Robey, E., & Axel, R. (1990) *Cell 60*, 697-700.
Smith, D. H., Byrn, R. A., Marsters, S. A., Gregory, T., Groopman, J. E., & Capon, D. J. (1987) *Science 238*, 1704-1707.
Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., & Klenk, D. C. (1985) *Anal. Biochem. 150*, 76-85.
Steiner, L. A., & Porter, R. R. (1967) *Biochemistry 6*, 3957-3970.

Incorporation of a Complete Set of Deoxyadenosine and Thymidine Analogues Suitable for the Study of Protein Nucleic Acid Interactions into Oligodeoxynucleotides. Application to the *Eco*RV Restriction Endonuclease and Modification Methylase†

Patrick C. Newman,‡ Victor U. Nwosu,‡,§ David M. Williams,‖,⊥ Richard Cosstick,‖ Frank Seela,# and Bernard A. Connolly*,‡

Department of Biochemistry (SERC Molecular Recognition Centre), University of Southampton, Southampton SO9 3TU, U.K., Department of Chemistry, University of Liverpool, Liverpool L69 3BX, U.K., and Laboratorium fur Organische and Bioorganische Chemie, Universitat Osnabruck, D-4500 Osnabruck, West Germany

Received April 11, 1990; Revised Manuscript Received July 20, 1990

ABSTRACT: A complete set of dA and T analogues designed for the study of protein DNA interactions has been prepared. These modified bases have been designed by considering the groups on the dA and T bases that are accessible to proteins when these bases are incorporated into double-helical B-DNA [Seeman, N. C., Rosenberg, J. M., & Rich, A. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 804-808]. Each of the positions on the two bases, having the potential to interact with proteins, have been subject to nondisruptive, conservative change. Typically a particular group (e.g., the 6-NH_2 of dA or the 5-CH_3 of T) has been replaced with a hydrogen atom. Occasionally keto groups (the 2- and 4-keto oxygen atoms of T) have been replaced with sulfur. The base set has been incorporated into the self-complementary dodecamer d(GACGATATCGTC) at the central d(ATAT) sequence. Melting temperature determination shows that the modified bases do not destabilize the double helix. Additionally, circular dichroism spectroscopy shows that almost all the altered bases have very little effect on overall oligodeoxynucleotide conformation and that most of the modified oligomers have a B-DNA type structure. d(GATATC) is the recognition sequence for the *Eco*RV restriction modification system. Initial rate measurements (at a single oligodeoxynucleotide concentration of 20 μM) have been carried out with both the *Eco*RV restriction endonuclease and modification methylase. This has enabled a preliminary identification of the groups of the dA and T bases within the d(GATATC) sequence that make important contacts to both proteins.

Many proteins show a very high specificity of interaction with particular DNA sequences. Two classes of proteins have been extensively studied and are responsible for many of the ideas concerning the mechanisms by which this high specificity arises. These are the repressor proteins that bind to DNA sequences between 14 and 17 base pairs in length (Ohlendorf & Matthews, 1983; Pabo & Sauer, 1984; Brennan & Matthews, 1989) and the type II restriction enzymes that cleave DNA at recognition sequences usually 4 or 6 base pairs long (Modrich, 1982; Chirikjian, 1987; Bennett & Halford, 1989). Results obtained with both groups have led to two general models to explain how specific protein DNA interactions occur. The first has been termed direct readout, and here proteins interact directly with the bases in DNA (Otwinowski et al., 1988; Matthews, 1988; Brennan & Matthews, 1989). At its simplest this model assumes that the overall B-DNA geometry does not change with base sequence and that specificity arises from a complementary binding of a protein to a unique set of functional groups that may only be presented by its cognate sequence. The direct interaction of proteins with bases has been observed for several repressor proteins (Anderson et al., 1987; Jordan & Pabo, 1988; Aggarwal et al., 1988; Wohlberger et al., 1988) and the *Eco*RI restriction endonuclease (McClarin et al., 1986) by using X-ray crystallography. The second model is called indirect readout. Here a particular sequence of bases changes the conformation of the DNA, for example, by altering the relative dispositions of the phosphate

† B.A.C. was supported by grants from the U.K. SERC and particularly by the setting up of a molecular recognition initiative centre by this body. Additional funds were contributed by the EEC (Science Plan). B.A.C. is a Lister Institute research fellow. R.C. was supported by the U.K. SERC. P.C.N. and D.M.W. were SERC-supported research students.

* Address correspondence to this author.
‡ University of Southampton.
§ Present address: European Molecular Biology Laboratory, D-6900 Heidelberg, West Germany.
‖ University of Liverpool.
⊥ Present address: Max-Planck-Institut fur experimentelle Medizin, D-3400 Göttingen, West Germany.
Universitat Osnabruck.

0006-2960/90/0429-9891$02.50/0

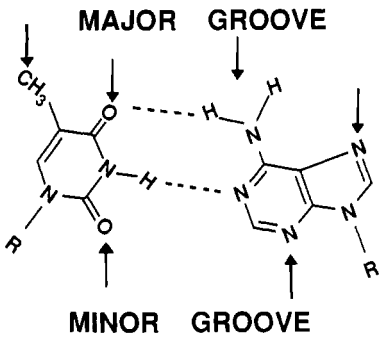

FIGURE 1: Structure of a T/dA base pair in double-stranded DNA. The arrows indicate the parts of each base that are capable of interacting with proteins in either the major or minor grooves. Most interactions are hydrogen bonds with the exception of a potential van der Waals contact between proteins and the 5-CH_3 group of T (Seeman et al., 1976). Similar considerations apply to base pairs containing dG and dC.

groups. The protein then makes specific interactions with the altered phosphate group array. So far, this has only been observed with the Trp repressor (Otwinowski et al., 1988). For the direct readout mechanism the groups on the bases in double-stranded B-DNA that are accessible to proteins have been determined (Seeman et al., 1976). Figure 1 shows these groups for a T/dA base pair [similar features are found for dA/T, dG/dC, and dC/dG base pairs; see Seeman et al. (1976)]. Most of the interactions between bases and proteins will be hydrogen bonds made via either the major or the minor groove of B-DNA. The only exception to this hydrogen-bonding rule is the 5-CH_3 group of T capable of a van der Waals interaction with proteins in the major groove. The direct readout mechanism can be probed by using base analogues, and Figure 1 gives the sites within the T and dA bases that must be altered to determine if they interact with a protein. At its most straightforward, this approach consists of the deletion of one of the potential contact sites shown in Figure 1, usually by replacement of the atom or groups of atoms concerned with hydrogen. Examples include the replacement of the 6-NH_2 group of dA or the 5-CH_3 group of T with hydrogen. As repressor proteins and restriction endonucleases show activity with short oligodeoxynucleotides, the easiest procedure is to prepare oligodeoxynucleotides containing appropriate base analogues. This approach has been used extensively with repressor proteins (Yansura et al., 1977; Goeddel et al., 1978; Fisher & Caruthers, 1979) and restriction endonucleases (Dwyer-Hallquist et al., 1982; Ono et al., 1984; Yolov et al., 1985; Fliess et al., 1986, 1988; Seela & Driller, 1986; Jiricny et al., 1986; Brennan et al., 1986; McLaughlin et al., 1987; Seela & Kehne, 1987; Ono & Ueda, 1987; Hayakawa et al., 1988; Mazzarelli et al., 1989). Usually the most difficult part of this approach is the preparation of suitable base analogues and their incorporation into oligodeoxynucleotides. None of the above studies have been completely systematic in the sense of using a complete set of base analogues thereby allowing the investigation of all the potential protein contact points shown for the dA and T bases in Figure 1. As a step in this direction, this paper reports the incorporation of a full set of dA and T analogues, which are suitable for the study of protein DNA interactions, into oligodeoxynucleotides.

The oligodeoxynucleotides prepared have been used in a preliminary study of the recognition of DNA by the restriction endonuclease and modification methylase *Eco*RV. These enzymes recognize d(GATATC) sequences with the nuclease cleaving between the central T and dA residues (Schildkraut et al., 1984; D'Arcy et al., 1985) and the methylase adding a CH_3 group to the first dA (Nwosu et al., 1988). Both the enzymes show activity with the self-complementary 12-mer d(GACGATATCGTC), and the dA and T analogues have been incorporated into the central d(ATAT) tetramer. The attraction of the *Eco*RV system is that the genes for both proteins have been cloned and sequenced (Bougueleret et al., 1984) and that overproducing strains for both the endonuclease (Bougueleret et al., 1985) and the methylase (Nwosu et al., 1988) are available. Simple high-yielding purification schemes for both enzymes have been described (D'Arcy et al., 1985; Luke et al., 1987; Nwosu et al., 1988). The endonuclease has been crystallized as the free enzyme and bound to cognate and noncognate sequences (D'Arcy et al., 1985; Winkler et al., 1986). One anticipates the eventual availability of crystallographic data which should complement the base modification studies described here. Kinetic studies with the endonuclease using plasmids containing cognate sites (Luke et al., 1987; Halford & Goodall, 1988) and nearly cognate sites (Taylor & Halford, 1989) and also under conditions of relaxed specificity (Halford et al., 1986) have shown that this enzyme discriminates between d(GATATC) cognate sites and the most favored nearly cognate site by a factor of 10^6. Mechanisms responsible for discrimination included (1) higher intrinsic catalytic rates and recognition of cognate as opposed to noncognate sites; (2) decrease in affinity for the essential cofactor Mg^{2+} with noncognate sites, and (3) switch from simultaneous double cutting of both strands with cognate sites to nicking of one strand with noncognate sites, allowing repair by DNA ligase. Finally, the endonuclease and methylase have no amino acid sequence homology (Bougueleret et al., 1984) and catalyze very different reactions. Thus, the base analogue containing oligodeoxynucleotides should elucidate the mechanisms by which two very different proteins recognize the same d(GATATC) sequences.

