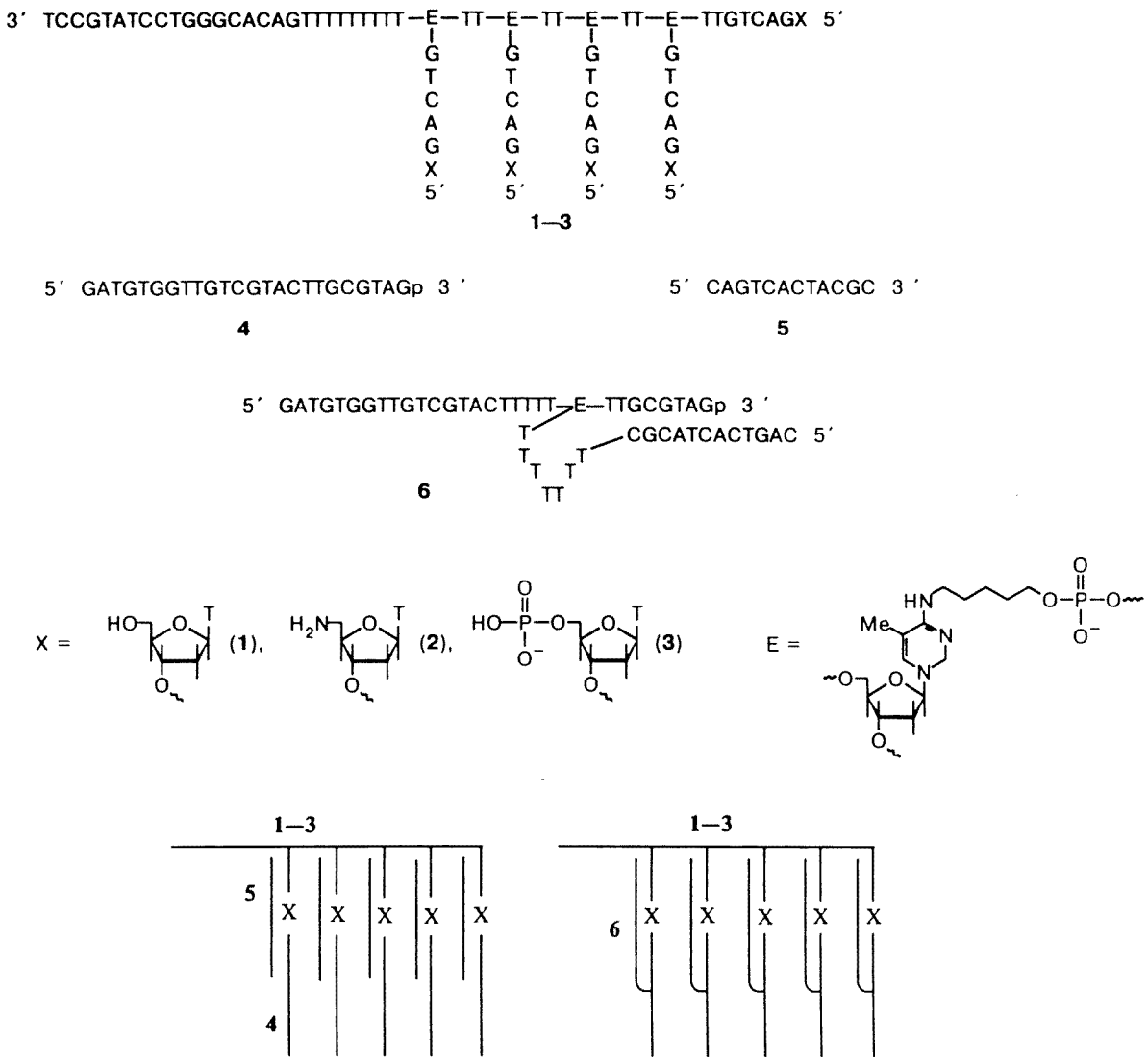


Fig. 15. Chemical assembly of branched DNAs. **1**–**6** are the oligonucleotide components of reaction mixtures. The duplexes based on BO, which brings together the reacting groups, are schematically presented below.

was the major product (33 %) (Fig. 16, *b*). It is interesting that the peak corresponding to the monosubstituted derivative exhibits fine structure indicating the presence of positional isomers having different mobilities in gel. The use of the hybrid autolink as the attachable component, which is both the template and the incoming oligomer (**6** in Fig. 15), significantly intensifies the chemical ligation. Thus, in the amplification multimer with the phosphodiester and phosphoramidate bonds, tetra- and pentasubstituted derivatives are the major products, and the pentasubstituted BO is the major pyrophosphate-containing product. This effect is achieved due to the formation of more stable complexes in this unusual combination of intra- and intermolecular complementary interactions.[61]

