Cold Spring Harbor Symposia on Quantitative Biology

Synthetic Sequence-specific Ligands

G.V. Gursky, A.S. Zasedatelev, A.L. Zhuze, et al.

Cold Spring Harb Symp Quant Biol 1983 47: 367-378
Access the most recent version at doi:10.1101/SQB.1983.047.01.043

Synthetic Sequence-specific Ligands

G.V. GURSKY, A.S. ZASEDATELEV, A.L. ZHUZE, A.A. KHORLIN, S.L. GROKHOVSKY, S.A. STRELTSOV, A.N. SUROVAYA, S.M. NIKITIN, A.S. KRYLOV, V.O. RETCHINSKY, M.V. MIKHAILOV, R.S. BEABEALASHVILI, AND B.P. GOTTIKH

Institute of Molecular Biology, Academy of Sciences of USSR, Moscow 117984, USSR

There is much interest in the design and synthesis of chemical compounds that possess binding specificity comparable with that of bacterial repressors. These compounds could be used to affect selectively the activity of certain genes in bacterial and eukaryotic cells and might have important implications in pharmacology. In this paper we report several approaches to the design and synthesis of sequence-specific ligands.

Present knowledge of the cccurrence of base-pair or base-sequence specificity among various antibiotics and drugs is restricted largely to the DNA complexes with antibiotics of the distamycin class (Zimmer et al. 1971; Zasedatelev et al. 1974, 1976; Wartell et al. 1974; Zimmer 1975; Kolchinsky et al. 1975; Gursky et al. 1977; Luck et al. 1977; Krylov et al. 1979) and to the interaction of actinomycin D (AMD) and related molecules with DNA (Gellert et al. 1965; Cerami et al. 1967; Müller and Crothers 1968, 1975).

Distamycin A (Dst) and netropsin (Nt) (Fig. 1) exhibit a pronounced specificity of binding to AT-rich regions on DNA (Zimmer et al. 1971; Zimmer 1975). Zasedatelev et al. (1976) have suggested a detailed stereochemical model for the binding of Dst and Nt to DNA that is strongly supported by several lines of evidence (for a compilation, see Zimmer 1975; Gursky et al. 1977). According to the model, the bound Dst molecule extends more than 5 bp in the minor DNA groove and is attached to DNA base pairs through hydrogen bonds that connect the four imido groups of the antibiotic peptide links with four possible acceptor sites, represented by adenine N3 and pyrimidine O2 atoms lying in one and the same polynucleotide strand. The binding is stereospecific in the sense that the N-C4-C′ sequence in the antibiotic molecule coincides with the C5′-C3′ direction in the polynucleotide chain.

In the first part of this paper, we summarize our attempts to evaluate the interaction free energies of a Dst amide group with each of the four DNA bases. We believe that the binding principles found for Dst may show a way to the synthesis of ligands with defined sequence specificity and that these principles, in a generalized form, may describe specific protein–nucleic acid interactions (Gursky et al. 1977).

The binding of AMD to DNA is known to require the proper conformation of DNA (B form) and to require the presence of a guanine 2-amino group faced into the minor groove of DNA (Gellert et al. 1965; Cerami et al. 1967). To enhance the binding specificity exhibited by Dst and AMD, one should increase the total number of specific reaction centers in the antibiotic molecules. An obvious way to accomplish this is to synthesize dimeric (oligomeric) compounds in which two or more monomers can interact specifically with DNA base pairs.

In this paper we describe DNA-binding properties of a series of bis-netropsins (bis-Nts) and bis-actinomycins, synthetic ligands of which each is composed of the two monomers linked by a flexible chain. We also demonstrate that certain oligopeptides with defined amino acid sequences can bind to DNA in a sequence-specific manner. At present it seems likely that the information determining the recognition of specific DNA sequences by regulatory proteins is inherent in the primary sequence of the polypeptide chain. Relatively short sequences of regulatory proteins are known to

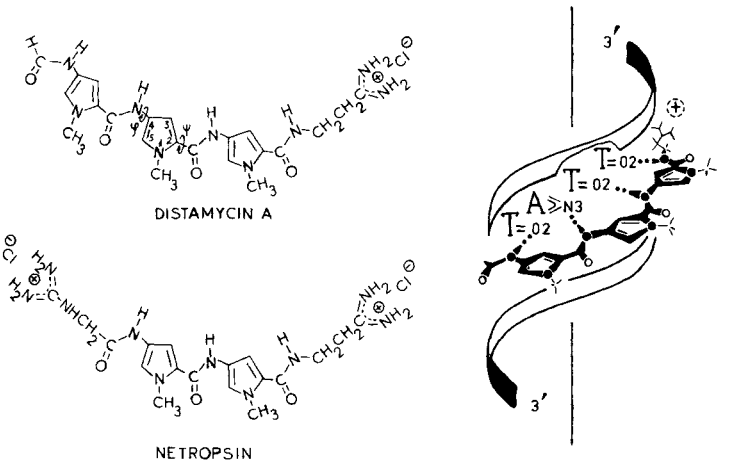


Figure 1. Chemical structure of Dst and Nt and schematic drawing illustrating the proposed model for distamycin-DNA complex. The amide groups of the antibiotic molecule are hydrogen bonded to the thymine C=O groups and adenine N3 atom lying in the same polynucleotide strand. The sequence TATT is shown as an attachment site for the antibiotic molecule. In the complex the antibiotic molecule possesses an overall curved shape with the imido-NH groups on the concave side of the molecule and the C=O groups on the convex side. Indicated are atom numbering and conformation angles ϕ and ψ. The conformation with $\phi = 140°$ and $\psi = 60°$ allows hydrogen bonding of the antibiotic amide groups to thymine O2 and adenine N3 atoms in the minor DNA groove (Gursky et al. 1977).

carry all the information necessary for recognition of the appropriate target sites on DNA (Adler et al. 1972; Müller-Hill 1975; Ogata and Gilbert 1979; Johnson et al. 1981).

A degenerated protein–nucleic acid recognition code proposed earlier (Gursky et al. 1975, 1976a, 1977, 1979) offers a possibility for synthesis of model oligopeptides with simplified sequences containing essential residues for the recognition of a specific DNA site. We describe here design and DNA-binding properties of a series of oligopeptides modeling the recognition site of the *lac* repressor.

Determination of Free-energy Changes for the Interaction of a Single Dst Amide Group with Four DNA Bases

To estimate the contributions of various chemical groups of Dst to the stability and specificity of Dst-DNA complexes, we have synthesized a series of Dst analogs (Fig. 2) and investigated their interaction with various natural and synthetic DNAs. These molecules contain various numbers of methylpyrrol carboxamide units and possess glycylamido or dansylglycylamido groups instead of a formamido group on the left end of the Dst molecule.

The binding constants $K = \lim_{r\to 0}(r/m)$ were determined from the intercepts of binding isotherms in Scatchard representation on the r/m axis and are summarized in Table 1. Here, r is the ratio of moles of bound ligand to moles of DNA base pairs, and m is the molar concentration of free ligand in the titration assays.

The free-energy change, ΔG_j, for binding of a ligand with n reaction centers to the jth site on a polynucleotide chain can be represented as a sum of contributions of individual reaction centers:

$$\Delta G_j = \Delta G_0 + \Delta G_{1,j} + \Delta G_{2,j+1} + \ldots \Delta G_{n,j+n-1} \quad (1)$$

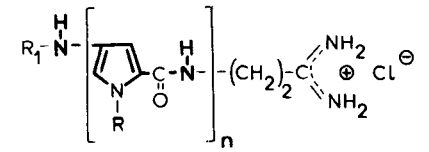
where ΔG_0 is the free-energy change corresponding to an unspecific component of the binding reaction. $\Delta G_{i,j+i-1}$ is the free-energy change accompanying the bonding of the ith ligand reaction center to a base at site $j+i-1$ on the polynucleotide chain ($1 \le i \le n$).

The intrinsic binding constant of the ligand to a stretch of n bases on the polynucleotide chain is given by the following standard relation:

$$K = \exp(-\Delta G_j/RT) = B \cdot \prod_{i=1}^{n} S_{j+i-1} \quad (2)$$

where $B = \exp(\Delta G_0/RT)$.