Experimental Procedures

High-pressure liquid chromatography (HPLC) was performed by using a Varian 5000 liquid chromatograph and a Varian UV 50 variable-wavelength detector. The C8 or C18 reverse-phase material Apex I octasilylsilica (or octadecylsilylsilica) (5-μm particle size) packed into columns 25 × 0.45 cm supplied by Jones Chromatography (Llanbradach, Glamorgan, U.K.) was used. Gradients were prepared from 0.1 M triethylammonium acetate, pH 6.5, containing 5% CH_3CN (HPLC buffer A) and 0.1 M triethylammonium acetate, pH 6.5, containing 65% CH_3CN (HPLC buffer B) at a flow rate of 1 mL min^{-1}. Columns were run at room temperature for deoxynucleoside composition analysis and at 50 °C for oligodeoxynucleotide purification and kinetic studies. Peaks were integrated by interfacing the detector output to an Apple IIe microcomputer fitted with a standard peak integration program. Oligodeoxynucleotide desalting was performed by using C-18 columns and gradients prepared from 0.1 M triethylammonium bicarbonate, pH 8, containing 5% (HPLC buffer C) and 65% (HPLC buffer D) CH_3CN at room temperature and a flow rate of 1 mL min^{-1}. Unless otherwise stated, all compounds were detected at 254 nm. The measurement of the thermal melting points (T_m) of the oligodeoxynucleotides and the recording of their circular dichroism (CD) spectra were performed as described (Connolly & Newman, 1989).

Synthesis of Oligonucleotides. All oligodeoxynucleotides were prepared by using an Applied Biosystems 381A automatic DNA synthesizer, and all the reagents required (with the exception of the modified bases) were purchased from this company. Synthesis and purification of the parent 12-mer

d(GACGATATCGTC) is given in more detail elsewhere (Connolly & Newman, 1989) as are the preparations of oligodeoxynucleotides containing 4-thiothymidine (^{4S}T), 2-thiothymidine (^{2S}T), and 5-methyl-2-pyrimidinone 1-β-D-(2-deoxyriboside) (^{4H}T). The preparation of oligodeoxynucleotides containing 7-deazadeoxyadenosine (d^{7C}A) (Seela & Kehne, 1987) and purine 1-β-D-(2-deoxyriboside) (dP) (Nwosu et al., 1988) have also been described. Oligodeoxynucleotides containing dU were prepared by using appropriately protected dU phosphoramidites purchased from Applied Biosystems. The synthesis of 12-mers containing 3-deazadeoxyadenosine (d^{3C}A) will be described later (information concerning this analogue is available on application from R.C.). All syntheses were performed on a 1-μmol scale by using the standard 1-μmol cyanoethyl phosphoramidite synthesis cycle. Modified deoxynucleoside phosphoramidites were dissolved in dry CH_3CN (to give the standard 0.1 M solutions) and passed through disposable 0.5-μm PTFE filters (Acro-LC3S, Gelman Sciences, Ann Arbor, MI) to remove any particulate material prior to use. These modified bases were introduced at the X position of the synthesizer, and no changes to the synthesis cycle were used for their coupling. All couplings, both of standard and modified bases, proceeded in yields of about 100% as monitored by dmt cation release. d(GACG-[^{6Me}A]TATCGTC) was prepared from d(GACGATATCGTC) by using *Eco*RV methylase as described (Nwosu et al., 1988). All oligodeoxynucleotides were purified by reverse-phase HPLC as previously described (Nwosu et al., 1988; Connolly & Newman, 1989). After these purification steps, the oligodeoxynucleotides were desalted by HPLC using a gradient prepared from HPLC buffers C and D (0–80% D over 20 min). The oligodeoxynucleotides eluted at around 12 min and were collected in Eppendorf tubes. Buffer components were removed by evaporation to dryness in a Sorvall Speed-Vac and last traces of triethylammonium bicarbonate removed by coevaporation from 3 × 0.3 mL of methanol. The resulting oligodeoxynucleotides were dissolved in 0.5 mL of H_2O and stored frozen at −20 °C.

Determination of the Extinction Coefficients of Oligodeoxynucleotides. A small aliquot (to give an absorbance of between 0.2 and 0.5 at 254 nm) of the appropriate oligodeoxynucleotide was dissolved in 0.5 mL of 10 mM KH_2PO_4, pH 7, containing 10 mM $MgCl_2$ and the absorbance noted. Snake venom phosphodiesterase (Boehringer-Mannheim, Lewes, East Sussex, U.K.) (5 μg) and alkaline phosphatase (special quality for molecular biology, Boehringer-Mannheim) (5 μg) were added and the mixtures incubated at 30 °C until the absorbance at 254 nm stopped increasing (about 1 h). The final absorbance was noted. The hyperchromicity of the oligodeoxynucleotide is defined as the value of the final absorbance divided by the initial absorbance. The extinction coefficient of the oligodeoxynucleotide is the sum of the extinction coefficients of the constituent monomeric deoxynucleosides divided by hyperchromicity. As all the 12-mers prepared are self-complementary and the double-stranded structures formed are the substrates for both endonuclease and methylase, the extinction coefficients have been determined for the double-stranded forms by summing the absorbances of 24, not 12, bases. The following extinction coefficients (M^{-1} cm^{-1}) at 254 nm for the individual bases were used: dC, 6 × 10^3; dG, 13.5 × 10^3; T, 7 × 10^3; dA, 14.3 × 10^3; dU, 9.3 × 10^3; d^{6Me}A, 12.1 × 10^3 (Fasman, 1975), dP, 6.5 × 10^3 (Nair & Chamberlain, 1984); d^{7C}A, 7.8 × 10^3 (Seela & Kehne, 1983); d^{3C}A, 9.3 × 10^3 (R. Cosstick, unpublished data); ^{4S}T, 2.9 × 10^3 (Fox et al., 1959); ^{2S}T, 6.3 × 10^3 (Faerber & Scheit, 1970); ^{4H}T, 0.4 × 10^3 (Wightman & Holy, 1973). The extinction coefficients of dP, d^{7C}A, ^{4S}T, ^{2S}T, and ^{4H}T at 254 nm were calculated from the extinction coefficients at the λ_{max} (given in the references) multiplied by the ratios of the absorbance values seen at 254 nm and the λ_{max}. These two absorbance values were obtained by measuring the UV spectra of the appropriate deoxynucleosides in water at pH 7.

Oligodeoxynucleotide Characterization. Oligodeoxynucleotide purity was checked by C18 HPLC using a gradient composed of 0–25% HPLC buffer B over 20 min. The identity of each oligodeoxynucleotide was determined by using base composition analysis. This was most conveniently performed on the same samples used in the determination of the hyperchromicity values. At the end of this experiment aliquots of the digested oligodeoxynucleotides were injected onto C18 columns to quantitatively determine the deoxynucleosides present. A gradient composed of 0% HPLC buffer B for 5 min followed by 0–25% B for 10 min and finally 25–75% B for 10 min was used. The deoxynucleosides eluted in the following order: dC, dG, ^{4H}T, T, dP, dA, d^{7C}A, d^{6Me}A, ^{2S}T, ^{4S}T. The exact elution times varied with different columns, small day to day variations in the buffers used, and small differences in the running temperatures. Although these elution times could be rather variable, the elution order remained unchanged. An exception was seen with d^{3C}A, which eluted either just before or just after dG. After this analytical run, the identities of each of the peaks were confirmed by coinjection of an aliquot of the digest sample with standards of the monomeric deoxynucleosides. Most bases were detected at 254 nm and quantified by peak integration using the above extinction coefficients. ^{4S}T and ^{4H}T were detected at 335 and 315 nm, respectively, as previously described (Connolly & Newman, 1989).

Purification of the EcoRV Endonuclease and Methylase. The endonuclease was purified from an overproducing *Escherichia coli* strain containing plasmid pTZ115 (Bougueleret et al., 1985) as described (D'Arcy et al., 1985). This is a two-column purification protocol, and after the second column, the enzyme is precipitated with $(NH_4)_2SO_4$. This precipitate was dissolved in a minimal volume (1–2 mL) of 20 mM KH_2PO_4, pH 7.5, containing 200 mM NaCl, 1 mM DTT, and 1 mM EDTA and dialyzed against this buffer to remove $(NH_4)_2SO_4$. The resulting enzyme solution was lyophilized and redissolved in 1 mL of a 50:50 water/glycerol mixture. This solution was stored at −20 °C. The methylase was purified from an overproducing *E. coli* strain containing the plasmia pVIC 1 as described (Nwosu et al., 1988). Once again this is a two-column purification that ends with an $(NH_4)_2SO_4$ precipitation. The precipitate was dissolved in 1 mL of 20 mM Tris, pH 7.4, containing 100 mM NaCl, 1 mM DTT, and 1 mM EDTA and dialyzed against this buffer to remove $(NH_4)_2SO_4$. Glycerol was added to the solution to give a final concentration of 50% and the solution stored at −20 °C.