The excellent possibilities of chemical ligation were demonstrated in the preparation of BO with 15 lateral "teeth". Branched DNAs with more than 12-fold substitution, form the majority of the pyrophosphate-linked product. The compound formed as a result of more than 12 parallel acts of chemical ligation contain >1000 nucleotide residues.[61]

10. Circularization of oligonucleotides using chemical ligation

The circular oligodeoxyribonucleotides (CO) are yet another example of unusual structure obtained successfully using chemical ligation. The interest in these compounds arises from their resistance to exonuclease degradation. Some peculiarities of the design of linear precursors of CO were predicted by us *a priori*. Thus, their length should be sufficient to form a stable duplex with overlapping of at least 6+6 b.p., whereas the bending site should contain *ca.* 22 nucleotide residues;

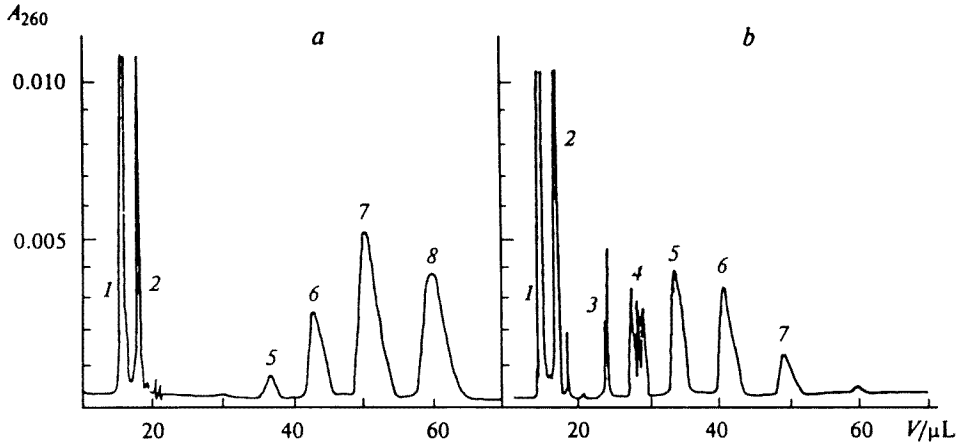


Fig. 16. Capillary gel electrophoresis of products of chemical ligation in branched DNAs. *a*. EDC-induced formation of derivatives with pyrophosphate bonds (**3+4+5** in Fig. 15); 16 h of incubation at 0 °C. *b*. BrCN-induced reaction in a mixture of oligomers **1+4+5** (formation of a phosphodiester bond); 5 min of incubation at 0 °C. Peaks *1*, *2* and 3 correspond to 12-membered template, 24-mer and the starting BO with 5 "teeth"; *4*—*8* correspond to BOs with 1—5 side substituents, respectively.

the phosphate group should be located at the 3′-end of the oligonucleotide, and in the ligation site, the most reactive nucleotides, —Cp↓T— or —Tp↓T—, should be in contact; BrCN should be used to form the phosphodiester bond. To meet these requirements, a model system **XIV** was synthesized, for which the main regularities of chemical circularization of oligonucleotides were revealed (Fig. 17). It was shown that the increase in the concentration of a nucleotide favors the dimerization of the initial oligonucleotide competing with concatemerization. At $C_0 = 10^{-4}$ mol L^{-1}, the yield of circular product having electrophoretic mobility intermediate between those of the linear precursor and its dimer (in 20 % denaturing polyacrylamide gel), was 80—90 %.[62] The circular nature of the ligation product was confirmed by its resistance to the mixture of enzymes, snake venom phosphodiesterase and bacterial alkaline phosphatase, resistance to the introduction of 5′-^{32}P-label and the unusual character of the cleavage at one of the bases (Maxam—Gilbert analysis). This analysis could be performed if the 5′-^{32}P-label was included in the linear precursor before the circularization; the pyrophosphate bond was formed as a result of chemical ligation. This original approach gives an unequivocal identification of the initial oligonucleotide, which gives a normal set of bands corresponding to the position of the given base in the chain, and the cyclic oligonucleotide, whose electrophoretic mobility changes abruptly after the first break of the chain so that it coincides with the mobility of the initial oligomer; its nucleotide sequence is "unreadable" due to the internal, not the terminal position of the radioactive label in the chain under question (Fig. 18).