$S_{j+i-1} = \exp(-\Delta G_{i,j+i-1}/RT)$ is the stability constant for a bond connecting the ith ligand reaction center with the base at site $j+i-1$. Since all reaction centers are equivalent, each term $\Delta G_{i,j+i-1}$ in Equation 1 assumes one of the four possible values ΔG_T, ΔG_A, ΔG_C, ΔG_G, depending on the nature of the base at each site $j+i-1$. (ΔG_T, ΔG_A, ΔG_C, and ΔG_G denote the free-energy changes accompanying the bonding of each amide group of Dst to thymine, adenine, cytosine, and guanine bases, respectively. Similarly, S_T, S_A, S_C, and S_G denote the



H-Gly-Dst(Pr)-2 (n=2, R=$CH_2CH_2CH_3$, R_1= $H_3\overset{+}{N}$-CH_2CO-)

Dns-Gly-Dst(Pr)-2 (n=2, R=$CH_2CH_2CH_3$, R_1= Dns-Gly-)

Dns-Gly-Dst-1 (n=1, R=CH_3, R_1= Dns-Gly-)

Dns-Gly-Dst-2 (n=2, R=CH_3, R_1= Dns-Gly-)

Dns-Gly-Dst-3 (n=3, R=CH_3, R_1= Dns-Gly-)

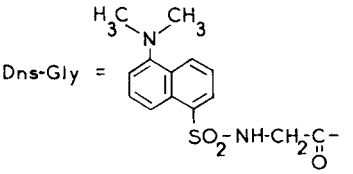


Figure 2. Chemical structure of Dst analogs.

corresponding bond-stability constants.) In the subsequent analysis of experimental data, we take into account that Dst can bind to DNA in the two opposite orientations by utilizing potential binding sites on the two polynucleotide strands. In the case of binding to a naturally occurring DNA, the intercept of the Scatchard isotherm on the vertical axis is equal to the average binding constant of a ligand to an isolated site on DNA (Zasedatelev et al. 1973; Livshitz et al. 1979):

$$K = \lim_{r\to 0}(r/m) = 2B[X_{AT}\cdot(S_T + S_A)/2 + (1-X_{AT})\cdot(S_C + S_G)/2]^n \quad (3)$$

In the limit, when $X_{AT} \to 1$,

$$K = 2BX_{AT}^n \cdot [(S_T + S_A)/2]^n \quad (4)$$

From Equations 3 and 4, one can conclude that the total number of ligand reaction centers can be determined from the slopes of experimental plots of $\ln K$ versus $\ln X_{AT}$ when $X_{AT} \to 1$. This procedure led us to the conclusion that the number of specific reaction centers in the molecules of Dst analogs is equal to the number of amide groups (Zasedatelev et al. 1974).

Once n is determined for every Dst analog, the binding constants to various synthetic DNAs can be readily calculated. Table 1 provides the necessary information for evaluation of stability constants S_T, S_A, S_C, and S_G from experimentally measured binding constants of Dst analogs to various DNAs. To accomplish this, we note that the unspecific interaction constant B takes the same value for a homologous series of Dst analogs (such as Dns-Gly-Dst-1, Dns-Gly-Dst-2, and Dns-Gly-Dst-3). (Dns is 5-dimethylamino naphthalene sulfonic acid.) We therefore used all data sets in Table 1 to calculate the experimental and theoretical values of f_i, which are de-

Table 1. Binding Constants of Dst Analogs to Various Natural and Synthetic DNAs

Polymer	X_{AT}	Compound	Binding constant K (M^{-1})	Theoretical expression for binding constant K
Poly(dA)•poly(dT)	1	Dns-Gly-Dst-1	$(6 \pm 1)\cdot 10^4$	$K^{DD1}_{A\bullet T} = B_{A\bullet T}(S_T^4 + S_A^2)$
		Dns-Gly-Dst-2	$(20 \pm 3)\cdot 10^5$	$K^{DD2}_{A\bullet T} = B_{A\bullet T}(S_T^3 + S_A^2)$
		Dns-Gly-Dst-3	$(5.5 \pm 1)\cdot 10^7$	$K^{DD3}_{A\bullet T} = B_{A\bullet T}(S_T^4 + S_A^4)$
		H-Gly-Dst(Pr)-2	$(11 \pm 1)\cdot 10^5$	$K^{D(Pr)2}_{A\bullet T} = B'_{A\bullet T}(S_T^3 + S_A^2)$
Poly(dA-dT)• poly(dA-dT)	1	Dns-Gly-Dst-2	$(18 \pm 3)\cdot 10^5$	$K^{DD2}_{A-T} = B_{A-T}S_TS_A(S_T + S_A)$
		Dns-Gly-Dst-3	$(18 \pm 3)\cdot 10^6$	$K^{DD3}_{A-T} = 2B_{A-T}(S_TS_A)^2$
		H-Gly-Dst(Pr)-2	$(7.2 \pm 1)\cdot 10^5$	$K^{D(Pr)2}_{A-T} = B'_{A-T}S_TS_A(S_T + S_A)$
Poly(dG)•poly(dC)	0	Dns-Gly-Dst-2	$(9 \pm 2)\cdot 10^3$	$K^{DD2}_{G\bullet C} = B_{G\bullet C}(S_C^3 + S_G^3)$
		Dns-Gly-Dst-3	$(5 \pm 3)\cdot 10^4$	$K^{DD3}_{G\bullet C} = B_{G\bullet C}(S_C^4 + S_G^4)$
		H-Gly-Dst(Pr)-2	$(25 \pm 5)\cdot 10^2$	$K^{D(Pr)2}_{G\bullet C} = B'_{G\bullet C}(S_C^3 + S_G^3)$
Poly(dG-dC)• poly(dG-dC)	0	Dns-Gly-Dst-2	$0 + 10^3$	$K^{DD2}_{G-C} = B_{G-C}S_CS_G(S_C + S_G)$
		H-Gly-Dst(Pr)-2	$0 + 10^3$	$K^{D(Pr)2}_{G-C} = B'_{G-C}S_CS_G(S_C + S_G)$
Poly(dA-dC)• poly(dT-dG)	0.5	Dns-Gly-Dst-2	$(10 \pm 5)\cdot 10^3$	$K^{DD2}_{(A\text{-}C)(T\text{-}G)} = B_{(A\text{-}C)(T\text{-}G)}[S_AS_C(S_A + S_C) + S_TS_G(S_T + S_G)]/2$
		Dns-Gly-Dst-3	$(2 \pm 1)\cdot 10^4$	$K^{DD3}_{(A\text{-}C)(T\text{-}G)} = B_{(A\text{-}C)(T\text{-}G)}(S_A^3S_C^3 + S_T^3S_G^3)$
Closteridium perfringens DNA	0.72	Dns-Gly-Dst-2	$(7 \pm 1)\cdot 10^5$	$K^{DD2}_{DNA} = 2B_{DNA}[X_{AT}(S_T + S_A)/2 + (1 - X_{AT})(S_C + S_G)/2]^3$
Calf thymus DNA	0.58		$(4 \pm 1)\cdot 10^5$	
Escherichia coli DNA	0.49		$(25 \pm 5)\cdot 10^4$	
Achromobacter agile DNA	0.38		$(11 \pm 5)\cdot 10^4$	
Closteridium perfringens DNA	0.72	H-Gly-Dst(Pr)-2	$(45 \pm 5)\cdot 10^4$	$K^{D(Pr)\text{-}2}_{DNA} = 2B'_{DNA}[X_{AT}(S_T + S_A)/2 + (1 - X_{AT})(S_C + S_G)/2]^3$
Calf thymus DNA	0.58		$(24 \pm 4)\cdot 10^4$	
Escherichia coli DNA	0.49		$(16 \pm 2)\cdot 10^4$	
Achromobacter agile DNA	0.38		$(10 \pm 1)\cdot 10^4$	
Micrococcus lysodeikticus DNA	0.28		$(5.3 \pm 1)\cdot 10^4$	

fined as the ratio of the appropriate binding constants determined from the binding of two homologous analogs to the same synthetic DNA (Table 2). This procedure allows us to eliminate Bs and provides us with experimental values of f_i for seven simultaneous equations, each containing the appropriate stability constants. To evaluate the parameters S_T, S_A, S_C, and S_G, we applied an iterative procedure in which the sum

$$\sum_{i=1}^{7} (f_i^e - f_i^c)/\Delta f_i^e \qquad (5)$$

is minimized. Here, f_i^e is the experimental value of f_i; Δf_i^e is the experimental uncertainty in f_i^e. f_i^c is the theoretically calculated value of f_i. Table 3 lists the estimated values for S_T, S_A, S_C, and S_G and the corresponding values for the interaction energies ΔG_T, ΔG_A, ΔG_C, and ΔG_G.

Table 2. Quantities Independent of the Magnitude of the Unspecific Component of Binding Free Energy

i	f_i	f_i^e	Δf_i^e	$(f_i^e - f_i^c)/\Delta f_i^e$
1	$f_1 = K^{DD2}_{A\bullet T}/K^{DD1}_{A\bullet T} = (S_T^3 + S_A^3)/(S_T^4 + S_A^2)$	33	9	0.21
2	$f_2 = K^{DD3}_{A\bullet T}/K^{DD2}_{A\bullet T} = (S_T^4 + S_A^4)/(S_T^3 + S_A^3)$	27.5	10	−0.43
3	$f_3 = K^{DD3}_{A\text{-}T}/K^{DD2}_{A\text{-}T} = 2S_TS_A/(S_T + S_A)$	10	3	0.01
4	$f_4 = K^{DD3}_{G\bullet C}/K^{DD2}_{G\bullet C} = (S_C^4 + S_G^4)/(S_C^3 + S_G^3)$	5.5	3	0.43
5	$f_5 = (\lim_{X_{AT}\to 0} \sqrt[3]{K^{DD2}_{DNA}})/(\lim_{X_{AT}\to 1} \sqrt[3]{K^{DD2}_{DNA}})$ $= (\lim_{X_{AT}\to 0} \sqrt[3]{K^{D(Pr)2}_{DNA}})/(\lim_{X_{AT}\to 1} \sqrt[3]{K^{D(Pr)2}_{DNA}})$ $= (S_C + S_G)/(S_T + S_A)$	0.115	0.02	−0.05
6	$f_6 = K^{DD3}_{(A\text{-}C)(T\text{-}G)}/K^{DD2}_{(A\text{-}C)(T\text{-}G)}$ $= 2(S_A^3S_C^3 + S_T^3S_G^3)/[S_AS_C(S_A + S_C) + S_TS_G(S_T + S_G)]$	2	4	−0.22
7	$f_7 = K^{DD2}_{G\text{-}C}/B_{G\text{-}C} = S_CS_G(S_C + S_G)$	0	10^2	−0.12

The experimental data for the calculation of f_i^e and Δf_i^e are taken from Table 1.