Protein Concentration Determination. The concentrations of the endonuclease and methylase in the above stock solutions were determined by using a Bio-Rad (Watford, Herts, U.K.) protein assay kit based on Coomassie blue G-250 with bovine serum albumin as standard. The endonuclease concentration was 20 mg mL^{-1} and the methylase 3 mg mL^{-1}. These were converted into molar terms by assuming that the nuclease is active as a dimer (D'Arcy et al., 1985; Luke et al.; 1987) of subunit molecular weight 29 000 (Bougueleret et al., 1984) (i.e., 58 000 daltons = 1 mol) and the methylase is active as a monomer (Nwosu et al., 1988; Garnett & Halford, 1988) of molecular weight 34 000 (i.e., 34 000 daltons = 1 mol).

Enzyme Assays. The endonuclease was assayed at 25 °C in 200-μL volumes of 50 mM Hepes–NaOH, pH 7.5, con-

taining 100 mM NaCl, 10 mM $MgCl_2$, and 20 μM (double-stranded concentration) oligodeoxynucleotide. The progress of the reaction was monitored by HPLC using a C18 column and a linear gradient composed of 0–25% HPLC buffer B over 20 min. Under these conditions the substrate 12-mers eluted at about 17 min and the two product hexamers (which were always resolved from each other) between 13 and 16 min. Initially, experiments were carried out by using a large quantity (40 μg, 3.5 μM dimeric enzyme) of the endonuclease for 5 h. After this time, a further 40 μg of enzyme was added to solutions showing no or partial reaction, and the mixtures were incubated for a further 25 h. After this time, the oligodeoxynucleotides were either completely converted to two hexameric products or unhydrolyzed. The hexameric products of those that showed substrate activity were purified by using the above HPLC system and their identities determined by base composition analysis. The rates at which the substrate oligodeoxynucleotides were cleaved were determined by using the minimal amount of endonuclease required to produce a measurable rate over a time course that varied from 0–20 min (good substrates) to 0–6 h (poor substrates). This was usually between 1.14 and 11.4 μg (i.e., 0.1–1 μM dimeric enzyme concentration). However, for the oligodeoxynucleotide that contained ^{4S}T in the first T position, 57 μm (5 μM) of endonuclease was used, and with the oligomer having dP in the first dA position 114 μg (10 μM) was required. At appropriate times 15-μL aliquots were withdrawn and added to 5 μL of 0.5 M NaOH to quench the reaction. This NaOH quench could not be used with oligodeoxynucleotides containing ^{4S}T as it destroys the ^{4S}T base, giving several breakdown products and so interfering with the HPLC assay. Fortunately, oligodeoxynucleotides containing this base were so slowly cleaved by the endonuclease that it was possible to directly inject aliquots onto the HPLC for rate determination without the NaOH quench. When aliquots are injected onto the HPLC, the 12-mer substrates and two 6-mer products are separated and the amounts of material can be determined by integration of the appropriate peaks. Initially, this integration gives a dimensionless figure not easily related to the number of moles of material present. Thus, the integrator was standardized for each substrate by injecting known amounts of the oligodeoxynucleotides onto the HPLC, noting the integrator response and thereby calibrating the number of integrator units that correspond to a mole of substrate. Cleavage of the parent 12-mer d(GACGATATCGTC) clearly gives 1 equiv of d(GACGAT) and dp(ATCGTC), but hydrolysis of all the 12-mers that contain dA or T analogues within the central d(ATAT) hexamer must also give one of these common products. Therefore, known amounts of each of these hexamers were injected, allowing the calibration of the integrator response for product.

The methylase was assayed at 37 °C in 100-μL volumes of 50 mM Hepes–NaOH, pH 7.5, containing 100 mM NaCl, 1 mM *S*-adenosylmethionine (SAM), and 20 μM oligodeoxynucleotide. SAM was purchased from Boehringer-Mannheim as its bisulfate salt. Stock solutions were made up in 1 mM H_2SO_4 immediately prior to use. The progress of the reaction was monitored by HPLC using a C8 column and a linear gradient composed of 0–25% HPLC buffer B over 20 min. The substrate 12-mers eluted between 15 and 17 min and the methylated products 1–2 min later. Initially, experiments were carried out by using 34 μg (10 μM monomeric enzyme) for a 5-h time period. After this time, solutions showing no or partial reaction were treated with a further 34 μg of the methylase and an additional aliquot (to give a concentration of 2 mM) of SAM for 25 h. Under these conditions either complete or no methylation was observed for the various members of the oligodeoxynucleotide set. With those that had shown complete reaction the methylated products were purified by using the above HPLC system and their identities determined by base composition analysis. The rates at which the substrate oligodeoxynucleotides were methylated were determined by using 3.4 μg (1 μM) of the methylase except when ^{2S}T or ^{4S}T replaced the first T, d^{3C}A the second dA, and dU the second T. Here 34 μg (10 μM) of the methylase was required. These reactions were measured over time periods that varied from 30 min to 5 h. Once again, 15-μL aliquots were withdrawn and quenched into 5 μL of 0.5 M NaOH. For oligodeoxynucleotides containing ^{4S}T the reaction was stopped by placing 20-μL aliquots in boiling water. The calibration of the integrator response for the substrate 12-mers has been described, and we have assumed that the product methylated 12-mers give the same response.

Results

Modified Base Selection. The structures of dA and T are shown in Figure 2, and the groups on these bases capable of interacting with proteins have been arrowed. The purpose of this paper was to introduce an entire set of dA and T analogues, in which each potential contact had been subject to conservative replacement, into oligodeoxynucleotides for the study of protein nucleic acid interactions. The base analogues used are also illustrated in Figure 2 and for dA consist of purine 1-β-D-(2-deoxyriboside) (dP) and 3- and 7-deazadeoxyadenosine (d^{3C}A, d^{7C}A). For T the analogues deoxyuridine (dU), 5-methyl-2-pyrimidinone 1-β-D-(2-deoxyriboside) (^{4H}T), and 2- and 4-thiothymidine (^{2S}T, ^{4S}T) have been used. In a conservative base analogue a potential protein contact is completely eliminated and no extra bulk is introduced. This is most simply accomplished by replacing the atoms or group of atoms concerned with hydrogen. Clearly, the analogues dP, dU, and ^{4H}T fall into this category. Also in this group are the bases d^{3C}A and d^{7C}A in which a ring nitrogen atom is replaced by the isosteric CH function. One consequence found with some of the analogues is that Watson–Crick base pairing to complementary bases in oligodeoxynucleotides may be reduced. This occurs with dP (only one instead of the usual two hydrogen bonds to T) and is particularly acute for ^{4H}T, which cannot form any (instead of the usual two) bonds to complementary dA bases. To combat potential problems, such as duplex instability, caused by this property of ^{4H}T, we have also used ^{4S}T. Here the usual 4-keto oxygen of T is replaced with a sulfur atom rather than with hydrogen as in ^{4H}T. Clearly, ^{4S}T differs from the conservative base analogues mentioned above such as ^{4H}T. First, potential protein contacting ability is reduced rather than eliminated as the oxygen to sulfur change gives an analogue that is still a potential hydrogen-bond acceptor albeit of far weaker strength than the parent. Second, the analogue is slightly larger than the parent. The advantage of ^{4S}T over ^{4H}T is that it retains some hydrogen-bonding ability to complementary dA residues and so may be less disruptive to the overall oligodeoxynucleotide structure. We have also used ^{2S}T to probe the role of the 2-keto oxygen of T, and the same arguments used for ^{4S}T apply to this base. ^{2S}T has been used as, in contrast to all the other base analogues, the appropriate hydrogen analogue (i.e., conversion of the 2-keto oxygen to hydrogen) has (to the best of our knowledge) never been prepared. Many of the analogues shown in Figure 2 have previously been incorporated into oligodeoxynucleotides and used to study DNA recognition by various restriction endonucleases. Examples include the fol-