The study of other oligonucleotide systems demonstrated that the length of the linear precursor should exceed the length of the template more than twice. The secondary structure of the initial oligonucleotide is the most impotant factor affecting the efficiency of the circularization. The presence of an intramolecular hair-

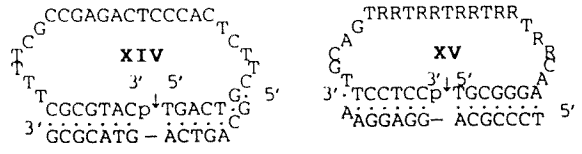


Fig. 17. Scheme of formation of cyclic oligonucleotides. R are the residues of 1,2-dideoxy-D-*erythro*-pentofuranose.

Fig. 18. Autoradiogram of electrophoresis in 20 % polyacrylamide gel of (*2*) the linear precursor and (*4*) the product of its circularization in duplex **XIV**; their cleavage at the residues dG under conditions of Maxam—Gilbert sequencing (*1* and *3* respectively).

pin, which leaves the terminal sites of the molecule free, is desirable to achieve a high yield of circular products. If these sites are involved in the formation of the secondary structure, chemical circularization is inhibited completely. It is clear that unfavorable intramolecular interactions are virtually unavoidable, if the linear precursor contains 30—40 nucleotide residues. One of the approaches to escape conformational problems can consist in the replacement of insignificant sequences in the original oligomer by nonnucleotide inserts. This approach also decreases the possibility of nonspecific interactions of the probe with the cellular nucleic acid, which is the main problem in therapeutic application of antisense oligonucleotides. The alternating residues of 1,2-dideoxy-D-*erythro*-pentofuranose (R) and thymidine were suggested as such inserts.[63] The $(RRT)_5$ block has been inserted into an oligonucleotide, whose terminal fragments after circularization should form a 18-membered region, which binds specifically with RNA of herpes simplex virus. Under conditions optimal for circularization of nucleotide analogs, the mixed oligomer forms the cyclic structure in a yield close to 100 % in the presence of a 14-membered template (duplex **XV**) (BrCN-induced chemical ligation).

Thus, chemical ligation is an efficient method for circularization of oligonucleotides not only in triple-stranded complexes,[64] but also in standard double helices.

* * *

In the present review, chemical ligation is discussed in two aspects. First, this reaction is of interest as a new method for the assembly of genetic structures including modified and anomalous ones. Using this method, several series of modified, functionally important DNA fragments and RNA—DNA block-copolymers have been obtained and an extended DNA duplex (gene) has been assembled. Efficient methods for the preparation of branched and circular oligonucleotides have been developed. In template-directed coupling of fragments of nucleic acids, water-soluble carbodiimides and BrCN were used. BrCN is a new, highly effective reagent for chemical ligation providing superfast reactions in DNA duplexes (1—3 min), which do not form side products.

Second, it was shown that chemical ligation can serve as a tool for probing the structure of nucleic acids. On the basis of the structural-kinetic concept developed, the data on the conformation of ribo- and arabinonucleotides in DNA duplexes have been obtained. The effect of noncomplementary pairs, single-strand deletions, and inserts on the structure of the double helix has been estimated. The interrelations between the reactivity of the interacting groups and sequence-dependent local conformation of the ligation site in B-DNA were determined. Chemical ligation was used as a sensitive monitor of stable duplex formation and third-strand binding. The development of chemical ligation stimulated studies of the physicochemical properties and the structure of DNA-like duplexes.

Evidently, the possibilities of chemical ligation are not limited to the uses studied. In future, the area of its application will expand due to the increasing demand of molecular biology and biotechnology for new methods of synthesis and investigation of artificial analogs of nucleic acids.

The authors are grateful to their colleagues N. I. Sokolova, T. S. Oretskaya, and E. M. Volkov from the Laboratory of the Chemistry of Nucleic Acids for their active assistance and the synthesis of oligodeoxyribonucleotides.

