DNA Binding Mechanism of O^6-Alkylguanine-DNA Alkyltransferase: Stoichiometry and Effects of DNA Base Composition and Secondary Structure on Complex Stability†

Michael G. Fried,*,‡ Sreenivas Kanugula,§ Jennifer L. Bromberg,‡ and Anthony E. Pegg§

Departments of Biochemistry and Molecular Biology and Physiology, The Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033

Received April 22, 1996; Revised Manuscript Received August 21, 1996⊗

ABSTRACT: O^6-Alkylguanine-DNA alkyltransferase (AGT) is an important cellular defense against the mutagenic effects of DNA alkylating agents. In humans this defense can contribute to the ability of some tumors to resist the effects of chemotherapeutic agents that act through DNA alkylation. We report here studies that characterize the interaction of AGT with DNA. We show that although AGT sediments as a monomer in the absence of DNA, it binds cooperatively to single stranded deoxyribonucleotides. The stoichiometries of complexes formed with 16-, 30-, and 80-base oligodeoxyribonucleotides are 3.8 ± 0.3, 5.3 ± 0.2, and 8.9 ± 0.2, respectively; the binding density decreasing from ∼4 nt/monomer to ∼9 nt/monomer as DNA length increases over this range. Binding competition assays show that DNA affinities depend only weakly on base composition or secondary structure, although in general G+C-rich sequences are bound with greater affinity than are A+T-rich ones and single-stranded DNA is bound with greater affinity than duplex forms. These results suggest mechanisms by which AGT may search for alkylated sites and interact with them to effect DNA repair.

O^6-Alkylguanine and O^4-alkylthymine are strongly mutagenic adducts that are formed in DNA as a result of exposure to alkylating agents (Loveless, 1969; Pegg & Singer, 1984; Pegg *et al*., 1995). Both lesions are repaired by O^6-alkylguanine-DNA alkyltransferase (AGT; EC 2.1.1.63). This protein, ubiquitous in bacteria and mammals (Lindahl *et al*., 1988; Pegg, 1990), catalyzes the stoichiometric, covalent transfer of the alkyl group to a cysteine residue within its own active site. This reaction returns the DNA base to an unmodified state and permanently inactivates the "enzyme". Therefore, the number of O^6-alkylguanine and O^4-alkylthymine adducts that can be repaired at one time depends on the steady-state cellular concentration of AGT (Pegg, 1990; Mitra & Kaina, 1993; Dolan *et al*., 1990) and on its ability to partition between relatively infrequent alkylated sites and the vast amount of structurally similar genomic DNA in which they are embedded.

In spite of the important roles played by AGT, remarkably little is known about its mechanisms of interaction with DNA. In the crystal structure of the C-terminal fragment of the homologous *Escherichia coli ada* protein, the active site cysteine 146 is deeply buried in a cleft that is far too small to accommodate a DNA strand (Moore *et al*., 1994). Thus, a conformational rearrangement must occur in order to bring this cysteine into contact with the surface of the DNA. In support of this notion, DNA-dependent conformational changes have been detected for both *ada* protein and human AGT (Takahashi *et al*., 1990; Chan *et al*., 1993). Data that we present below suggest that a DNA-mediated conformational change promotes interaction between AGT molecules, resulting in cooperative DNA binding. In addition, both the sequences flanking an O^6-alkylguanine adduct and their secondary state influence the repair rates of bacterial and human proteins (Dolan *et al*., 1988; Liem *et al*., 1993), although their roles in AGT binding remain to be discovered. Here we present evidence that AGT binds cooperatively to single stranded DNAs, although no protein-protein interaction is detectable in free solution. The complex stoichiometry changes with the length of the DNA tested, suggesting that AGT possesses more than one mode of DNA binding. We show that the protein binds with greater affinity to single stranded than to duplex DNAs and hence must function as a DNA melting protein. Finally, our results indicate that AGT binds with greater affinity to CG-rich sequences than to AT-rich ones, and that while there are sites present in calf genomic DNA that are bound with unusually low affinity, the range of binding affinities is small. This indicates that AGT possesses only weak ability to discriminate among competing genomic sites. As discussed below, these results are likely to have significant implications for the mechanism of AGT function.