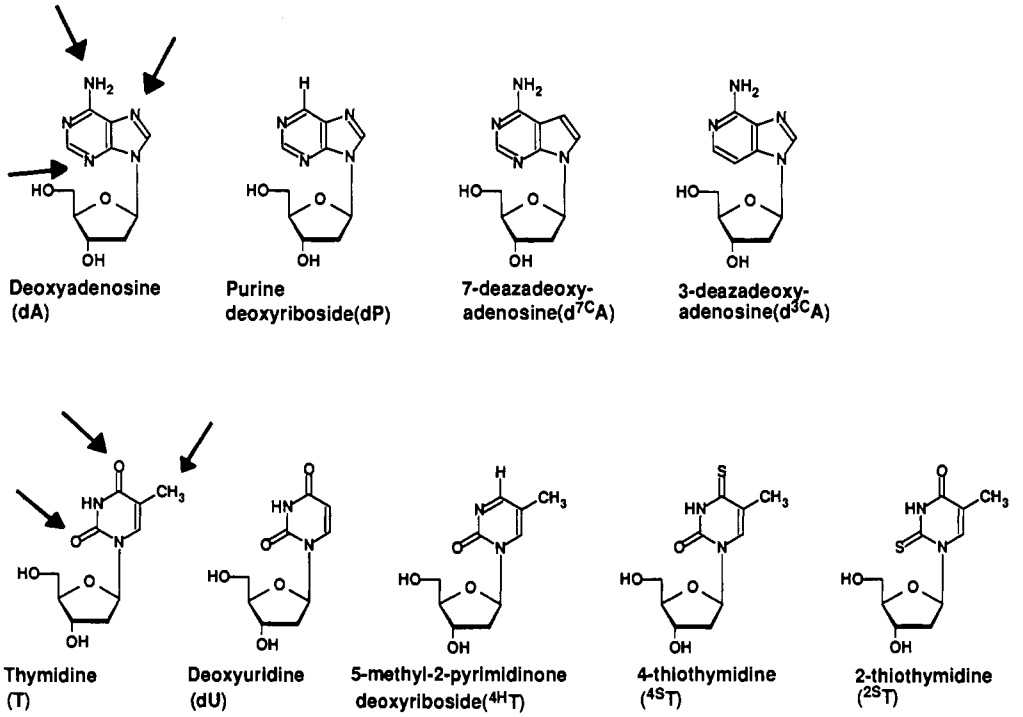

FIGURE 2: Structures of dA and T. The arrows show the parts of each base capable of interacting with proteins when they are incorporated into double-helical B-DNA. The base analogues used to probe each of these potential contacts are also illustrated.

lowing: dP, *Hin*dII, *Sal*I, and *Taq*I (Jiricny et al., 1986); *Eco*RI (McLaughlin et al., 1987); *Eco*RV (Nwosu et al., 1988; Mazzarelli et al., 1989); $d^{3C}A$, *Bgl*II, *Sau*3AI, and *Mbo*I (Ono & Ueda, 1987); $d^{7C}A$, *Bgl*II, and *Sau*3AI (Ono et al., 1984); *Hin*dII, *Sal*I, and *Taq*I (Jiricny et al., 1986); *Eco*RI (Seela & Kehne, 1987); *Eco*RV (Fliess et al., 1988); dU, *Hpa*I (Dwyer-Hallquist et al., 1982); *Eco*RII (Yolov et al., 1985); *Eco*RI (Fliess et al., 1986; Brennan et al., 1986; McLaughlin et al., 1987); *Bgl*II, *Sau*3AI, and *Mbo*I (Hayakawa et al., 1988); *Eco*RV (Fliess et al., 1986, 1988; Mazzareli et al., 1989). We have recently described the incorporation of ^{4H}T, ^{4S}T, and ^{2S}T into oligodeoxynucleotides (Connolly & Newman, 1989), and a similar synthesis for ^{4H}T has also been published (Gildea & McLaughlin, 1989). As far as we know, however, none of these last three analogues have been used as substrates for restriction endonucleases. This paper is the first time all the analogues have been brought together and incorporated into a single oligodeoxynucleotide for an extensive study of a particular enzyme system.

Properties of the Oligodeoxynucleotides Containing Modified Bases. The *Eco*RV endonuclease and methylase both recognize d(GATATC) sequences, and as a parent unmodified oligodeoxynucleotide the self-complementary 12-mer d(GACGATATCGTC) has been used. The dA and T analogues have been separately incorporated into the dA and T sites with the central recognition hexamer. This results in 14 modified oligodeoxynucleotides (listed in Tables I and II), each of which contains a single base analogue and has one potential protein contact site deleted. Each of the oligodeoxynucleotides was >97.5% pure by HPLC (data not shown). All the oligodeoxynucleotides gave correct base compositions after complete hydrolysis to their constituent monodeoxynucleosides. In general, this proved an excellent method of characterization and confirmation that a modified base was present. As typical examples Figure 3 shows that on hydrolysis d(GACGATPTCGTC) yields dC, dG, T, dA, and dP in the expected molar ratios of 3, 3, 3, 2, 1. Similarly, Figure 3 also shows that for d(GACGA[^{4S}T]ATCGTC) 1 equiv of ^{4S}T can be seen as well as the expected amounts of the four usual bases. The base compositions found for the entire oligodeoxynucleotide set are summarized in Table I.

The extinction coefficients of the oligodeoxynucleotides are also given in Table I. These are quoted per mole of double-stranded dodecamer, the true substrate for both enzymes. Due to the hypochromic effect, i.e., the decrease in absorbance seen when monodeoxynucleosides are polymerized into DNA, these extinction coefficients cannot simply be obtained by summing the absorbances of the individual bases. However, these values can be determined by addition of the extinction coefficients of the 24 constituent bases and dividing the value found by the hyperchromicity of the oligodeoxynucleotide. Here the hyperchromicity is the increase in absorbance seen when the oligodeoxynucleotide is enzymatically digested to its deoxynucleoside constituents. The hyperchromicities found for each oligomer are given in Table I and vary between 1.25 and 1.53. Hyperchromicity arises because of the base stacking in double-helical oligodeoxynucleotides and consequent π–π orbital overlap. These small differences in hyperchromicities may be due to perturbations in the base stacking consequent on the presence of the modified bases. However, the hyperchromicities were determined mainly to allow an accurate determination of concentration and are not a very good method of oligodeoxynucleotide structural characterization.

As both endonuclease and methylase require double-stranded structures, it is important to determine that all the 12-mers are in the double-helical form under enzyme assay conditions, i.e., 50 mM Hepes, pH 7.5, containing 100 mM NaCl at 25 or 37 °C. The T_m values for the oligodeoxynucleotide set are given in Table I. The parent 12-mer has a T_m of 52 °C, in excellent agreement with a value of 53 °C seen previously (Ott et al., 1985). All the modified base containing oligodeoxynucleotides have similar T_m values that lie between 48 and 56 °C. The melting curves of oligodeoxynucleotides containing ^{4H}T, ^{4S}T, and ^{2S}T have been published (Connolly & Newman, 1989), and those containing dU, dP, $d^{3C}A$, and $d^{7C}A$ were similar in appearance (data not

Table I: Hyperchromicities, Extinction Coefficients, T_m Values, and Base Composition for the Oligodeoxynucleotide Set[a]

oligonucleotide	hyperchromicity	extinction coefficient (M^{-1} cm^{-1} per double strand)	T_m (°C)	base composition
d(GACGATATCGTC)	1.48	1.66×10^5	53	dA (2.8), dG (3), dC (2.8), T (2.7)
---GPTATC---	1.25	1.83×10^5	48	dA (1.9), dG (3), dC (3.3), T (2.9), dP (0.8)
---G[^{7C}A]TATC--	1.44	1.61×10^5	48	dA (1.8), dG (3), dC (3.2), T (2.7), d^{7C}A (0.9)
---G[^{3C}A]TATC---	1.34	1.75×10^5	52	dA (1.7), dG (3), dC (3.2), T (3.0), d^{3C}A (0.9)
---GAUATC---	1.43	1.75×10^5	52	dA (3.1), dG (3), dC (3.4), T (2.2), dU (1.1)
---GA[^{4S}T]ATC---	1.53	1.55×10^5	52	dA (2.9), dG (3), dC (3.1), T (2.1), ^{4S}T (1.0)
---GA[^{4H}T]ATC---	1.40	1.65×10^5	50	dA (2.9), dG (3), dC (3.0), T (2.0), ^{4H}T (1.1)
---GA[^{2S}T]ATC---	1.31	1.85×10^5	56	dA (3.0), dG (3), dC (3.1), T (1.8), ^{2S}T (1.0)
---GATPTC---	1.29	1.79×10^5	48	dA (1.8), dG (3), dC (3.5), T (2.9), dP (0.8)
---GAT[^{7C}A]TC---	1.38	1.68×10^5	52	dA (1.8), dG (3), dC (3.4), T (3.0), d^{7C}A (0.9)
---GAT[^{3C}A]TC---	1.33	1.77×10^5	53	dA (1.6), dG (3), dC (3.4), T (3.2), d^{3C}A (1.0)
---GATAUC---	1.47	1.70×10^5	52	dA (3.2), dG (3), dC (3.4), T (2.1), dU (1.1)
---GATA[^{4S}T]C---	1.51	1.56×10^5	52	dA (2.8), dG (3), dC (3.1), T (1.8), ^{4S}T (1.2)
---GATA[^{4H}T]C---	1.37	1.69×10^5	50	dA (2.9), dG (3), dC (3.0), T (2.1), ^{4H}T (1.0)
---GATA[^{2S}T]C---	1.35	1.81×10^5	56	dA (2.7), dG (3), dC (3.3), T (1.8), ^{2S}T (0.9)

[a] The full sequence is only given for the parent dodecamer. Only the central hexameric d(GATATC) core is given for the modified dodecamers. The flanking six bases were identical with the parent. For the base composition analysis, the value for dG was set to 3 and the other bases ratioed to this number.