References

1. R. Naylor and P. T. Gilham, *Biochemistry*, 1966, **5**, 2722.
2. H. Schneider-Bernloehr, R. Lohrmann, J. Sulston, B. J. Weimann, and L. E. Orgel, *J. Mol. Biol.*, 1968, **37**, 151.
3. R. Lohrmann and L. E. Orgel, *Nature*, 1973, **244**, 418.
4. Z. A. Shabarova and M. A. Prokofiev, *FEBS Letters*, 1970, **11**, 237.
5. A. G. Badashkeeva, G. N. Kabasheva, D. G. Knorre, G. G. Shamovskii, and T. N. Shubina, *Dokl. Akad. Nauk SSSR*, 1972, **206**, 870 [*Dokl. Chem.*, 1972, **206** (Engl. Transl.)].
6. B. S. Sproat and M. J. Gait, in *Oligonucleotide Synthesis: A Practical Approach*, Ed. M. J. Gait, IRL Press, Oxford, UK, 1984, Ch. 4.
7. V. K. Potapov, V. P. Veiko, O. N. Koroleva, and Z. A. Shabarova, *Nucleic Acids Res.*, 1979, **6**, 2040.
8. D. G. Knorre, N. V. Melamed, S. G. Popov, G. G. Shamovskii, and T. N. Shubina, *Dokl. Akad. Nauk SSSR*, 1974, **218**, 113 [*Dokl. Chem.*, 1974, **218** (Engl. Transl.)].
9. N. G. Dolinnaya, E. S. Gromova, E. V. Il'ina, N. F. Sergeeva, Z. A. Shabarova, and M. A. Prokofiev, *Bioorg. Khim.*, 1975, **1**, 1296 [*Sov. J. Bioorgan. Chem.*, 1975, **1** (Engl. Transl.)].
10. N. G. Dolinnaya, E. S. Gromova, S. N. Mikhailov, and Z. A. Shabarova, *Bioorg. Khim.*, 1978, **4**, 535 [*Sov. J. Bioorgan. Chem.*, 1978, **4** (Engl. Transl.)].
11. N. G. Dolinnaya and Z. A. Shabarova, *Bioorg. Khim.*, 1980, **6**, 209 [*Sov. J. Bioorgan. Chem.*, 1980, **6** (Engl. Transl.)].
12. Z. A. Shabarova, N. G. Dolinnaya, S. I. Turkin, and E. S. Gromova, *Nucleic Acids Res.*, 1980, **8**, 2413.
13. N. G. Dolinnaya, V. L. Drutsa, and Z. A. Shabarova, *Nucleic Acids Res., Symposium Series*, 1981, No. 9, 255.
14. Z. A. Shabarova, N. G. Dolinnaya, V. L. Drutsa, N. P. Melnikova, and A. A. Purmal, *Nucleic Acids Res.*, 1981, **9**, 5747.
15. D. Porschke and F. Eggers, *Eur. J. Biochem.*, 1972, **26**, 490.
16. O. N. Koroleva, V. L. Drutsa, N. G. Dolinnaya, A. V. Tsytovich, and Z. A. Shabarova, *Molekulyar. Biologiya*, 1984, **18**, 146 [*Molecular Biology*, 1984, **18** (Engl. Transl.)].
17. E. S. Gromova, M. N. Vinogradova, A. A. Elov, V. P. Veiko, N. G. Dolinnaya, V. L. Drutsa, V. G. Metelev, T. S. Oretskaya, and Z. A. Shabarova, *Molekulyar. Biologiya*, 1984, **18**, 370 [*Molecular Biology*, 1984, **18** (Engl. Transl.)].
18. N. G. Dolinnaya, A. A. Purmal, V. L. Drutsa, and Z. A. Shabarova, *Vestn. Mosk. Univ., Ser. Khim.*, 1986, **27**, 520 [*Moscow Univ. Bull., Chem.*, 1986, **27** (Engl. Transl.)].
19. E. S. Gromova and Z. A. Shabarova, in *Progress in Nucleic Acid Research and Molecular Biology*, Academic Press, Inc., 1990, **39**, 1.