Table 3. Stability Constants and Free-energy Changes for the Interaction of a Single Amide Group of Dst with Thymine, Adenine, Cytosine, and Guanine Bases

Base	Stability constant	Free-energy change (kcal/M)
Thymine	$S_T = 32 \pm 6$	$\Delta G_T = -2.1 \pm 0.1$
Adenine	$S_A = 5.9 \pm 2$	$\Delta G_A = -1.1 \pm 0.2$
Cytosine	$S_C = 4.2 \pm 1.5$	$\Delta G_C = -0.9 \pm 0.2$
Guanine	$S_G = 0.2^{+1}_{-0.2}$	$\Delta G_G = +0.9^{-1}_{+\infty}$

These results deserve some comment. The observed differences in the magnitude of the interaction energies listed in Table 3 appear to reflect the order of the reactivities of purine N3 and pyrimidine O2 atoms with respect to their interaction with amide groups of Dst. The interaction energies are decreased in the order thymine O2 > adenine N3 > cytosine O2 ⋙ guanine N3. The guanine actually acts as a repulsive element. From examination of molecular models, one can conclude that the guanine 2-amino group causes a sterical hindrance for hydrogen bond formation between the antibiotic imido group and the guanine N3 atom, thus preventing the attachment of the antibiotic molecule to a section of the polynucleotide chain that contains guanine (Zasedatelev et al. 1976; Gursky et al. 1977). The decrease in the reactivity of the O2 position of cytosine, compared with that of the thymine O2 atom, results probably from the fact that the cytosine O2 atom is implicated in hydrogen bond formation with the guanine 2-amino group in a GC pair. The replacement of guanine by inosine unshields the N3 position of the purine and makes more reactive the O2 position of cytosine in an IC pair, thereby increasing the strength of the antibiotic binding to poly(dI)·poly(dC) (Krey et al. 1973; Zasedatelev et al. 1976).

Having resolved the four free energies that define the interaction between a single amide group of Dst and DNA bases, it is possible to evaluate unspecific interaction energies ΔG_0 using the data in Table 1. Table 4 shows that the estimated ΔG_0 values are not identical for binding to various synthetic DNAs but vary around the average value of about 2.6 kcal/mole. This may indicate that unspecific binding of Dst analogs is to some extent sensitive to structural differences that are known to exist between various synthetic DNAs (Arnott and Chandrasekaran 1981). However, the variation of ΔG_0 with DNA sequence within about 1 kcal/mole does not markedly affect the antibiotic-binding specificity, which is basically governed by hydrogen-bonding interactions between the four Dst amide groups and DNA bases (see Table 3).

Design and DNA-binding Properties of Bis-Nts

The binding specificity shown by Nt and Dst can be improved by their di- and oligomerization. Figure 3 shows the chemical formulas of the bis-Nts studied here. Each molecule consists of the two Nt-like fragments linked by a flexible chain. In principle, the two Nt-like fragments can be aligned in either a parallel or antiparallel manner. The latter class can further be subdivided into two groups corresponding to head-to-head and tail-to-tail associations of the two fragments. These molecules will be referred to as bis-Nt(←→) and bis-Nt(→←), respectively, with the notation bis-Nt(→→) for the molecule in which the two fragments are aligned in the parallel fashion. A detailed account of the synthesis of these compounds will be published elsewhere.

The CD spectra of bis-Nt(→←) and bis-Nt(→→) complexed with poly(dA)·poly(dT) are very similar to those reported for binding of bis-Nt(←→) (Khorlin et al. 1980). The molar ellipticities of all these complexes are found to be twice as large as that of the complex with the corresponding monomeric fragment. Figure 4 shows typical CD titration curves obtained for binding of bis-Nt(←→), bis-Nt(→←), and bis-Nt(→→) to poly(dA)·poly(dT) and calf thymus DNA. From measured saturation levels of the binding of these ligands to poly(dA)·poly(dT), the conclusion can be drawn that each bound bis-Nt molecule occupies 10–11 bp. In contrast, the estimated site size for binding of the monomeric fragment is equal to 5 bp. The saturation levels of binding of bis-Nts to calf thymus DNA are markedly lower than that of Nt, reflecting a greater binding specificity of bis-Nts in comparison with the monomeric fragment. The ratio of the measured dichroisms for saturated complexes of each ligand with poly(dA)·poly(dT) and calf thymus DNA can be considered as a measure of binding specificity. According to this criterion, the binding specificity is decreased in the following order: bis-Nt(→→) > bis-Nt(←→) ≅ bis-Nt(→←) > Nt.

The binding principles found for Dst and Nt suggest

Table 4. Stability Constants and Free-energy Changes Characterizing Unspecific Interactions between Dst Analogs and Various Nucleic Acids

Polymer	Compound	Stability constant B	Free-energy change ΔG_0 (kcal/M)
Poly(dA)·poly(dT)	Dns-Gly-Dst-n	$B^{DDn}_{A\cdot T} = 57 \pm 4$	-2.43 ± 0.05
Poly(dA-dT)·poly(dA-dT)	Dns-Gly-Dst-n	$B^{DDn}_{A\text{-}T} = 250 \pm 20$	-3.31 ± 0.05
Poly(dG)·poly(dC)	Dns-Gly-Dst-n	$B^{DDn}_{G\cdot C} = 120 \pm 10$	-2.89 ± 0.07
Poly(dA-dC)·poly(dT-dG)	Dns-Gly-Dst-n	$B^{DDn}_{(A\text{-}C)(T\text{-}G)} = 35 \pm 8$	-2.1 ± 0.2
Natural DNAs	Dns-Gly-Dst-n	$B^{DDn}_{DNA} = 110 \pm 10$	-2.83 ± 0.06

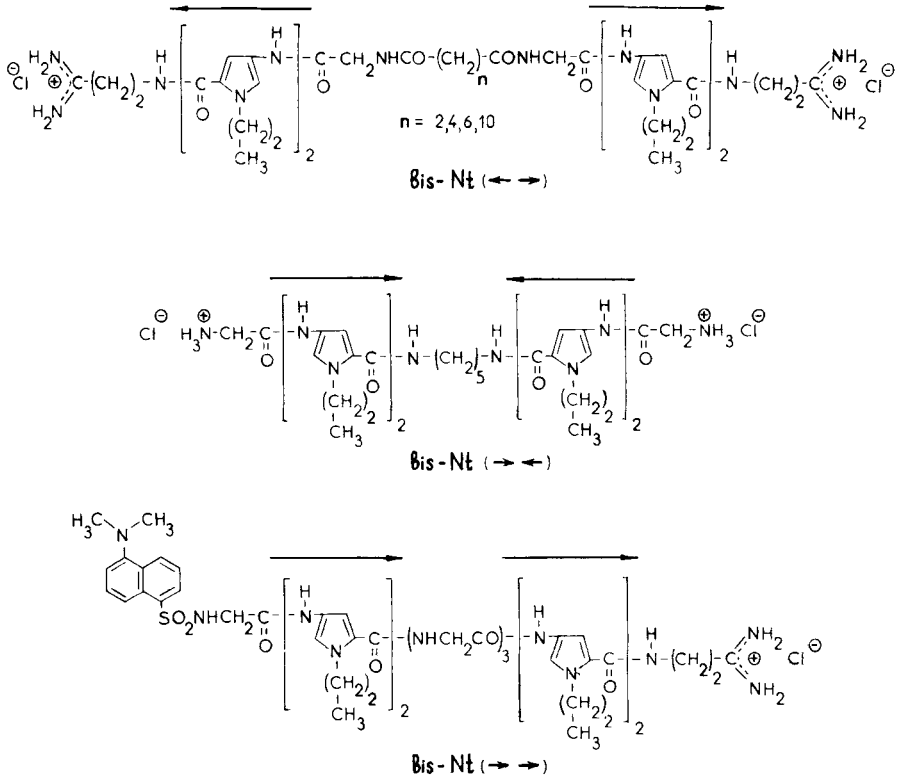


Figure 3. Chemical formulas of bis-Nt with antiparallel and parallel alignments of two Nt-like fragments.

that the two Nt-like fragments of bis-Nt(→ →) appear to interact preferentially with two stretches of thymine residues lying in the same polynucleotide strand. In contrast, bis-Nt(← →) and bis-Nt(→ ←) appear to bind preferentially to two stretches of thymine bases lying in the opposite polynucleotide strands. In the complex, the two Nt-like fragments are probably related by twofold rotation symmetry. The distance between the two Nt-like fragments can be adjusted by changing the length of the connecting chain (Krylov et al. 1980). Each Nt-like fragment can interact with any base sequence containing no guanine.

As expected for the molecules with six specific reaction centers, bis-Nt(← →), bis-Nt(→ ←), and bis-Nt(→ →) bind very tightly to DNA with the intrinsic binding constant greater than 10^8 M^{-1}. If the two Nt-like fragments of bis-Nt act independently upon binding, the estimated value of the binding constant of bis-Nt(→ →) to poly(dA)·poly(dT) is on the order of 10^{12} M^{-1}.