MATERIALS AND METHODS

Reagents. Restriction endonuclease *Bam*H1 and T_4 polynucleotide kinase were from New England Biolabs. [γ-^{32}P]-ATP was from DuPont-New England Nuclear. Acrylamide and *N*,*N*′-methylenebisacrylamide were from Aldrich.

AGT. Recombinant human AGT was prepared as previously described (Pegg *et al*., 1993). The protein was homogeneous as judged by electrophoresis and was 100% active in debenzoylating O^6-benzylguanine. AGT concentrations were measured spectrophotometrically using a molar extinction coefficient, $\epsilon_{280} = 3.93 \times 10^4$ M^{-1} cm^{-1},

† Supported by NSF Grant DMB 91-48816 (to M.G.F.) and by NIH Grant CA-18137 (to A.E.P.).

* Corresponding author. Tel: (717) 531-5250. FAX: (717) 531-7072. E-mail: mfried@cor-mail.biochem.hmc.psu.edu.

‡ Department of Biochemistry and Molecular Biology.

§ Department of Physiology.

⊗ Abstract published in *Advance ACS Abstracts,* November 1, 1996.

S0006-2960(96)00971-3 CCC: $12.00

Table 1: Oligodeoxyribonucleotides Used in This Study

oligo 1:
5′-GAC TGA CTG ACT GAC T-3′

oligo 2:
5′-GTG CCG CCA ACC CGC TGC CTA TCG TTA TAC-3′

oligo 3:
5′-GTG GTC TGC AGC AGC GGA GCC GTG GGC AAC
TAC CGC TGG GGC GTG TCG CGT AAG GAA TGG
CTT CTG GCC CAT GAA GGC-3′

calculated from data of Roy *et al*. (1995). Values of $\epsilon_{215}/\epsilon_{280} = 8.2$ and $\epsilon_{260}/\epsilon_{280} = 0.63$ were obtained from UV spectra; extinction coefficients $\epsilon_{215} = 3.2 \times 10^5$ M^{-1} cm^{-1} and $\epsilon_{260} = 2.5 \times 10^4$ M^{-1} cm^{-1} were calculated from these ratios.

DNA. Poly d(A-T), poly dA-poly dT, poly d(G-C), and poly dG-poly dC were obtained from Sigma. Samples were extensively dialyzed at 4 °C against 10 mM Tris (pH 8.0), 1 mM EDTA, and used without further purification. On heating, samples from these preparations gave hyperchromicities in the range of 1.30−1.35 at 260 nm, indicating significant duplex content. Calf thymus DNA was obtained from Sigma and sonicated and fractionated as described (Fried & Bloomfield, 1984). Plasmid pBR322 (Sutcliffe, 1978b) was propagated in *E. coli* DH5α. Supercoiled DNA was purified by CsCl density gradient centrifugation. The linear form was obtained by digestion with endonuclease *Bam*H1. Single-stranded plasmid DNA was obtained by heating to 95 °C for 5 min followed by quenching on ice, immediately prior to use. Oligodeoxyribonucleotides containing 16, 30, and 80 residues were synthesized in the institutional biomolecular core facility and purified by NENsorb chromatography. Their sequences are given in Table 1. Where appropriate, DNA samples were labeled at 5′ termini with ^{32}P as described (Maxam & Gilbert, 1977), ethanol precipitated, and resuspended in 10 mM Tris (pH 8.0), 1 mM EDTA, prior to use. Stock DNA concentrations were measured spectrophotometrically, using $\epsilon_{260} = 1.3 \times 10^4$ M^{-1} cm^{-1} (per base pair) for duplex samples and $\epsilon_{260} = 1.04 \times 10^4$ M^{-1} cm^{-1} (per base) for single-stranded samples.