20. Z. A. Shabarova, V. P. Veyko, N. G. Dolinnaya, V. L. Drutsa, V. G. Metelev, T. S. Oretskaya, and A. A. Purmal', *Bioorg. Khim.*, 1987, **13**, 628 [*Sov. J. Bioorgan. Chem.*, 1987, **13** (Engl. Transl.)].
21. N. I. Sokolova, D. T. Ashirbekova, N. G. Dolinnaya, and Z. A. Shabarova, *Bioorg. Khim.*, 1987, **13**, 1286 [*Sov. J. Bioorgan. Chem.*, 1987, **13** (Engl. Transl.)].
22. N. I. Sokolova, D. T. Ashirbekova, N. G. Dolinnaya, and Z. A. Shabarova, *FEBS Letters*, 1988, **232**, 153.
23. E. Kanaya and H. Yanagawa, *Biochemistry*, 1986, **25**, 7423.
24. O. I. Gryaznova, N. G. Dolinnaya, M. G. Isagulyants, V. G. Metelev, T. S. Oretskaya, N. I. Udalov, N. I. Sokolova, and Z. A. Shabarova, *Bioorg. Khim.*, 1986, **12**, 124 [*Sov. J. Bioorgan. Chem.*, 1986, **12** (Engl. Transl.)].
25. A. V. Tsytovich, T. S. Oretskaya, N. G. Dolinnaya, N. I. Sokolova, and Z. A. Shabarova, *Khim. Prir. Soedin.*, 1987, **3**, 421 [*Chem. Nat. Compd.*, 1987, **3** (Engl. Transl.)].
26. N. G. Dolinnaya, N. I. Sokolova, O. I. Gryaznova, and Z. A. Shabarova, *Nucleic Acids Res.*, 1988, **16**, 3721.
27. N. G. Dolinnaya and O. I. Gryaznova, *Usp. Khim.*, 1989, **58**, 1318 [*Russ. Chem. Rev.*, 1989, **58** (Engl. Transl.)].
28. N. G. Dolinnaya, O. I. Gryaznova, N. I. Sokolova, and Z. A. Shabarova, *Bioorg. Khim.*, 1986, **12**, 755 [*Sov. J. Bioorgan. Chem.*, 1986, **12** (Engl. Transl.)].
29. N. G. Dolinnaya, N. I. Sokolova, D. T. Ashirbekova, and Z. A. Shabarova, *Nucleic Acid Res.*, 1991, **19**, 3067.
30. Z. A. Shabarova, N. G. Dolinnaya, and N. I. Sokolova, in *Nucleic Acid Chemistry. Improved and New Synthetic Procedures, Methods and Techniques*, Eds. L. B. Townsend and R. S. Tipson, J. Wiley and Sons, New York, 1991, 348.
31. N. G. Dolinnaya, A. V. Tsytovich, S. G. Tevosyan, V. N. Sergeev, and Z. A. Shabarova, *Bioorg. Khim.*, 1990, **16**, 1195 [*Sov. J. Bioorgan. Chem.*, 1990, **16** (Engl. Transl.)].
32. N. G. Dolinnaya, A. V. Tsytovich, V. N. Sergeev, and T. S. Oretskaya, *Nucleic Acids Res.*, 1991, **19**, 3073.
33. D. T. Ashirbekova, N. I. Sokolova, N. G. Dolinnaya, and Z. A. Shabarova, *Bioorg. Khim.*, 1989, **15**, 166 [*Sov. J. Bioorgan. Chem.*, 1989, **15** (Engl. Transl.)].
34. N. G. Dolinnaya, O. I. Gryaznova, N. I. Sokolova, and Z. A. Shabarova, *Bioorg. Khim.*, 1986, **12**, 921 [*Sov. J. Bioorgan. Chem.*, 1986, **12** (Engl. Transl.)].
35. N. G. Dolinnaya, A. V. Tsytovich, V. N. Sergeev, M. N. Gertsuk, and Z. A. Shabarova, *Bioorg. Khim.*, 1990, **16**, 1183 [*Sov. J. Bioorgan. Chem.*, 1990, **16** (Engl. Transl.)].
36. N. V. Melamed, S. G. Popov, and G. G. Shamovskii, *Izv. Sib. Otd. Akad. Nauk SSSR, Ser. Khim. Nauk*, 1974, 90 [*Bull. Sib. Branch Acad. Sci. USSR, Div. Chem. Sci.*, 1974 (Engl. Transl.)].
37. M. N. Gertsuk, V. N. Sergeev, A. V. Tsytovich, N. G. Dolinnaya, and V. P. Kukhar', *Bioorg. Khim.*, 1990, **16**, 1268 [*Sov. J. Bioorgan. Chem.*, 1990, **16** (Engl. Transl.)].
38. W. Saenger, in *Principles of Nucleic Acids Structure*, Springer-Verlag, New York, 1984, Ch. 17.
39. M. Salazar, J. J. Champoux, and B. R. Reid, *Biochemistry*, 1993, **32**, 739.