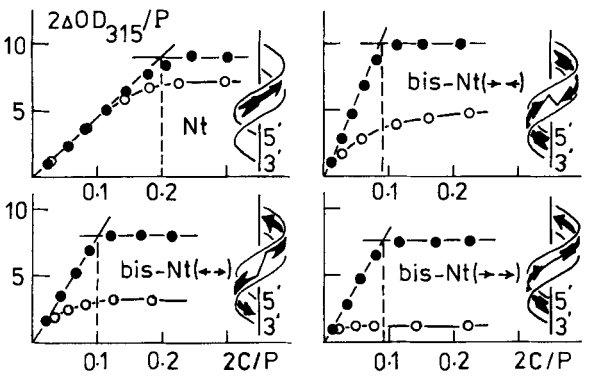


Figure 4. CD titrations of poly(dA)·poly(dT) (●) and calf thymus DNA (○) with the monomeric Nt-like fragment (H-Gly-Dst[Pr]-2) and various bis-Nts. Indicated is the proposed geometry of complexes between ligand and DNA. Concentrations of nucleic acids (P/2) were 5 × 10^{-5} M (base pairs). C is the total molar concentration of ligand. $2\Delta OD_{315}/P$ is the measured dichroism expressed per mole of base pairs and per 1 cm path length. Experiments were made in 0.06 M phosphate buffer (pH 6.0) at 20°C.

Bis-Nts as Selective Inhibitors of DNA-dependent RNA Synthesis

Retchinsky et al. (1981b) have previously showed that bis-Nt(← →) selectively inhibited the in vitro RNA synthesis from certain prokaryotic promoters. Therefore, we have begun a series of studies to test inhibition effects of bis-Nts in highly purified systems that use as templates short, sequenced DNA fragments, each containing a well-characterized promoter such as the UV5 *lac* promoter (Dickson et al. 1975) or early phage T7 promoters A2 and A3 (Dunn and Studier 1981). The isolated DNA fragments containing these promoter sequences were inserted in a modified pBR322 plasmid instead of the original tetracycline promoter. The cloning and the properties of some of the plasmids were described elsewhere (Retchinsky et al. 1981a). The cloned promoters together with the sequences originating in pBR322 were cut out with *Hae*III restriction nuclease. The resulting fragments of about 250 bp were isolated via polyacrylamide gel electrophoresis and purified by

DEAE-cellulose chromatography and gel filtration. Each fragment contains (1) the 18-bp *Hae*III-*Eco*RI restriction fragment situated upstream from a promoter-containing fragment and carrying the nucleotides present in positions 4344–4362 of the pBR322 DNA (Sutcliffe 1979), (2) the 144-bp *Hin*dIII-*Hae*III fragment situated downstream from the promoter and containing the nucleotides 31–175 of the pBR322 DNA, and (3) a promoter-containing fragment inserted between the above-mentioned pBR322 DNA sequences. The DNA fragments carrying promoter sequences were (1) the *Hpa*II-*Hpa*II fragment located between nucleotides 547 and 651 of T7 DNA and containing the A2 promoter, (2) the *Hpa*II-*Hin*fI fragment located between nucleotides 651 and 761 of T7 DNA and containing the A3 promoter (Dunn and Studier 1981), and (4) the UV5 *lac* promoter sequence ranging from nucleotide −58 to nucleotide +5 of the *lac* control region (Dickson et al. 1975). The last sequence will be referred to as the modified UV5 *lac** promoter, with the abbreviation UV5 *lac* for the entire *lac* control region ranging from nucleotide −58 to nucleotide +37 (Dickson et al. 1975).

Figure 5 shows the effects of bis-Nts on the RNA synthesis directed by various promoters. Experiments were done as described by Retchinsky et al. (1981b) and outlined in the figure legend. From Figure 5, one can conclude that all bis-Nts selectively inhibit transcription from various promoters. As shown in Figure 5a, the inhibition effect of bis-Nt(←→) is decreased in the order UV5 *lac* ≅ UV5 *lac** > A2 > A3 > A1.

To estimate the potential contribution of the inhibition of RNA chain elongation, the ternary complexes between RNA polymerase, whole D111 T7 DNA, and nascent, unlabeled RNA were prepared and treated for 10 minutes with various concentrations of inhibitor. RNA synthesis was then restarted by adding $MgCl_2$ and [α-^{32}P]CTP. The maximal inhibition level seen under these conditions was about 30% in the presence of 1.2×10^{-2} M of either bis-Nt(←→) or Nt (Fig. 5a). This suggests that bis-Nt(←→) preferentially inhibits initiation of transcription. It is worth noting that in the restriction fragments carrying the UV5 *lac** promoter and promoters A2 or A3, the transcribed regions are nearly identical. Since bis-Nt(←→) binds to DNA in a sequence-specific manner, the observed difference in the inhibition levels of various promoters appears to be attributed to specific interaction between bis-Nt(←→) and promoter. Analysis of the distribution of putative binding sites for bis-Nt(←→) on various promoter sequences suggests that the inhibiting effect can be correlated with the strength of bis-Nt(←→) binding to putative sites located in a small DNA region near the transcription start site. Bis-Nt(←→) appears to bind to two stretches of nonguanines lying in the opposite polynucleotide strands and separated by two arbitrary base pairs (Krylov et al. 1980). Such sequences exist in the UV5 *lac* promoter and the T7 A2 promoter near the start site of transcription. Our previous experiments show that bis-Nt(←→) at a concentration of approximately 5×10^{-6} M inhibits the formation of a binary complex between RNA polymerase and the UV5 *lac* promoter if added to the reaction mixture prior to the addition of the enzyme (Retchinsky et al. 1981b).

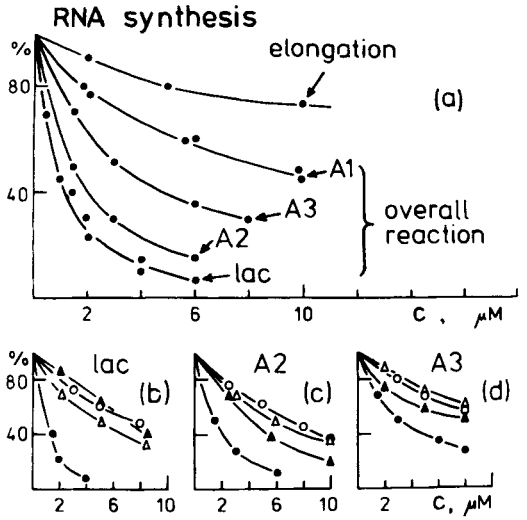


Figure 5. Effect of bis-Nts on the incorporation of [α-^{32}P]CMP into RNA at various concentrations (C) of bis-Nts. CMP incorporation in the absence of ligand was taken as 100%. (*a*) Effect of bis-Nt(←→) on the RNA synthesis directed by T7 bacteriophage promoters (A1, A2, and A3) and promoters UV5 *lac* and UV5 *lac**. The inhibition levels of the last two promoters were coincident (the lower curve). The upper curve shows unspecific inhibition levels of RNA-chain-elongation reaction at various concentrations of bis-Nt(←→). Data for the A1 promoter were obtained using whole D111 T7 DNA as a template (Retchinsky et al. 1981b). (*b, c, d*) Effect of various bis-Nts on the RNA synthesis directed by the UV5 *lac** promoter, A2 promoter, and A3 promoter, respectively. (●) Bis-Nt(←→) with 10 methylene links in the connecting chain (see Fig. 3); (▲) bis-Nt(→→); (△) bis-Nt(→←); (○) Nt-like fragment (H-Gly-Dst[Pr]-2). Experiments were made as reported by Retchinsky et al. (1981b). The concentration of promoter in the reaction mixture was 40 nM for each restriction fragment and 2 nM for whole D111 T7 DNA. Experiments were made at 37°C in 50 mM Tris-HCl (pH 7.5) buffer containing 75 mM NaCl; 3 mM $MgCl_2$; 1 mM β-mercaptoethanol; 10 mg/liter of bovine serum albumin; 100 μM ATP, GTP, and UTP; and 10 μM [α-^{32}P]CTP. Bis-Nt and DNA restriction fragment (or whole D111 T7 DNA) were preincubated for 10 min in the complete reaction mixture without RNA polymerase. The transcription was initiated by adding RNA polymerase holoenzyme up to the concentration of 40 nM and terminated by adding EDTA.

Figure 5, b, c, and d, shows variations in the inhibition effects produced by various bis-Nts on the transcription from the same promoters. The observed effects for bis-Nt(→←) and bis-Nt(→→) are of smaller magnitudes than that found for bis-Nt(←→). In some cases, inhibition effects are comparable to the activity of the monomeric Nt analog. The selective influence of each ligand on the RNA synthesis directed by various promoters is apparent from the data presented in Figure 5.

Bis-Actinomycin D

Bis-actinomycins D (bis-AMD-1 and bis-AMD-2) are synthetic sequence-specific ligands (Fig. 6) in which

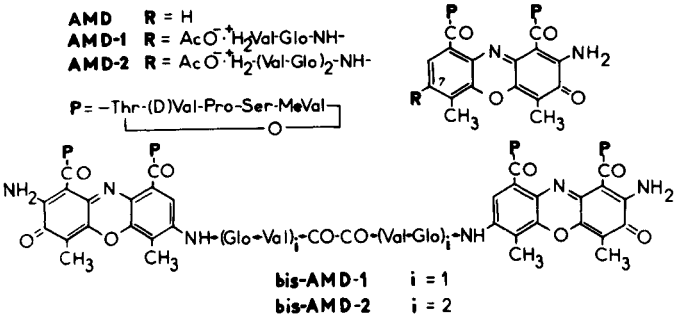


Figure 6. Chemical formulas of AMD, AMD analogs (AMD-1 and AMD-2), and dimeric compounds bis-AMD-1 and bis-AMD-2. Indicated are $C' \rightarrow C^{\alpha} \rightarrow N$ directions in the two halves of the connecting chain of bis-AMD. Glo is the glycolic acid residue.

two actinomycinlike fragments are linked covalently by a flexible chain attached to position 7 of the antibiotic chromophore (Mikhailov et al. 1981; Nikitin et al. 1981). Both bis-AMD-1 and bis-AMD-2 display a complex pattern of sequence selectivity with considerable specificity for GC pairs. Physicochemical studies on the binding of bis-AMD-1 and bis-AMD-2 to various natural and synthetic DNAs show that the two AMD-like fragments of bis-AMD bind to DNA in virtually the same manner as AMD does (Mikhailov et al. 1981). Each AMD fragment carries one GC-specific reaction center. In the complex, the two AMD-like fragments of bis-AMD-1 are probably related by twofold symmetry and interact specifically with two guanine bases lying in the opposite polynucleotide strands and separated by 4 bp.