Electrophoretic Analysis of Protein−DNA Complexes. Binding reactions were carried out at 21 (±1) °C in 10 mM Tris (pH 7.6), 1 mM dithiothreitol, 1 mM EDTA, 10 μg/mL bovine serum albumin or in that buffer supplemented with 100 mM NaCl, as indicated. Protein−DNA complexes were formed by adding appropriate amounts of AGT to solutions containing ^{32}P-labeled oligodeoxyribonucleotides. Mixtures were equilibrated at 21 (±1) °C for 2 h. In binding competition experiments, competing DNA was directly added to tubes containing protein-DNA complexes, mixed gently and allowed to equilibrate at 21 (±1) °C for 2 h. Duplicate samples incubated for longer periods gave identical results, indicating that equilibrium had been attained. Electrophoresis was performed using 10% polyacrylamide gels, cast and run at 8 V/cm in buffer consisting of 10 mM Tris-acetate, 1 mM EDTA, pH 7.6. Autoradiograms were obtained with DuPont Cronex film, exposed at 4 °C. Gel segments containing individual electrophoretic species were excised using the developed film as a guide and counted in a scintillation counter by the Cerenkov method.

Complex stoichiometries and association constants were estimated from mobility shift assay data as follows. The binding of n molecules of AGT protein (P) to one molecule of 16-mer oligodeoxyribonucleotide (D) may be represented by $nP + D \rightleftarrows P_nD$. For this reaction the association constant is $K_n = [P_nD]/[P]^n[D]$; separating variables and taking logarithms gives

$$\ln\frac{[P_nD]}{[D]} = n \ln[P] + \ln K_n \quad (1)$$

A graph of $\ln[P_nD]/[D]$ as a function of $\ln[P]$ has a slope equal to the protein stoichiometry n; at half-saturation, $\ln[P_nD]/[D] = 0$ and the association constant can be evaluated from $\ln K_n = -n \ln[P]$. The ratio $[P_nD]/[D]$ was varied by serial dilution of AGT−DNA mixtures. Self-consistent values of n and K_n were obtained by calculating [P] using

$$[P] = [P]_o - n[P_nD] \quad (2)$$

in which $[P]_o$ is the total AGT concentration in the reaction mixture and an initial value of $n = 4$ was used on the basis of sedimentation equilibrium results. A new value of n was estimated from the slope of a plot of $\ln[P_nD]/[D]$ *versus* $\ln[P]$ about $\ln[P_nD]/[D] = 0$; this was used to calculate [P] using eq 2; this cycle was iterated until values of n converged.

Relative binding affinities were measured by binding competition assay, performed as follows. Addition of competing DNA (C) to a reference AGT−DNA complex (P_nD) establishes the equilibrium

$$P_nD + C \rightleftarrows D + P_nC \quad (3)$$

The distribution of AGT between reference and competitor DNAs depends on the ratio of equilibrium constants

$$\frac{K_{ref}}{K_{comp}} = \frac{[P_nD][C]}{[D][P_nC]} \quad (4)$$

Here $K_{ref} = [P_nD]/[P]^n[D]$, the equilibrium constant for formation of the initial AGT−DNA complex and $K_{comp} = [P_nC]/[P]^n[C]$, the equilibrium constant for formation of the complex with competitor. The concentrations $[P_nD]$, [D], and $[P_nC]$ are readily calculated from band intensity data. The free competitor concentration, $[C] = [C]_o - mn[P_nC]$, in which $[C]_o$ is its input concentration, m is the number of bases (or base pairs) of competitor occupied by each protein, and the other terms are defined as above.

Sedimentation Equilibrium. AGT protein and oligodeoxyribonucleotides were dialyzed against 10 mM Tris (pH 7.6), 1 mM DTT, 1 mM EDTA, 100 mM NaCl. Analytical ultracentrifugation was performed at 20 °C in a Beckman XL-A centrifuge equipped with absorbance optics, using an AN 60 Ti rotor, six-sector charcoal-epon centerpieces, and fused silica windows. Scans were obtained at 215, 260, and 280 nm with a step size of 0.001 cm. Equilibrium was taken to be attained when scans made 24 h apart were indistinguishable. Typically, equilibration times ≥48 h met this criterion for AGT−DNA mixtures. Five scans were averaged for each sample at each wavelength and rotor speed. For analysis of AGT protein alone, models incorporating different assembly stoichiometries were based on the general equation

$$A(r) = \sum_n \alpha_P{}^n K_n \exp[n\sigma_P(r^2 - r_o{}^2)] + \epsilon \qquad (5)$$