40. J. Doornbos, J.-L. Barascut, H. Lazrek, J.-L. Imbach, J. van Westrenen, G. M. Visser, J. H. van Boom, and C. Altona, *Nucleic Acids Res.*, 1983, **11**, 4583.
41. E. Ohtsuka, Y. Ishino, K. Ibaraki, and M. Ikehara, *Eur. J. Biochem.*, 1984, **139**, 447.
42. N. G. Dolinnaya, D. T. Ashirbekova, N. I. Sokolova, and Z. A. Shabarova, *Bioorg. Khim.*, 1989, **15**, 1346 [*Sov. J. Bioorgan. Chem.*, 1989, **15** (Engl. Transl.)].
43. R. Lohrmann and L. E. Orgel, *J. Mol. Biol.*, 1980, **142**, 555.
44. P. K. Bridson and L. E. Orgel, *J. Mol. Biol.*, 1980, **144**, 567.
45. J. H. G. van Roode and L. E. Orgel, *J. Mol. Biol.*, 1980, **144**, 579.
46. A. V. Tsytovich, N. G. Dolinnaya, and Z. A. Shabarova, *Molekulyar. Biologiya*, 1988, **22**, 690 [*Molecular Biology*, 1988, **22** (Engl. Transl.)].
47. I. N. Merenkova, N. G. Dolinnaya, T. S. Oretskaya, N. I. Sokolova, and Z. A. Shabarova, *Bioorg. Khim.*, 1992, **18**, 85 [*Russ. J. Bioorg. Chem.*, 1992, **18** (Engl. Transl.)].
48. N. G. Dolinnaya, I. N. Merenkova, and Z. A. Shabarova, *Nucleosides and Nucleotides*, 1994, **13**, 2169.
49. C. R. Calladine, *J. Mol. Biol.*, 1982, **161**, 343.
50. A. Sarai, J. Mazur, R. Nussinov, and R. L. Jernigan, *Biochemistry*, 1988, **27**, 8498.
51. J. Aymami, M. Coll, G. A. van der Marel, J. H. van Boom, A. H.-J. Wang, and A. Rich, *Proc. Natl. Acad. Sci. USA*, 1990, **87**, 2526.
52. J. M. L. Pieters, R. M. W. Mans, H. van den Elst, G. A. van der Marel, J. H. van Boom, and C. Altona, *Nucleic Acids Res.*, 1989, **17**, 4551.
53. D. G. Gorenstein, S. A. Schroeder, J. M. Fu, J. T. Metz, V. Roongta, and C. R. Jones, *Biochemistry*, **27**, 7223.
54. I. N. Merenkova, T. S. Oretskaya, N. I. Sokolova, and Z. A. Shabarova, *Bioorg. Khim.*, 1993, **19**, 1205 [*Russ. J. Bioorg. Chem.*, 1993, **19** (Engl. Transl.)].
55. J. Kypr, V. Sklenar, and M. Vorličkova, *Biopolymers*, 1986, **25**, 1803.
56. Z. A. Shabarova, I. N. Merenkova, T. S. Oretskaya, N. I. Sokolova, E. A. Skripkin, E. V. Alexeyeva, A. G. Balakin, and A. A. Bogdanov, *Nucleic Acids Res.*, 1991, **19**, 4247.
57. K. J. Luebke and P. B. Dervan, *J. Am. Chem. Soc.*, 1989, **111**, 8733.
58. N. G. Dolinnaya, O. V. Pyatrauskene, and Z. A. Shabarova, *FEBS Letters*, 1991, **284**, 232.
59. P. Rajagopal and J. Feigon, *Nature*, 1989, **339**, 637.
60. T. Horn and M. S. Urdea, *Nucleic Acids Res.*, 1989, **17**, 6959.
61. N. G. Dolinnaya, S. Gryaznov, Z. A. Shabarova, M. S. Urdea, and T. Horn, *Bioorg. and Med. Chem. Lett.*, 1994, **4**, 1011.
62. N. G. Dolinnaya, M. Blumenfeld, I. N. Merenkova, T. S. Oretskaya, N. F. Krynetskaya, M. G. Ivanovskaya, M. Vasseur, and Z. A. Shabarova, *Nucleic Acids Res.*, 1993, **21**, 5403.
63. T. S. Oretskaya, N. G. Dolinnaya, I. N. Merenkova, N. F. Krynetskaya, E. A. Romanova, E. M. Volkov, M. Blumenfel'd, M. Vasseur, and Z. A. Shabarova, *Bioorg. Khim.*, 1994, **20**, 63 [*Russ. J. Bioorg. Chem.*, 1994, **20** (Engl. Transl.)].
64. G. Prakash and E. T. Kool, *J. Chem. Soc., Chem. Commun.*, 1991, 1161.

Received March 29, 1996