Figure 7 shows typical titration curves obtained for binding of bis-AMD-1 and the corresponding monomer analog to poly(dG-dC)•poly(dG-dC). From measured saturation levels of binding, we conclude that the site size for strong binding of bis-AMD-1 (9 bp) is about two times greater than that of AMD. Synthesis of bis-AMD seems to be important for further progress in designing sequence-specific ligands.

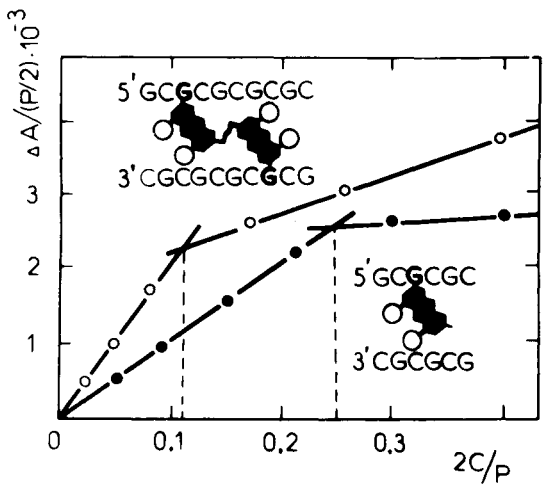


Figure 7. Titration of poly(dG-dC)•poly(dG-dC) with AMD-1 (●) and bis-AMD-1 (○). ΔA is the difference in the absorbance at 425 nm of ligand solutions in the absence and presence of poly(dG-dC)•poly(dG-dC). Concentration of poly(dG-dC)•poly(dG-dC) (P/2) was 2×10^{-4} M (base pairs). C is the total molar concentration of the ligand in the titration assay.

Design and DNA-binding Properties of Oligopeptides Modeling the Recognition Site of *lac* Repressor

It is known that the 59 aminoterminal residues of the *lac* repressor carry all (or almost all) of the information needed for the recognition of the specific *lac* operator sequence on DNA (Müller-Hill 1975; Ogata and Gilbert 1979). Here, we describe our attempts to define which residues of the *lac* repressor polypeptide chain are essential for the recognition of the operator sequence. We have synthesized a series of oligopeptides modeling the aminoterminal sequence of the *lac* repressor. The ideas of synthesis stem from the stereochemical model that has been advanced for the binding of the *lac* repressor to the *lac* operator (Gursky et al. 1975, 1976b). According to the model, the aminoterminal residues 19–32 and 53–71 of the *lac* repressor are implicated in specific interaction with operator DNA. A correspondence has been found between the *lac* repressor sequence 19–30 and the nucleotide sequence of the *lac* operator (Fig. 8). This correspondence points toward the existence of a degenerated protein–nucleic acid recognition code in which threonine, serine, asparagine, and histidine residues (potential hydrogen bond donors) serve as AT-coding residues, whereas inert valyl and alanyl residues serve as GC-coding residues (Gursky et al. 1976a, 1977). This type of correlation, in a generalized form, has also been found between the aminoterminal sequences of λcI and *cro* repressors and corresponding operator sequences (Gursky et al. 1979).

To check these rules, we have synthesized a series of oligopeptides modeling the sequence 19–31 of the *lac* operator (Fig. 8) and investigated their binding to natural and synthetic DNAs. As shown in Figure 8, in the model tripeptide (TVP), hexapeptide (HEXP), tridecapeptide (TDP), and eicosapeptide (ECP), the valyl and threonyl residues are used as potential GC- and AT-coding residues, respectively. The sequence of ECP involves a TDP fragment linked to the peptide Lys-Gly-Ala-Gln-Gln-Leu-Ala which contains residues 53 to 57 of the *lac* repressor. These residues are probably implicated in the formation of an intersubunit β sheet in the repressor tetramer (Gursky et al. 1976b).

The monodisperse, chemically and optically pure oligopeptides were synthesized according to the conventional stepwise elongation procedures in solution with

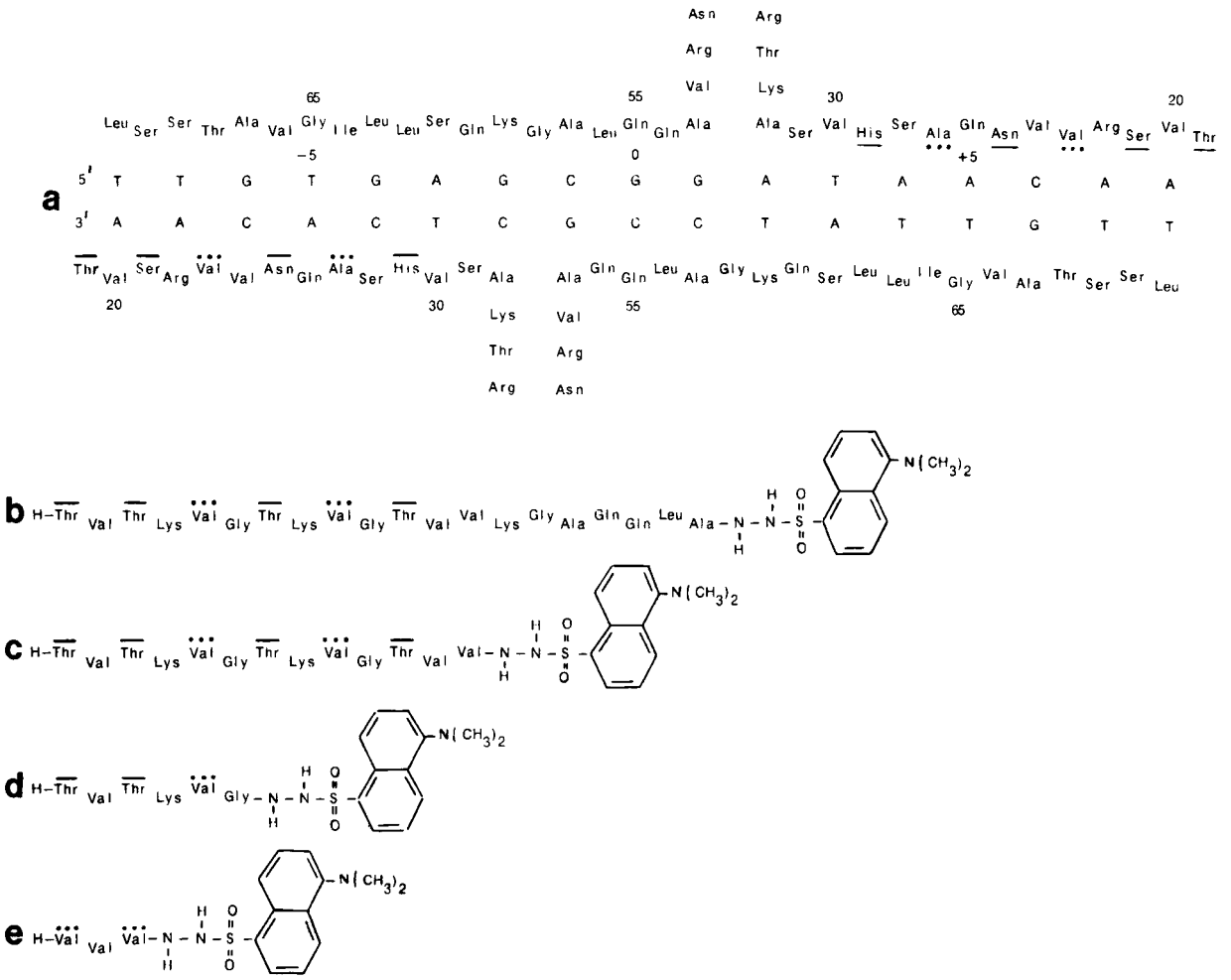


Figure 8. Design and chemical formulas of oligopeptides modeling the recognition site of *lac* repressor. (*a*) A model for the binding of *lac* repressor to the *lac* operator (Gursky et al. 1976b). The repressor polypeptide chain sequences 19–32 and 53–71 are involved in a specific interaction with operator DNA. A pair of twofold-related repressor subunits interact with symmetrically disposed base-pair sequences on the *lac* operator. There is a correlation between the sequence 19–30 of *lac* repressor subunit and base-pair sequences lying on the peripheral parts of the *lac* operator. Amino acid residues coding for A-T and G-C pairs are underlined by solid and dotted lines, respectively. The sequences 53–57 in the two repressor subunits are implicated in the formation of an intersubunit β sheet that probably interacts with the central part of the *lac* operator sequence. (*b–e*) Chemical formulas of oligopeptides modeling the recognition site of the *lac* repressor. Each oligopeptide carries a fluorescent dansyl label.

modifications that will be reported in detail elsewhere. All of the oligopeptides used have a limited solubility in aqueous solution and tend to form aggregates of various sizes that are in concentration-dependent equilibrium with each other and monomers. The existence of associated and nonassociated oligopeptide forms in aqueous solution is strongly supported by our observations showing that the absorbance and fluorescence spectra of the oligopeptides in free solution are concentration-dependent (Streltsov et al. 1980). All of the oligopeptides bind to DNA. Figure 9 shows the absorbance and fluorescence spectra of TDP and HEXP in the presence and absence of DNA. The spectral changes observed on adding DNA to TDP (or ECP) are similar to those reported for binding of TVP to DNA (Streltsov et al. 1980). The fluorescence spectrum of TDP (or ECP) shows a shift to a shorter wavelength on binding to DNA, which is attributable to the hydrophobic environment of the Dns group in the complex. The fluorescence intensity per mole of TDP (or ECP) is greatly enhanced on binding to DNA. Similar enhancement is observed for HEXP at a concentration 100 times higher.