Here $A(r)$ is the absorbance at radial position r, α_P is the monomer absorbance at the reference radius (r_o), and n is the protein stoichiometry of each reaction component. The parameter σ_P is the reduced molecular weight [$\sigma_P = M_P(1 - \bar{v}\rho)\omega^2/(2RT)$], M_P is the protein molecular weight, $\bar{v}$ its partial specific volume, ρ the solvent density, ω the rotor angular velocity, R the gas constant, T the absolute temperature, and ϵ is the base line offset. The term K_n is the association constant for the multimerization reaction nAGT $\rightleftarrows$ AGT$_n$. Solvent density (1.004 g/mL) was measured using a Mettler density meter. The partial specific volume of AGT was calculated by the method of Cohn and Edsall (1943). Using the partial specific volumes of amino acids tabulated by Laue and co-workers (1992), a value of 0.744 was obtained.

Equation 6 gives the absorbance distribution at sedimentation equilibrium for a system in which protein−DNA complex formation depends on the local concentrations of macromolecules throughout the cell according to the equilibrium nP + D $\rightleftarrows$ P$_n$D.

$$A(r) = \alpha_P \exp[\sigma_P(r^2 - r_o{}^2)] + \alpha_D \exp[\sigma_D(r^2 - r_o{}^2)] + \alpha_{PD} \exp[\sigma_{PD}(r^2 - r_o{}^2)] + \epsilon \qquad (6)$$

Here, most terms are defined as in eq 5; α_D and α_{PD} are absorbances of DNA and protein−DNA complex at r_o, and the reduced molecular weights of DNA and protein−DNA complex are given by $\sigma_D = M_D(1 - \overline{v_D}\rho)\omega^2/(2RT)$ and $\sigma_{PD} = (nM_P + M_D)(1 - \overline{v_{P_nD}}\rho)\omega^2/(2RT)$ and n is the protein:DNA ratio of the complex. In this analysis, the molecular weights of recombinant AGT protein (M_r = 21 519) and DNAs (M_r = 4881, 9079, and 24 860 for oligos 1−3, respectively) were used as constants. The partial specific volume of NaDNA at 0.1 M NaCl (0.502 mL/g) was estimated by interpolation of the data of Cohen and Eisenberg (1968). Partial specific volumes of each of the possible protein−DNA complexes were estimated using

$$\overline{v_{PnD}} = (nM_P\overline{v_P} + M_D\overline{v_D})/(nM_P + M_D) \qquad (7)$$

where n is the stoichiometric ratio of protein to DNA in the complex. Equation 7 implicitly assumes that the partial specific volume of a complex is a weight-average of the partial specific volumes of the constituent species and thus that there is no significant volume change upon association. Although we do not know whether such a volume change occurs, it seems reasonable that the partial specific volumes of complexes containing a large mass-proportion of protein (like those analyzed here) should have partial specific volumes that reflect that proportion. Data sets obtained for samples run at three different rotor speeds and three input concentrations of AGT and DNA were fit simultaneously to eq 6 using the global fitting approach first described by Johnson and co-workers (1981). In this method, the values of α_P, α_D, α_{PD}, and ϵ are unique for each cell but the value of n must be common to all data sets. Nonideality was not considered, as there was no evidence of nonideal effects.