Table II: Initial Rates Seen at 20 μM (Double Stranded) Oligonucleotide Concentration for the Entire Oligonucleotide Set with the *Eco*RV Endonuclease and Methylase[a]

oligodeoxynucleotide	endonuclease [nmol min^{-1} mg^{-1} (%)]	methylase [nmol min^{-1} mg^{-1} (%)]
d(GACGATATCGTC)	115 (100)	35 (100)
---GPTATC---	0.12 (0.1)	0 (0)
---G[^{7C}A]TATC---	0 (0)	0 (0)
---G[^{3C}A]TATC---	2.9 (2.5)	0 (0)
---GAUATC---	3.5 (3.0)	0 (0)
---GA[^{4S}T]ATC---	0.23 (0.2)	1.8 (5.0)
---GA[^{4H}T]ATC---	0 (0)	0 (0)
---GA[^{2S}T]ATC---	35 (30)	1.4 (4.0)
---GATPTC---	1.7 (1.5)	74.2 (212)
---GAT[^{7C}A]TC---	23 (20)	29.8 (85)
---GAT[^{3C}A]TC---	3.5 (3.0)	0.8 (2.2)
---GATAUC---	1.7 (1.5)	0.1 (0.4)
---GATA[^{4S}T]C---	0 (0)	74.2 (212)
---GATA[^{4H}T]C---	0 (0)	47.2 (135)
---GATA[^{2S}T]C---	6.9 (6)	55.0 (157)

[a] The rates are given as nmol of double-stranded oligonucleotide hydrolyzed or methylated min^{-1} (mg of enzyme)$^{-1}$. The figures in parentheses are percent rates with the rate of the parent 12-mer set to 100. For the endonuclease the turnover number with the parent 12-mer is 6.7 min^{-1} and for the methylase 1.2 min^{-1}. Zero rates mean <0.01% of the rate seen with the parent 12-mer for the endonuclease and <0.05% for the methylase.

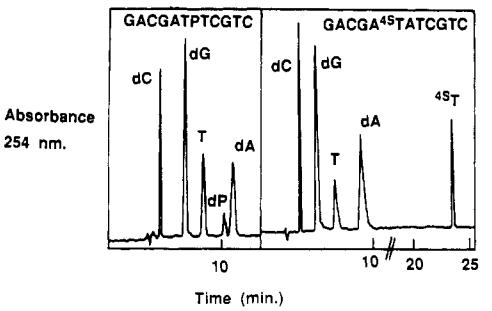

FIGURE 3: Results of the base composition analysis for d(GACGATPTCGTC) and d(GACGA[^{4S}T]ATCGTC). The mole fractions of each base present obtained from these data (and that for all the other oligodeoxynucleotides) are summarized in Table I. In the base composition analysis of d(GACGA[^{4S}T]ATCGTC) the UV detector was switched from 254 to 335 nm after dA elution for ^{4S}T detection.

shown). In all these melting profiles the absorbance begins to increase at around 30 °C and reaches a maximum at about 75 °C, giving T_m values in the 50 °C range. These results imply that the entire oligodeoxynucleotide set will be predominantly in a double-helical form at temperatures of 25 and 37 °C used to assay the endonuclease and methylase, respectively. Most of the curves were smooth, indicating a single melting transition; however, the exact shape of the curves (indicative of the degree of cooperativity) varied between oligodeoxynucleotides. The only exception to this was oligomers containing dP, which showed a distinct transition between 30 and 40 °C, probably indicative of premelting of dP/T base pairs. Perhaps the most surprising aspect of these results is the absence of any drop in T_m values seen with base analogues that cannot form their normal complement of Watson–Crick base pairs such as dP (loss of one of two hydrogen bonds to T) and ^{4H}T (loss of both hydrogen bonds to dA). This is especially significant as the self-complementary 12-mers used mean that the analogue occurs once in each strand and so affects two rather than a single base pair. A single dP/T base pair in a 16-mer has been reported to cause a T_m drop of 6 °C (Jiricny et al., 1986), and two dP/T base pairs in a 10-mer caused a lowering of T_m by 11 °C (McLaughlin et al., 1987). Considerable destabilization has been reported for ^{4H}T; in a 10-mer two ^{4H}T/dA base pairs lowered the T_m from 41 to <15 °C, and with a 12-mer the presence of two ^{4H}T/dA base pairs gave an anomalous melting curve from which a T_m could not be abstracted (Gildea & McLaughlin, 1989). In our hands dP and ^{4H}T have very little destabilizing effect, and this presumably reflects the different sequences and position of the incorporated analogue that we have used in comparison to other workers. It has been suggested that double-helix destabilization seen with mismatches (i.e., non-Watson–Crick combinations of the four normal bases) is due as much to steric clashes as simple hydrogen-bond loss (Fersht, 1987). As dP and ^{4H}T simply result in disappearance of Watson–Crick hydrogen bonds but do not induce any steric clashes between opposing strands, it is perhaps not

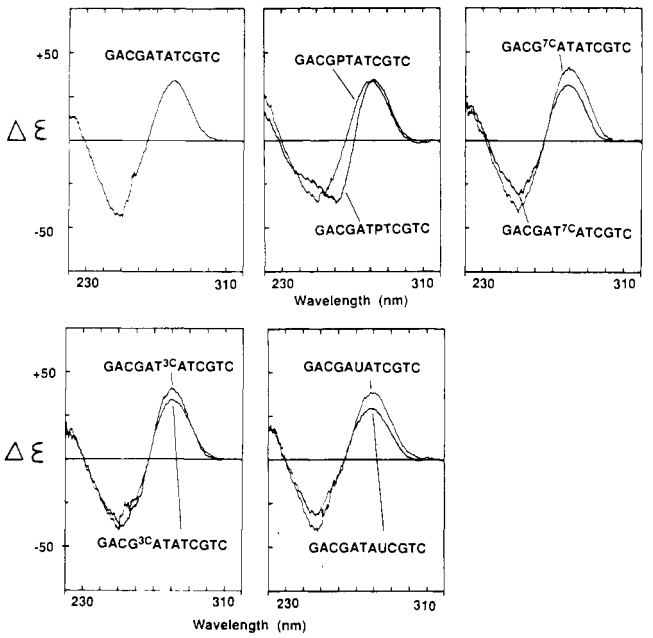

FIGURE 4: CD spectra of the parent 12-mer and dodecamers containing dA analogues and dU. For the spectra of dodecamers containing ^{4S}T, ^{4H}T, and ^{2S}T, see Connolly and Newman (1989).

surprising that under some conditions no drop in T_m is seen. At the present time we have not further analyzed the T_m data, e.g., by measuring T_m values as a function of concentration. This enables further characterization of the melting transition such as determination of the transition enthalpy and a study of the cooperativity of melting.

Finally we have measured the circular dichroism (CD) spectra of all the oligodeoxynucleotides. Those for the parent 12-mer and oligomers containing ^{4H}T, ^{4S}T, and ^{2S}T have been published (Connolly & Newman, 1989). Those for the parent 12-mer and dodecamers containing dU, dP, d^{3C}A, and d^{7C}A are shown in Figure 4. As expected, the parent 12-mer shows a typical B-DNA spectrum with equally sized positive and negative peaks centered around 280 and 250 nm, respectively (Ivanov et al., 1973; Saenger, 1984; Fairall et al., 1989). Almost identical spectra were seen for oligodeoxynucleotides containing dU, ^{4S}T, d^{3C}A, d^{7C}A, and d(GACGPTATCGTC), suggesting that these also form typical B-DNA structures. This has previously been seen for oligomers containing dU (Hayakawa et al., 1988), d^{3C}A (Ono & Ueda, 1987), and d^{7C}A (Ono et al., 1984). With dodecamers containing ^{4H}T and d(GACGATPTCGTC) slight deviations away from the usual B-DNA type CD spectra were seen with the negative peak shifted from 250 to 260 nm. In each case, however, a shoulder was visible at 250 nm. These spectra are, however, still reminiscent of B-DNA, and the most likely explanation is that the oligodeoxynucleotides adopt this form with some slight distortions away from the classical B-DNA structure, probably due to the lack of base pairing seen with dP and ^{4H}T. Previous CD spectra of ^{4H}T-containing oligodeoxynucleotides did not give the results expected for B-DNA (Gildea & McLaughlin, 1989). However, in that study, in contrast to here, the oligomers produced did not form double-helical structures. Only with ^{2S}T-containing oligomers were the CD spectra not typical of B-DNA. The spectra observed with a large positive peak at 265 nm are more typical of the A conformation (Ivanov et al., 1973; Saenger, 1985; Fairall et al., 1989). However, at present we are unable to say whether these 12-mers do exist as the A-form or are a distorted B-form with an unusual CD spectra. Nevertheless, because of this discrepancy, caution must be exercised in interpreting enzymatic results seen with ^{2S}T-containing oligodeoxynucleotides.