The addition of DNA to diluted solutions of HEXP and TDP induces only minor changes in the absorbance and fluorescence spectra, even at a DNA/oligopeptide ratio of 100:1. There is a small increase in the fluorescence intensity at 550 nm. These spectral changes can be attributed to the binding of the monomeric oligopeptide species to DNA (Streltsov et al. 1980).

The CD spectra for TDP in the presence and absence of DNA are shown in Figure 9c. In the peptide absorption region, the CD pattern of free oligopeptide shows a negative band at 197 nm. This indicates that a significant portion of the oligopeptide molecules is in a random-coiled conformation. On adding DNA to TDP,

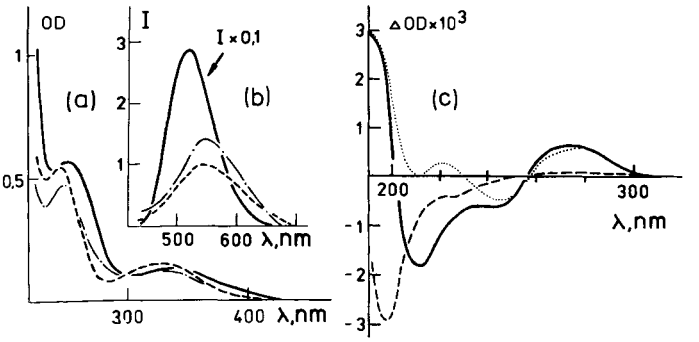


Figure 9. Changes observed in the absorption, fluorescence, and CD spectra of TDP and HEXP upon binding to calf thymus DNA. (*a*) Absorption spectrum of free TDP (----) and difference absorption spectra obtained by subtracting the spectrum of DNA from the spectra of complexes of TDP (——) and HEXP (—•—•—) with DNA. (*b*) Fluorescence emission spectra (uncorrected) of free TDP (----), TDP (——), and HEXP (—•—•—) in the presence of calf thymus DNA. The fluorescence and absorption spectra of free HEXP were practically coincident with those of TDP at the same concentration. Excitation wavelength was 380 nm. (*c*) CD spectra of free TDP (----), of TDP in the presence of calf thymus DNA (——), and of DNA alone (••••••). (ΔOD) Measured dichroism calculated per 1 cm path length. The complexes between oligopeptides and DNA were prepared as reported by Streltsov et al. (1980). Concentrations of TDP and HEXP were 5×10^{-5} M. DNA concentration was 1×10^{-4} (base pairs). Experiments were made at 20°C in 0.001 M Na-cacodylate buffer (pH 7.0) in the presence of 10% (v/v) methanol.

the CD spectrum changes completely. There is a great increase in the CD amplitude at 190 nm. In addition, the two negative CD bands appear at 213 nm and 255 nm. The CD band around 255 nm is likely to be related to a contribution of the aromatic Dns group. There is a remarkable similarity of CD spectra obtained for DNA complexes with TDP and TVP (Streltsov et al. 1980). The β structure is known to be the preferred conformation for the N-protected oligo(L-valine)$_7$ in trifluoroethanol, whereas oligo(L-valine)$_3$ assumes an unordered conformation (Toniolo and Bonora 1974). In aqueous solution the equilibrium between associated and nonassociated species is shifted toward the formation of associated species (Toniolo and Bonora 1974; Streltsov et al. 1980). An increase in the CD amplitude at 190 nm observed for TDP-DNA complexes may be partially attributed to the formation of β-associated TDP species upon binding to DNA.

Figure 10 shows typical binding isotherms of TVP to poly(dG)•poly(dC) and poly(dA)•poly(dT). The binding isotherms were determined for two different DNAs simultaneously in three-part dialysis cells similar to those described by Müller and Crothers (1975). From Figure 10, one can conclude that at low concentrations of TVP in free solution, the binding is negligible. Scatchard plots show that the oligopeptide in the monomeric form binds weakly ($K \cong 10^3$ M^{-1}) and unspecifically to DNA. Increasing the free oligopeptide concentration from 6×10^{-5} M to 1.5×10^{-4} M results in a remarkable increase in the apparent binding constant. In this region of binding isotherm, TVP binds more strongly to poly(dG)•poly(dC) than to poly(dA)•poly(dT). It seems likely that β-associated oligopeptide species are implicated in specific interactions with DNA base pairs.

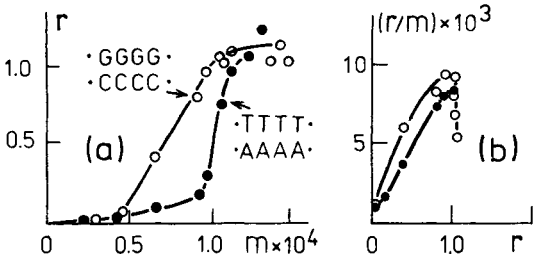


Figure 10. Binding isotherms of TVP to poly(dG)•poly(dC) (○) and poly(dA)•poly(dT) (●) as determined by equilibrium dialysis. (*a*) Plots of *r* (the ratio of moles of bound TVP to moles of DNA base pairs) against *m* (the molar concentration of free oligopeptide). (*b*) Scatchard plots of the same data. Concentrations of nucleic acids were 8×10^{-5} M (base pairs). Conditions: 0.001 M Na cacodylate (pH 7.0), 5% (v/v) methanol, 20% (v/v) 2,2,2-trifluoroethanol.

The binding is a cooperative phenomenon. As soon as a certain critical concentration is exceeded, DNA binds a significant amount of the oligopeptide in a cooperative manner. In an aqueous solution, β sheets are known to have a tendency to form β sandwiches stabilized by hydrophobic interactions (Cohen et al. 1981). It seems likely that sequence-specific binding of β-associated oligopeptide species in one of the DNA grooves favors subsequent β strand–β sheet–β sandwich condensation. In agreement with our binding experiments, this model predicts that in a saturated complex, the two TVP dimers forming a sandwich occupy a DNA region extending over 3 bp.

The binding takes place in the minor DNA groove, as revealed from our observations that both TDP and TVP bind tightly to phage T2 DNA, in which the major groove is occupied by massive glucose and diglucose residues. In addition, distamycin A, which binds to AT base pairs in the minor groove, can displace TVP and TDP from DNA (Fig. 11).

A combination of equilibrium dialysis and fluorescence measurements shows that the fluorescence intensity at 520 nm is very sensitive to the formation of complexes between β-associated oligopeptide species and DNA. In the excess of DNA over TVP, the fluorescence intensity is proportional to the square of the oligopeptide concentration, thereby suggesting that the dimeric oligopeptide species are implicated in specific interaction with DNA (Streltsov et al. 1980). The fluorescence intensity of TVP-DNA complexes with higher *r* values

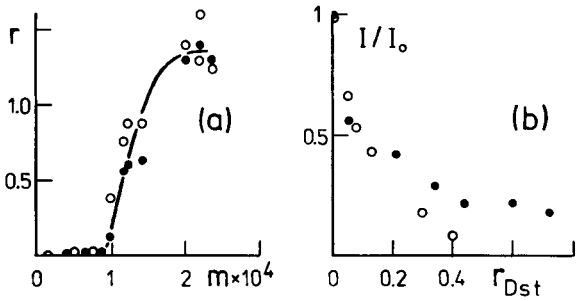


Figure 11. Data showing that TVP and TDP bind to DNA in the minor groove. (*a*) Plots of r against m for the binding of TVP to calf thymus DNA (○) and glycosylated bacteriophage T2 DNA (●). The binding isotherms were obtained by equilibrium dialysis in three-part dialysis cells. Concentration of nucleic acids was 8×10^{-5} M (base pairs). Conditions were the same as those in Fig. 10. (*b*) Displacement of TVP (○) and TDP (●) from calf thymus DNA by Dst. The normalized fluorescence intensity I/I_0 of oligopeptide-DNA complexes is plotted against r_{Dst} (the ratio of moles of bound Dst to moles of DNA base pairs). I_o is the fluorescence intensity of the oligopeptide-DNA complexes in the absence of Dst. Concentrations of TVP and TDP were 4×10^{-4} M and 2.5×10^{-5} M, respectively. Concentration of DNA was 1×10^{-4} M (base pairs). Conditions were the same as those in Fig. 9.

$(0.1 < r \leq 1)$ exceeds the intensity of the fluorescence of the free oligopeptide by approximately two orders of magnitude. The contribution of bound monomers to the fluorescence intensity can also be neglected in the region $0.1 < r \leq 1$. Figure 12 shows that for complexes with high r values, there is a linear correlation between the fluorescence intensity and the concentration of bound oligopeptide, as determined from equilibrium dialysis measurements. From Figure 12a, one can also conclude that the fluorescence intensities per mole of oligopeptide bound are equal for the complexes with poly(dA)•poly(dT) and poly(dG)•poly(dC). These experiments demonstrate that the fluorescence intensity at 520 nm can be regarded as a measure of the amount of oligopeptide bound to DNA in the associated form. Since the fluorescence quantum yield of the bound Dns group does not depend on the G+C content of DNA, the fluorescence measurements provide a rather simple and sensitive tool for comparing the affinities of TVP and TDP to various DNAs. Figure 10 shows that there is a region of binding isotherm where the binding process is sequence-specific.

Figure 12 displays the plots of fluorescence intensity at 520 nm versus nucleic acid concentration for titrations of the two oligopeptide solutions with various DNAs. For several titration curves presented in Figure 12, the fluorescence intensity is initially observed to increase with an increase of nucleic acid concentration until the binding becomes maximum, beyond which the fluorescence intensity decreases with a further increase of DNA concentration. This type of behavior reflects the existence of binding cooperativity for which interaction between bound oligopeptide molecules is responsible. Increasing the DNA concentration beyond a certain critical value leads to a decrease in the amount of oligopeptide bound to DNA in the associated form with

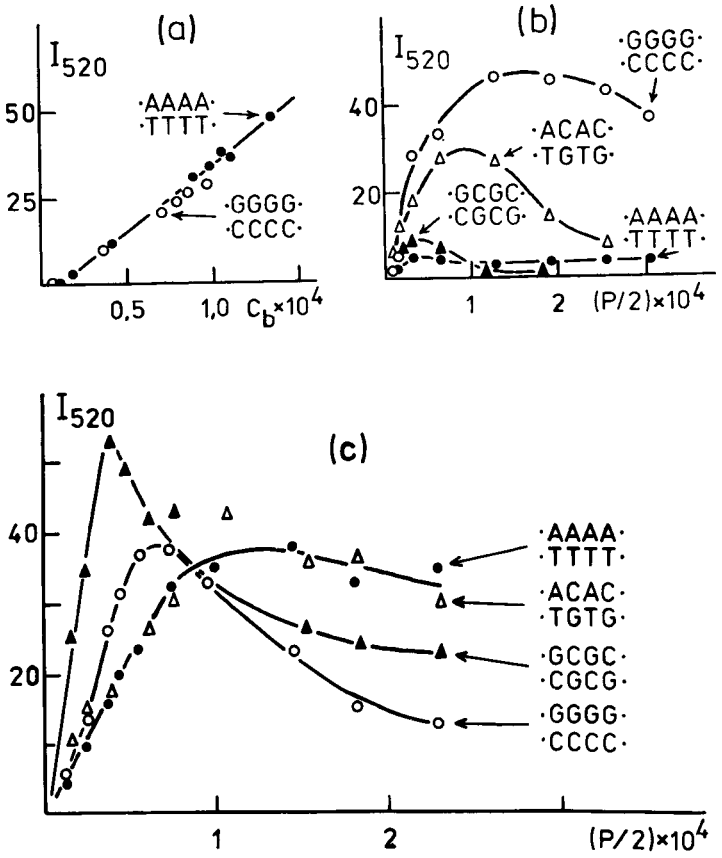


Figure 12. Fluorimetric titrations of TVP and TDP with various synthetic DNAs. (*a*) A plot of the fluorescence intensity measured at 520 nm against the molar concentration of bound TVP for complexes of TVP with poly(dA)•poly(dT) (●) and poly(dG)•poly(dC) (○). The data points were obtained by a combination of equilibrium dialysis and fluorescence measurements. Conditions were the same as those in Fig. 10. (*b*, *c*) Variation of the fluorescence intensity at 520 nm with the nucleic acid concentration (P/2) for complexes of TVP (*b*) and TDP (*c*) with various DNAs. (○) Poly(dG)•poly(dC); (△) poly(dA-dC)•poly(dT-dG); (▲) poly(dG-dC)•poly(dG-dC); (●) poly(dA)•poly(dT). Concentrations of TDP and TVP were 4×10^{-5} M and 8×10^{-5} M, respectively. Conditions were the same as those in Fig. 9 and Fig. 10 for titrations of TDP and TVP, respectively.

a concomitant increase in the amount of bound monomers.

At a relatively high DNA/oligopeptide ratio, there is a marked difference between the titration curves obtained with various DNAs. For TVP-DNA complexes, the fluorescence intensity is found to decrease in the following order: poly(dG)·poly(dC) > poly(dA-dC)·poly(dT-dG) > poly(dG-dC)·poly(dG-dC) > poly(dA)·poly(dT). This suggests the same order of sequence preferences for binding of TVP to DNA. This is confirmed by equilibrium dialysis experiments (Fig. 10 and similar data for other DNAs).

From the experimental curves obtained from the binding of TDP to various nucleic acids, the conclusion can be drawn that this oligopeptide exhibits a quite distinct order of sequence preferences compared with that of TVP (Fig. 12). At relatively high DNA/oligopeptide ratios, the fluorescence intensity of TDP-DNA complexes is found to decrease in the following order: poly(dA-dC)·poly(dT-dG) ≥ poly(dA)·poly(dT) > poly(dG-dC)·poly(dG-dC) > poly(dG)·poly(dC), thereby indicating that the amount of TDP bound to these sequences in a self-associated form is decreased in the same order. The observed preference of TVP for binding to poly(dG)·poly(dC) is consistent with the proposed protein–nucleic acid recognition code in which valine is regarded as a GC-coding residue (Gursky et al. 1976a, 1977, 1979). Since valyl side chains are inert, the observed specificity is probably attributed to hydrogen-bonding interactions between the backbone C=O groups of TVP and guanine 2-amino groups. Further support comes from our observations that TDP binds more strongly to poly(dA-dC)·poly(dT-dG) than to poly(dG)·poly(dC). The former sequence matches perfectly the sequence of coding valyl and threonyl residues in TDP (Fig. 8), thereby explaining the observed sequence specificity. A prominent feature of DNA sequences recognized by various repressors and activators is the existence of long DNA stretches with an asymmetric distribution of guanine bases between the two polynucleotide strands. Protein recognition sites appear to be complementary to such DNA sequences (Gursky et al. 1976a, 1977). This is consistent with our observations that at high DNA/oligopeptide ratios, poly(dG-dC)·poly(dG-dC) exhibits a low affinity for associated oligopeptide species (Fig. 12) binding TDP and TVP preferentially in the nonassociated form. Since the sequence 5′ TGTGAG represents a dominating motif in the *lac* operator, our experiments suggest that TDP and ECP are capable of selective binding to the *lac* operator. For binding of ECP, this was recently confirmed by the footprinting technique using the above-described restriction fragments containing the UV5 *lac* promoter-operator region (A.V. Skamrov and R.S. Beabealashvilly, unpubl.). It is of interest that there is a low extent of homology between the aminoterminal sequence of the *lac* repressor (residues 1–59) and that of ECP, although these two sequences can recognize the same base-pair sequence on DNA. Evidently, ECP contains essential residues for recognition of *lac* operator.

Determination of the structures of two regulatory proteins has greatly added to our knowledge (Anderson et al. 1981; McKay and Steitz 1981). Detailed models have been developed for complexes of these proteins with DNA that imply α-helical regions of the two proteins in the protein recognition sites. However, there is no direct evidence that α-helical regions can recognize specific DNA sequences. In addition, the model of McKay and Steitz and that of Anderson et al. are in disagreement with recent experimental data (Gutte et al. 1980; Kolb and Buc 1982). The aminoterminal sequence 1–22 of the *cro* repressor appears to recognize the specific λ operator sequence (Gutte et al. 1980), even when the postulated *cro* recognition helix at sequence 27–36 (Anderson et al. 1981) is eliminated. The aminoterminal sequence of the *cro* repressor exhibits a correspondence with the base-pair sequence of the λ operator of the same type as that shown in Figure 8 for binding of *lac* repressor to the *lac* operator (Gursky et al. 1979). Residues 1–22 of the *cro* protein form a β strand and two α helices. If these short and relatively unstable α helices undergo a transition to a deformed β structure upon binding to DNA, the suggested modes of binding of the *cro* protein and *lac* repressor are basically similar.

Our present observations show that β-associated oligopeptide species bind to DNA and exhibit base-sequence preferences that point toward the existence of a degenerated protein–nucleic acid recognition code.

REFERENCES

ADLER, K., K. BEYREUTHER, E. FANNING, N. GEISLER, B. GRONENBORN, A. KLEMM, B. MÜLLER-HILL, M. PFAHL, and A. SCHMITZ. 1972. How *lac* repressor binds to DNA. *Nature* **237:** 322.

ANDERSON, W.F., D.H. OHLENDORF, Y. TAKEDA, and B.W. MATTHEWS. 1981. Structure of the cro repressor from bacteriophage λ and its interaction with DNA. *Nature* **290:** 754.

ARNOTT, S. and R. CHANDRASEKARAN. 1981. Fibrous polynucleotide duplexes have very polymorphic secondary structures. In *Biomolecular stereodynamics* (ed. R. Sarma), vol. 1, p. 99. Adenine Press, New York.