Interaction of the Oligodeoxynucleotides with the EcoRV Endonuclease. Table II shows the rates at which the endonuclease cleaves 20 μM concentrations of each of the oligodeoxynucleotides. These can be divided into two groups, first, nonsubstrates {d(GACG[^{7C}A]TATCGTC), d(GACGATA-[^{4S}T]CGTC), and both ^{4H}T-containing 12-mers} and, second, substrates (all other oligodeoxynucleotides) in which the rates vary from 115 nmol min^{-1} mg^{-1} for the parent to 0.12 nmol min^{-1} mg^{-1} for d(GACGPTATCGTC). All these rates were obtained from HPLC traces of the time course of the reactions (data not shown). These HPLC traces showed that all the substrate oligodeoxynucleotides, even the poorest d-(GACGPTATCGTC), could be completely cleaved to two hexameric products, providing enough endonuclease was added and the mixtures were incubated for a long time. In contrast, no cleavage was observed with the four nonsubstrate oligodeoxynucleotides. On the basis of the limits of our assay system the four nonsubstrates must be cleaved at less than 0.01% of the rate of the parent 12-mer rate. We have also determined (data not shown) that d(GACG[^{6Me}A]-TATCGTC) is a nonsubstrate. This oligodeoxynucleotide which contains *N*-6-methyldeoxyadenosine in the first dA position is the product of *Eco*RV methylase action (Nwosu et al., 1988). This is clearly not an oligodeoxynucleotide containing a conservatively modified base due to the bulk of the CH_3 group. However, as the physiological role of the methylase is to add a methyl group to the recognition sequence and thereby prevent endonuclease action, it is of interest to compare the properties of this methylated 12-mer with our analogue set. Clearly, methylation completely abolishes substrate activity with the endonuclease rather than reducing it to an extremely slow rate. The significance of the complete cleavage of all the substrate oligodeoxynucleotides (including those hydrolyzed at a very low rate) is twofold. First, it eliminates artifacts that may arise due to trace contamination of a nonsubstrate oligodeoxynucleotide with the parent 12-mer. Second, the product hexamers could be purified and their identities established by base composition analysis. The analysis of these hexameric products was always consistent with the *Eco*RV endonuclease cleaving between the central T and dA residues in the d(GATATC) recognition sequence as expected (Schildkraut et al., 1984; D'Arcy et al., 1985). In no case did the presence of a modified base change the cleavage position.

Interaction of the Oligodeoxynucleotides with the EcoRV Methylase. Table II shows the rates at which the methylase adds a CH_3 group to the oligodeoxynucleotides. Again, the set can be divided into nonsubstrates {oligodeoxynucleotides containing a dA analogue at the first position, d-(GACGAUATCGTC), and d(GACGA[^{4H}T]ATCGTC)} and substrates (all other 12-mers). The rates at which the substrates are methylated vary widely and can be much less, e.g., 0.14 nmol min^{-1} mg^{-1} for d(GACGATAUCGTC), or more, e.g, 74.2 nmol min^{-1} mg^{-1} for d(GACGATPTCGTC)} than the 35 nmol min^{-1} mg^{-1} observed with the parent 12-mer. As with the endonuclease these rates were obtained from HPLC traces of the time course of the reactions (data not shown). Once again with all the substrate oligodeoxynucleotides including the poorest d(GACGATAUCGTC) it was possible to observe 100% methylation providing long times and large quantities of the methylase were used. In contrast, no reaction was seen with the nonsubstrates, and on the basis of detection limits these react at <0.05% of the parent 12-mer. All the product oligodeoxynucleotides arising from methylation were

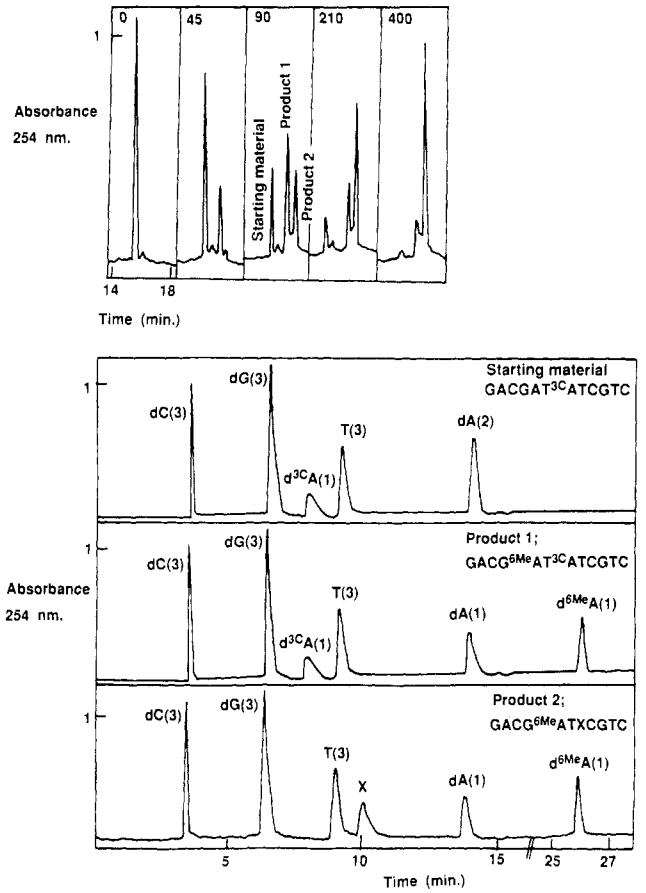

FIGURE 5: Time course of methylation of d(GACGAT[^{3C}A]TCGTC) by the *Eco*RV methylase. As can be seen, the starting material is converted initially to an intermediate product (product **1**). This is then converted to the final product (product **2**). The base composition analysis of the starting material and products **1** and **2** are also shown and indicate that product **1** arises from methylation of the dA residue in the d(GAT[^{3C}A]TC) sequence, whereas product **2** is due to a second methylation at d^{3C}A.

analyzed by base composition analysis. In each case (except one) reaction was accompanied by the production of 1 equiv of d^{6Me}A and loss of 1 equiv of dA. We have previously shown that the first dA residue within d(GATATC) is the target for the methylase (Nwosu et al., 1988). Although we have not strictly shown this for all the analogues, we have assumed that the production of d^{6Me}A with the stoichiometric loss of dA results from methylation at this site. An anomalous result was found with d(GACGAT[^{3C}A]TCGTC). The time course for its methylation is shown in Figure 5. It is obvious that an initial slightly later eluting product (product **1**) is first formed and that this disappears with the production of a final later eluting product (product **2**). We have purified both the intermediate and the final product and subjected them to base composition analysis as shown in Figure 5. For comparison this figure also gives the analysis of the starting material d(GACGAT[^{3C}A]TCGTC). It is clear that the initial product is d(GACG[^{6Me}A]T[^{3C}A]TCGTC) and arises from the normal methylase reaction, i.e., addition of a CH_3 group to the first dA in the recognition sequence. The final product is d(GACG[^{6Me}A]TXTCGTC). Clearly, d^{3C}A has disappeared at the expense of X. We suspect that this product arises from a second methylation of d^{3C}A to give X, which is presumably the *N*-6-methyl derivative of d^{3C}A (unfortunately, we do not have a standard of this compound to allow unambiguous determination). When d(GACG[^{6Me}A]T[^{3C}A]TCGTC) was incubated with the methylase, it also gave the dimethylated final product as would be expected. Thus, it appears that after addition of a CH_3 to its normal target dA, the methylase adds a further methyl group to the second dA in the target sequence when this position is occupied by d^{3C}A. This is the only case with both endonuclease and methylase where the presence of a base analogue affects the position of enzymatic action. Both methylations require the presence of the methylase and *S*-adenosylmethionine, and in particular neither takes place in the presence of SAM alone. This eliminates the possibility that especially the second unusual CH_3 group addition is due to the reactive SAM acting as a simple chemical methylating agent.

Discussion

The most important issue raised in this paper is, how disruptive are the supposedly conservative base analogues when incorporated into oligodeoxynucleotides? It is fairly easy to prepare free deoxynucleosides in which potential protein contacts are deleted and to do this in an apparently conservative manner. As mentioned previously, this involves simply removing the potential contact and avoiding the introduction of bulky groups into the analogue. Almost all the modified bases shown in Figure 2 meet these two criteria. The only exceptions are the sulfur analogues ^{2S}T and ^{4S}T. Even these are near conservative in greatly reducing potential protein contacting ability and being only slightly bulkier than the parents. When these analogues are incorporated into oligodeoxynucleotides, however, a new problem arises. To be a truly conservative base analogue, the derivative should not change the conformation of the oligodeoxynucleotide as compared to the parent which contains the usual four bases. This is a fundamental problem that plagues all analogue studies, be they substrate analogues, the oligodeoxynucleotide analogues mentioned here, or protein analogues produced by site-directed mutagenesis. It is very difficult in all these cases (and when faced with hardened devil's advocacy, probably impossible) to show conclusively that the analogue does not cause global conformational effects in addition to just altering the locus under study. The problem may be especially acute with oligodeoxynucleotides as it is becoming apparent that even with just the four normal bases the overall DNA structure shows some dependence on base sequence (Drew et al., 1988; Kennard & Hunter, 1988; Travers, 1989). Thus, DNA does not simply consist of an unvarying Watson–Crick B structure throughout its length but shows sequence-dependent differing conformations along its backbone. Therefore, there exists the possibility with our 12-mers that the presence of a modified base causes overall changes in the global structure. These changes would be expected to alter the sugar phosphate backbone and perhaps alter the relative disposition of neighboring bases. For a protein using just the direct readout mechanism this change in neighboring base orientations might lead to differential interaction with loci of the bases in addition to the one being probed. Proteins that use the direct readout mechanism also tend to make nonspecific contacts to the sugar phosphate backbone, and so a global conformational change would also be expected to disrupt some of those contacts. One would expect the problem to be particularly acute for a protein using the indirect readout mechanism and so recognizing a peculiar sugar phosphate arrangement. Here it would be particularly important for a base analogue not to change the overall structure. Fortunately, this mechanism appears to be the exception rather than the rule. A final consideration is that many proteins, for example, *Eco*RI (McClarin et al., 1986), DNase I (Suck et al., 1988), and 434 repressor (Aggarwal et al., 1988), cause distortions to DNA on binding. It