CERAMI, A., E. REICH, D.C. WARD, and I.H. GOLDBERG. 1967. The interaction of actinomycin with DNA: Requirement for 2-amino group of purines. *Proc. Natl. Acad. Sci.* **57:** 1036.

COHEN, F.E., M.J.E. STERNBERG, and W.R. TAYLOR. 1981. Analysis of the tertiary structure of protein β-sheet sandwiches. *J. Mol. Biol.* **148:** 253.

DICKSON, R.C., J. ABELSON, W.M. BARNES, and W.S. RESNIKOFF. 1975. Genetic regulation—The *lac* control region. *Science* **187:** 27.

DUNN, J.J. and F.W. STUDIER. 1981. Nucleotide sequence from the genetic left end of bacteriophage T7 DNA to the beginning of gene 4. *J. Mol. Biol.* **148:** 303.

GELLERT, M., C.E. SMITH, D. NEVILLE, and G. FELSENFELD. 1965. Actinomycin binding to DNA: Mechanism and specificity. *J. Mol. Biol.* **11:** 445.

GURSKY, G.V., V.G. TUMANYAN, A.S. ZASEDATELEV, A.L. ZHUZE, S.L. GROKHOVSKY, and B.P. GOTTIKH. 1975. A code governing specific binding of regulatory proteins to DNA and structure of stereo-specific sites of regulatory proteins. *Mol. Biol.* **9:** 635.

——. 1976a. A code controlling specific binding of regulatory proteins to DNA. *Mol. Biol. Rep.* **2:** 413.

——. 1976b. A model for the binding of *lac* repressor to the *lac* operator. *Mol. Biol. Rep.* **2:** 427.

——. 1977. A code controlling specific binding of proteins to double helical DNA and RNA. In *Nucleic acid-protein recognition* (ed. H.J. Vogel), p. 189. Academic Press, New York.

GURSKY, G.V., A.S. ZASEDATELEV, V.G. TUMANYAN, A.L. ZHUZE, S.L. GROKHOVSKY, and B.P. GOTTIKH. 1979. Complementarity and recognition code between regulatory proteins and DNA. *FEBS Proc. Meet.* **51:** 23.

GUTTE, B., S. SCHINDLER, F. STANDAR, and E. WITTSCHIEBER. 1980. Interaction of synthetic NH_2-terminal fragments of bacteriophage λ cro protein with nucleic acids. *Biochem. Biophys. Res. Commun.* **95:** 1071.

JOHNSON, A.D., A.R. POTEETE, G. LAUER, R.T. SAUER, G.K. ACKERS, and M. PTASHNE. 1981. λ Repressor and cro: Components of an efficient molecular switch. *Nature* **294:** 217.

KHORLIN, A.A., A.S. KRYLOV, S.L. GROKHOVSKY, A.L. ZHUZE, A.S. ZASEDATELEV, G.V. GURSKY, and B.P. GOTTIKH. 1980. A new type of AT-specific ligand constructed of two netropsin-like molecules. *FEBS Lett.* **118:** 311.

KOLB, A. and H. BUC. 1982. Is DNA unwound by cAMP receptor protein? *Nucleic Acids Res.* **10:** 473.

KOLCHINSKY, A.M., A.D. MIRZABEKOV, A.S. ZASEDATELEV, G.V. GURSKY, S.L. GROKHOVSKY, A.L. ZHUZE, and B.P. GOTTIKH. 1975. Structure of complexes of antibiotics of the distamycin type and actinomycin D with DNA: New experimental data on the location of the antibiotics within the narrow groove of DNA. *Mol. Biol.* **9:** 14.

KREY, A.K., R.G. ALLISON, and F.E. HAHN. 1973. Interactions of the antibiotic distamycin A with native DNA and with synthetic duplex polydeoxyribonucleotides. *FEBS Lett.* **29:** 58.

KRYLOV, A.S., S.L. GROKHOVSKY, A.S. ZASEDATELEV, A.L. ZHUZE, G.V. GURSKY, and B.P. GOTTIKH. 1979. Quantitative estimation of the contribution of pyrrolcarboxamide groups of the antibiotic distamycin A into specificity of its binding to DNA AT pairs. *Nucleic Acids Res.* **6:** 289.

KRYLOV, A.S., A.A. KHORLIN, S.L. GROKHOVSKY, A.L. ZHUZE, A.S. ZASEDATELEV, G.V. GURSKY, and B.P. GOTTIKH. 1980. Synthetic ligands capable of recognizing DNA AT sequences related by two-fold symmetry axis. *Dokl. Akad. Nauk SSSR* **254:** 234.

LIVSHITZ, M.A., G.V. GURSKY, A.S. ZASEDATELEV, and M.V. VOLKENSTEIN. 1979. Equilibrium and kinetic aspects of protein-DNA recognition. *Nucleic Acids Res.* **6:** 2217.

LUCK, G., C. ZIMMER, K.-E. REINERT, and F. ARCAMONE. 1977. Specific interactions of distamycin A and its analogs with (A•T)-rich and (G•C)-rich duplex regions of DNA and deoxyribonucleotides. *Nucleic Acids Res.* **4:** 2655.

MCKAY, D.B. and T.A. STEITZ. 1981. Structure of catabolite gene activatior protein at 2.9 Å resolution suggests binding to left-handed B-DNA. *Nature* **290:** 744.

MIKHAILOV, M.V., S.M. NIKITIN, A.L. ZHUZE, A.S. ZASEDATELEV, G.V. GURSKY, and B.P. GOTTIKH. 1981. Bis-actinomycin D: A new synthetic sequence-specific ligand. *FEBS Lett.* **136:** 53.

MÜLLER, W. and D.M. CROTHERS. 1968. Studies of the binding of actinomycin and related compounds to DNA. *J. Mol. Biol.* **35:** 251.

——. 1975. Interactions of heteroaromatic compounds with nucleic acids. *Eur. J. Biochem.* **54:** 267.

MÜLLER-HILL, B. 1975. *Lac* repressor and *lac* operator. *Prog. Biophys. Mol. Biol.* **30:** 227.

NIKITIN, S.M., S.L. GROKHOVSKY, A.L. ZHUZE, M.V. MIKHAILOV, A.S. ZASEDATELEV, G.V. GURSKY, and B.P. GOTTIKH. 1981. DNA base pair sequence specific ligands. V. Actinomycin D analogs substituted at position 7 of phenoxazone chromophore. *Bioorg. Khim.* **7:** 542.

OGATA, R.T. and W. GILBERT. 1979. DNA-binding site of *lac* repressor probed by dimethylsulfate methylation of *lac* operator. *J. Mol. Biol.* **148:** 253.

RETCHINSKY, V.O., L.P. SAVOTCHKINA, and R.S. BEABEALASHVILLY. 1981a. Construction of promoter-probe plasmid vector. *Mol. Biol.* **15:** 950.

RETCHINSKY, V.O., R.S. BEABEALASHVILLY, A.A. KHORLIN, S.L. GROKHOVSKY, A.S. KRYLOV, A.S. ZASEDATELEV, A.L. ZHUZE, G.V. GURSKY, and B.P. GOTTIKH. 1981b. "Bis-netropsin" as a selective inhibitor of DNA dependent RNA synthesis. *Dokl. Akad. Nauk SSSR* **259:** 244.

STRELTSOV, S.A., A.A. KHORLIN, A.N. SUROVAYA, G.V. GURSKY, A.S. ZASEDATELEV, A.L. ZHUZE, and B.P. GOTTIKH. 1980. Specific interaction between oligovaline and double-stranded nucleic acids. *Biophyzik* **25:** 929.

SUTCLIFFE, J.G. 1979. Complete nucleotide sequence of the *E. coli* plasmid pBR322. *Cold Spring Harbor Symp. Quant. Biol.* **43:** 77.

TONIOLO, C. and G.M. BONORA. 1974. Conformational properties of homo-oligo(L-valine)s in solution. Comparison with isoleucine. *Makromol. Chem.* **175:** 2203.

WARTELL, R.M., J.E. LARSON, and R.D. WELLS. 1974. Netropsin. A specific probe for A-T regions of duplex deoxyribonucleic acid. *J. Biol. Chem.* **249:** 6719.

ZASEDATELEV, A.S., G.V. GURSKY, and M.V. VOLKENSTEIN. 1973. Binding isotherms of small molecules to DNA. *Stud. Biophys.* **40:** 79.

ZASEDATELEV, A.S., G.V. GURSKY, C. ZIMMER, and H. THRUM. 1974. Binding of netropsin to DNA and synthetic polynucleotides. *Mol. Biol. Rep.* **1:** 337.

ZASEDATELEV, A.S., A.L. ZHUZE, C. ZIMMER, S.L. GROKHOVSKY, V.G. TUMANYAN, G.V. GURSKY, and B.P. GOTTIKH. 1976. Stereo-chemical model of the molecular mechanism of "recognition" of AT-pairs in the binding of antibiotic distamycin A and netropsin. *Dokl. Akad. Nauk SSSR* **231:** 1006.

ZIMMER, C. 1975. Effects of the antibiotics netropsin and distamycin A on the structure and function of nucleic acids. *Prog. Nucleic Acids Res. Mol. Biol.* **15:** 285.

ZIMMER, C., K.-E. REINERT, G. LUCK, U. WÄHNERT, G. LÖBER, and H. THRUM. 1971. Interaction of oligopeptide antibiotics netropsin and distamycin A with nucleic acids. *J. Mol. Biol.* **58:** 329.