may be the case that only the recognition sequence is able to be correctly distorted on enzyme binding to form a tight enzyme substrate complex. At present we do not known if and how our base analogues affect the flexibility and distortability of the oligodeoxynucleotides and whether or not they can hinder the attainment of a peculiar enzyme bound conformation. What effects then do the modified bases have on the overall conformation of the oligodeoxynucleotides, and is this conformation changed with respect to the parent 12-mer? The T_m values obtained show that the entire set forms stable double-helical structures at 25 or 37 °C and that all the duplexes have similar thermal stability. This is important as both endonuclease and methylase require double-stranded forms. With the exception of ^{2S}T-containing oligomers, the CD spectra of the parent were almost identical with or very similar to the entire modified set and typical of B-DNA. We are at present not sure why ^{2S}T-containing oligodeoxynucleotides have odd CD spectra and whether or not they are conformationally different from the parent. Because of this, results obtained with this oligomer must be interpreted with caution. On the basis of the T_m and CD data we have assumed (with the exception of ^{2S}T) that the structures of the parent and the modified oligodeoxynucleotides are the same and are double-helical B-DNA. We realize that this is probably an oversimplification as T_m and CD are rather crude measures of overall conformation and at best the oligodeoxynucleotide set will be similar rather than identical. A corollary of this is that although we have interpreted the enzymatic data simply in terms of potential contact group deletion, we cannot entirely eliminate the possibility that some effects arise due to global conformational differences. At present we do not know whether or not the *Eco*RV enzymes distort DNA on binding. Information with the endonuclease should be forthcoming from crystallography of enzyme substrate complexes (Winkler et al., 1986). In the absence of this information we have ignored the possibility that some of our analogue-containing oligomers are poor substrates because they cannot achieve the required bound conformation. Once again, this is an oversimplification and requires some caution in the interpretation of our results. One final possibility is the formation of an intramolecular hairpin rather than a double-stranded duplex for the parent oligodeoxynucleotide or some of the analogues containing modified bases. Fully self-complementary oligodeoxynucleotides can form hairpins (Marky et al., 1983; Wemmer et al., 1985), but these forms are usually much less stable than the duplex and only found at low salt concentrations (<10 mM NaCl), high temperatures, and low oligodeoxynucleotide levels. With our parent dodecamer d(GACGATATCGTC) perhaps the best evidence that it exists as a duplex rather than a hairpin under the conditions we have used for enzymological studies (100 mM NaCl, 25 °C, 0.5–20 μM oligomer concentration) comes from the linear plot seen for its K_m and k_{cat} determination with the *Eco*RV endonuclease (following paper). Restriction endonucleases require double-stranded oligodeoxynucleotides as substrate as shown by X-ray crystallography for both *Eco*R1 (McClarin et al., 1986) and *Eco*RV (F. Winkler, personal communication). Any hairpin–duplex equilibrium is concentration dependent, and if this equilibrium was changing over the 0.5–20 μM range, the fraction of the active duplex form present would vary; i.e., the concentration of the duplex would not change in a simple linear fashion over the 0.5–20 μM levels. This would be expected to result in nonlinear plots for K_m and k_{cat} determination. There still exists the possibility that some of the analogue-containing oligodeoxynucleotides exist as hairpins, especially those containing bases that result in Watson–Crick hydrogen-bond loss (dP, ^{4H}T, and perhaps ^{4S}T). For most of our dodecamer set this would not seem to be a problem. First, wherever a modified base containing oligodeoxynucleotide was a substrate for the *Eco*RV endonuclease, linear plots were obtained when K_m and k_{cat} were evaluated (following paper). This includes d(GACGATPTCGTC) and d(GACGA[^{4S}T]ATCGTC). In addition, dP/T base pairs have been reported not to destabilize double-helical structures and to be perfectly compatible with B-DNA (Ikuta et al., 1987; Clore et al., 1988). The greatest chance of hairpin formation would seem to be when T is replaced by ^{4H}T as no hydrogen bonds to complementary dA residues are formed. Table II shows that although d(GACGATA[^{4H}T]CGTC) is not a substrate for the endonuclease, it is as good a substrate for the methylase as is the parent dodecamer. As this methylase requires double-stranded DNA, this would imply that the above dodecamer exists as mainly the double-stranded form. With d(GACGAT[^{4H}T]ATCGTC) both endonuclease and methylase are inactive. Here there does exist the possibility that this is due to formation of a nonsubstrate hairpin rather than a substrate duplex. We are currently investigating whether or not this oligonucleotide is a hairpin by performing T_m and CD measurements as a function of concentration.

Keeping the above provisos in mind, what do the results we have obtained tell us about the endonuclease and the methylase and how do these enzymes recognize the d(ATAT) hexamer in the d(GATATC) recognition sequence? At present we have only measured a single rate at 20 μM oligodeoxynucleotide concentration. Therefore, we have kept this discussion brief. Much more information is available from K_m and k_{cat} determination, and these are presented for the endonuclease in the subsequent paper. We have some but not yet all of these parameters for the methylase, and they will be reported when complete. Table II shows that for the endonuclease the presence of a modified base always slows down the rate of reaction. With the methylase a different pattern is seen in that some analogues slow the rate, whereas others cause an increase as compared to the parent. Although it is common for substrate analogues to slow reaction rates, an increase is rare. One possibility is that for the methylase product release is rate limiting and some oligodeoxynucleotides containing modified bases result in a speedup of this slowest step. This is currently under investigation. Considering each of the four bases within the d(ATAT) sequence in turn, it is clear that the first dA is extremely important for the activity of both the endonuclease and methylase. With the methylase zero rates are seen with all three analogue-containing oligodeoxynucleotides. As the 6-NH_2 group of this dA is the site of CH_3 addition, dP clearly cannot be a substrate. Nevertheless, the zero rates seen with d^{7C}A and d^{3C}A emphasize how critical this base is for methylase activity. With the endonuclease a zero rate is obtained with d^{7C}A and very low rates are obtained with dP and d^{3C}A. We therefore propose that both enzymes contact this dA in both the major groove (to the 6-NH_2 and 7-N groups) and the minor groove (to the 3-N group) via hydrogen bonds. Essentially similar results have previously been seen for d^{7C}A (Fliess et al., 1988) and dP (Mazzarelli et al., 1989) in this position. With the first T residue conversion to dU completely inhibits the methylase and gives a very poor substrate for the endonuclease. Once again, similar results have been previously seen with the endonuclease for the T to dU conversion (above references). We propose a van der Waals contact between both enzymes and this CH_3 group. The O4 carbonyl oxygen of this T is also important as when it is converted to hydrogen (T → ^{4H}T) both enzymes are completely inhibited and on conversion to sulfur (T → ^{4S}T)

poor substrates are obtained. Again the simplest explanation is that both proteins make a hydrogen bond to this position. A slight worry with ^{4H}T is that it cannot base pair with dA, and thus the local position of the ^{4H}T/dA base pair is changed relative to the normal T/dA. This could alter the position of other important functional groups (the 5-CH_3 of T and the 6-NH_2 and 7-N of dA) and be the cause of the inhibition seen rather than the loss of an interaction to the carbonyl O4. This cannot strictly be excluded, but the low activities seen with ^{4S}T (which can hydrogen bond with dA and is presumably less disruptive of overall structure) would seem to argue for an important contact at this position. Due to the peculiar CD spectrum seen with oligodeoxynucleotides containing ^{2S}T we would not like to draw any conclusions about the interaction of the two enzymes with this position; i.e., any changes of activity could be due to an overall conformational change in the oligodeoxynucleotides rather than to alteration in a specific contact to the 2-keto oxygen. With the second dA and the methylase both dP and d^{7C}A are excellent substrates. Both analogues involve changes to major groove contacts, and we therefore suggest that the methylase does not contact the 6-NH_2 and the 7-N groups of this second dA residue via the major groove. With the endonuclease dP is strongly inhibitory, whereas d^{7C}A is a reasonable substrate. It therefore appears that the endonuclease, in contrast to the methylase, makes a single hydrogen bond to the 6-NH_2 group of this dA via the major groove. Once again similar results have been reported for d^{7C}A (Fliess et al., 1988) and dP (Mazzarelli et al., 1989) at this position. Both proteins give similar slow rates with d^{3C}A, consistent with a hydrogen bond to the ring 3N via the minor groove. Removal of the 5-CH_3 group of the second T (T → dU) strongly inhibits both enzymes, suggesting an important van der Waals contact to this CH_3 group. Similar effects have previously been seen for this T → dU conversion (Fliess et al., 1986, 1988; Mazzarelli et al., 1989). The two enzymes show very different behavior to the 4-keto group of this T. With the methylase this function is unimportant as both the T → ^{4H}T and the T → ^{4S}T changes give excellent substrates. Furthermore, this result argues against a large conformational change for the ^{4H}T/dA base pair with ^{4H}T in this position. With the methylase all the potential contacts on the dA residue and the 5-CH_3 on the T are crucial for activity giving nonsubstrates or very poor substrates. If this ^{4H}T/dA base pair was perturbed relative to the parent T/dA, one would expect changes in the positions of these important enzyme contacts and as a result a poor substrate in contrast to the excellent one seen. With the endonuclease the position is vital as both ^{4H}T and ^{4S}T are nonsubstrates. We would propose a hydrogen bond between the endonuclease and this oxygen atom, although potential problems with ^{4H}T/dA base pairs that complicate this interpretation have been discussed above. Once again because of the CD spectrum seen with ^{2S}T we feel unable to draw any conclusions resulting from placing ^{2S}T at the second T site.

In the final conclusion Table II indicates that the endonuclease and methylase show behavior similar to but not identical with that of the dA and T analogue set. The first dA in T bases within the d(ATAT) sequence is very important to both proteins, and most changes result in poor substrates. With the second dA and T the behavior diverges somewhat with the methylase being, in general, more tolerant to analogues than the endonuclease.

Acknowledgments

We thank Dr. Alex Drake (Birkbeck College, London) for help with the CD spectra, Dr. Geoff Kneale (Portsmouth Polytechnic, Portsmouth) for assistance with the T_m determinations, and Dr. David Corina and Mrs. Joan Street (Southampton University) for running mass spectra and NMR spectra, respectively, during the preparation of modified bases. We are especially grateful to Dr. S. Halford (University of Bristol) for discussions and making unpublished data available and to Dr. F. Winkler (Hoffman-La Roche Ltd., Basel) for making available unpublished crystallographic data. We also thank Mrs. Sue Broomfield and Miss Clare Philo for typing the manuscript.

References

Aggarwal, A. K., Rogers, D., Drotter, M., Ptashne, M., & Harrison, S. C. (1988) *Science 242*, 899–907.

Anderson, J. E., Ptashne, M., & Harrison, S. C. (1987) *Nature 326*, 846–852.

Bennett, S. P., & Halford, S. E. (1989) *Curr. Top. Cell. Regul. 30*, 57–104.

Bougueleret, L., Schwartzstein, M., Tsugita, A., & Zabeau, M. (1984) *Nucleic Acids Res. 12*, 3654–3676.

Bougueleret, L., Tenchini, M. L., Botterman, J., & Zabeau, M. (1985) *Nucleic Acids Res. 13*, 3823–3839.

Brennan, C. A., VanCleve, M. D., & Gumport, R. I. (1986) *J. Biol. Chem. 261*, 846–852.

Brennan, R. G., & Matthews, B. W. (1989) *J. Biol. Chem. 264*, 1903–1906.

Chirikjian, J. G. (1987) *Restriction Endonucleases and Methylases*, Elsevier, New York.

Clore, G. M., Oschkinat, M., McLaughlin, L. W., Benseler, F., Happ, C. S., Happe, E., & Gronenborn, A. W. (1988) *Biochemistry 27*, 4185–4197.

Connolly, B. A., & Newman, P. (1989) *Nucleic Acids Res. 17*, 4957–4974.

D'Arcy, A., Brown, R. S., Zabeau, M., Van Resandt, R. W., & Winkler, F. K. (1985) *J. Biol. Chem. 260*, 1987–1990.

Drew, H. R., McCall, M. J., & Calladine, C. R. (1988) *Annu. Rev. Cell Biol. 4*, 1–20.

Dwyer-Hallquist, P., Kezdy, F. J., & Agarwal, K. L. (1982) *Biochemistry 21*, 4693–4700.

Faerber, P., & Scheit, K. H. (1970) *Chem. Ber. 103*, 1307–1311.

Fairall, L., Martin, S., & Rhodes, D. (1989) *EMBO J. 8*, 1809–1817.

Fasman, G. D. (1975) *CRC Handbook of Biochemistry and Molecular Biology*, 3rd ed., Nucleic Acids Vol. I, pp 65–213 and 419–449, CRC, Cleveland, OH.

Fersht, A. R. (1987) *Trends Biochem. Sci. 12*, 301–304.

Fisher, E. F., & Caruthers, M. H. (1979) *Nucleic Acids Res. 7*, 401–416.

Fliess, A., Wolfes, H., Rosenthal, A., Schwellus, K., Blocker, H., Frank, R., & Pinground, A. (1986) *Nucleic Acids Res. 14*, 3463–3474.

Fliess, A., Wolfes, H., Seela, F., & Pingoud, A. (1988) *Nucleic Acids Res. 16*, 11781–11793.

Fox, J. J., Praag, D. V., Wempen, I., Doerr, I. L., Cheong, L., Knoll, J. E., Eidinoff, M. L., Bendich, A., & Brown, G. B. (1959) *J. Am. Chem. Soc. 81*, 178–187.

Garnett, J., & Halford, S. E. (1988) *Gene 74*, 73–76.

Gildea, B. & McLaughlin, L. W. (1989) *Nucleic Acids Res. 12*, 4539–4557.

Goeddel, D. V., Yansura, D. G., Winston, C., & Caruthers, M. H. (1978) *J. Mol. Biol. 123*, 723–737.

Halford, S. E., & Goodall, A. J. (1988) *Biochemistry 27*, 1771–1777.

Halford, S. E., Loveday, B. M., & McCallum, S. A. (1986) *Gene 41*, 173–181.

Hayakawa, T., Ono, A., & Ueda, T. (1988) *Nucleic Acids Res. 16*, 4761–4776.
Ikuta, S., Eritja, R., Kaplan, B. E., & Itakura, K. C. (1987) *Biochemistry 26*, 5646–5650.
Ivanov, V. I., Minchenkova, L. E., Schyolkima, A. K., & Poletayer, A. I. (1973) *Biopolymers 12*, 89–110.
Jiricny, J., Wood, S. G., Martin, D., & Ubasawa, A. (1986) *Nucleic Acids Res. 14*, 6579–6590.
Jordan, S. R., & Pabo, C. O. (1988) *Science 242*, 893–899.
Kennard, O., & Hunter, W. (1988) *Landolt-Borstein Data New Series Group VII*, Vol. 1a, pp 255–360, Springer-Verlag, Berlin.
Luke, P. A., McCallum, S. A., & Halford, S. E. (1987) *Gene Amplif. Anal. 5*, 183–205.
Matthews, B. W. (1988) *Nature 335*, 294–295.
Mazzarelli, J., Scholtissek, S., & McLaughlin, L. W. (1989) *Biochemistry 28*, 4616–4622.
McClarin, J. A., Fredrick, C. A., Wang, B. C., Greene, P., Boyer, H. W., Grable, J., & Rosenberg, J. M. (1986) *Science 234*, 1526–1541.
McLaughlin, L. W., Benseler, F., Greaser, E., Piel, N., & Scholtissek, S. (1987) *Biochemistry 26*, 7238–7245.
Modrich, P. (1982) *CRC Crit. Rev. Biochem. 13*, 287–323.
Nair, V., & Chamberlain, S. D. (1984) *Synthesis*, 401–403.
Nwosu, V. U., Connolly, B. A., Halford, S. E., & Garnett, J. (1988) *Nucleic Acids Res. 16*, 3705–3720.
Ohlendorf, D. H., & Matthews, B. W. (1983) *Annu. Rev. Biophys. Bioeng. 12*, 259–284.
Ono, A., & Ueda, T. (1987) *Nucleic Acids Res. 15*, 3059–3072.
Ono, A., Sato, M., Ohtani, Y., & Ueda, T. (1984) *Nucleic Acids Res. 12*, 8939–8947.
Ott, J., Eckstein, F., & Connolly, B. A. (1985) *Nucleic Acids Res. 13*, 6317–6329.
Otwinowski, Z., Schevitz, R. W., Zheng, R. G., Lawson, C., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., & Sigler, P. B. (1988) *Nature 335*, 321–329.
Pabo, C. O., & Sauer, R. T. (1984) *Annu. Rev. Biochem. 53*, 293–321.
Saenger, W. (1985) in *Principles of Nucleic Acid Structure*, p 371, Springer-Verlag, Berlin.
Schildkraut, I., Banner, C. D., Rhodes, C. S., & Parekh, S. (1984) *Gene 27*, 327–329.
Seela, F., & Driller, H. (1986) *Nucleic Acids Res. 14*, 2319–2332.
Seela, F., & Kehne, A. (1983) *Justus Liebigs Ann. Chem.*, 876–884.
Seela, F., & Kehne, A. (1987) *Biochemistry 26*, 2232–2238.
Seeman, N. C., Rosenberg, J. M., & Rich, A. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 804–808.
Such, D., Lahm, A., & Oefner, C. (1988) *Nature 332*, 465–468.
Taylor, J. D., & Halford, S. E. (1989) *Biochemistry 28*, 6198–9207.
Travers, A. A. (1988) *Annu. Rev. Biochem. 58*, 427–452.
Wightman, R., & Holy, A. (1973) *Collect. Czech. Chem. Commun. 38*, 1381–1397.
Winkler, F. K., Brown, R. S., Leonard, K., & Berriman, J. (1986) *Crystallography and Molecular Biology*, NATO Advanced Study Institute, pp 345–352, Plenum Press, New York.
Wohlberger, C., Dang, Y., Ptashne, M., & Harrison, S. C. (1988) *Nature 35*, 789–795.
Yansura, D. G., Goeddel, D. V., Kundu, A., & Caruthers, M. H. (1977) *Nucleic Acids Res. 4*, 723–737.
Yolov, A. A., Vinogradova, M. N., Gromova, E. S., Rosenthal, A., Cech, D., Veiko, V. P., Metelev, V. G., Kosyhyk, V. A., Buryanov, Y. I., Bayeu, A. A., & Shaborova, Z. A. (1985) *Nucleic Acids Res. 13*, 8983–8998.