RESULTS

Native AGT Protein Is a Monomer in Solution. AGT protein was sedimented to equilibrium from solutions containing 13, 6.5, and 1.2 μM protein at three different centrifuge speeds (15 000, 20 000, and 40 000 rpm), to determine its degree of self-association in the absence of DNA. A representative data set for a sample with nominal AGT concentration of 13 μM is shown in Figure 1. The model in which only monomer AGT is present in solution (eq 5 with $n = 1$, $K_n = 1$) was fit to the data, giving M_r = 21 350 ± 400, in excellent agreement with the monomer molecular weight derived from the protein sequence (M_r = 21 519). The small, uniformly distributed residuals indicate that the monomer model is consistent with the data over the entire concentration range present in the centrifuge cell. Extension of the model to include oligomers of AGT (eq 1 with $n > 1$ and $K_n \neq 0$) did not improve the quality of the fit as judged by the correlation coefficient or the magnitude of the residuals (result not shown). The range of rotor speeds, protein concentrations, and wavelengths examined gave equilibrium distributions spanning the concentration range 0.25−38 μM. Over this range, no systematic deviation from the monomer molecular weight could be detected. Previously reported sedimentation velocity and gel permeation chromatography results (Bhattacharyya *et al.*, 1990) are consistent with the presence of monomeric AGT under highly dilute conditions, but do not rule out the formation of oligomers at higher concentrations. The results shown here indicate that under our buffer conditions, AGT is rigorously monomeric at concentrations ≤38 μM.

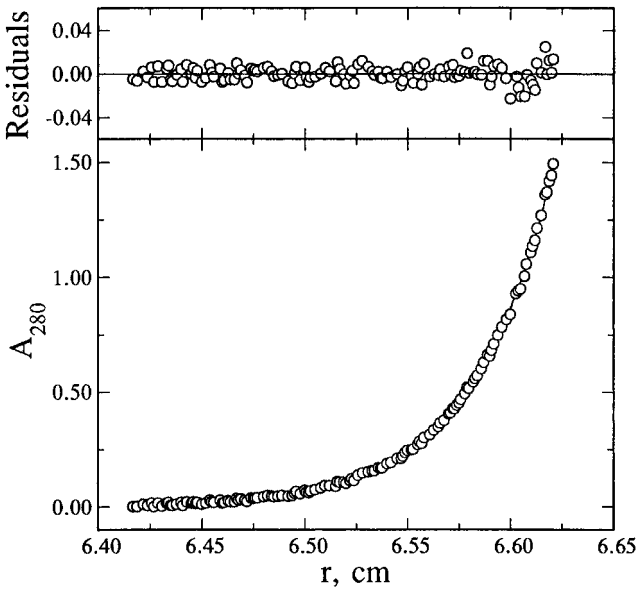

FIGURE 1: Sedimentation equilibrium of AGT at 20 ± 0.1 °C. Representative data taken at 40 000 rpm. The initial concentration of protein was 13 μM, in buffer consisting of 10 mM Tris (pH 7.6), 1 mM DTT, 1 mM EDTA, 100 mM NaCl. The data were fit to eq 5 with monomer molecular weight M as the adjustable parameter and both complex stoichiometry n and equilibrium constant K_n set equal to 1. The line through the data represents the global best fit obtained for data sets taken at 15 000, 20 000, and 40 000 rpm. This fit gave M_r = 21 350 ± 400, which is in good agreement with the monomer molecular weight derived from sequence data (21 519). Including an assembly stoichiometry n and equilibrium constant for assembly K_n as adjustable parameters did not improve the quality of the fit.

Sedimentation Equilibrium Analysis of AGT Interactions with 16-, 30- and 80-Residue Oligodeoxyribonucleotides. Addition of the 16-residue oligo 1 (Table 1) to solutions containing AGT resulted in the formation of a single protein−DNA complex as visualized by gel mobility shift

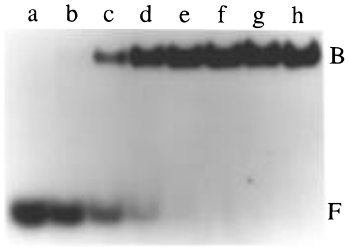

FIGURE 2: Titration of oligo 1 DNA with AGT. Analysis by gel electrophoresis mobility shift assay. Binding was carried out at 20 ± 1 °C in 10 mM Tris (pH 7.6), 1 mM DTT, 1 mM EDTA. All samples contained 5.1×10^{-7} M oligo 1; samples in lanes b−h contained AGT protein at final concentrations of 2.3×10^{-6}, 4.7×10^{-6}, 7×10^{-6}, 9.3×10^{-6}, 1.2×10^{-5}, 1.4×10^{-5}, and 1.6×10^{-5} M, respectively. Electrophoresis was carried out as described in Materials and Methods. Band designations: B, bound DNA; F, free DNA.

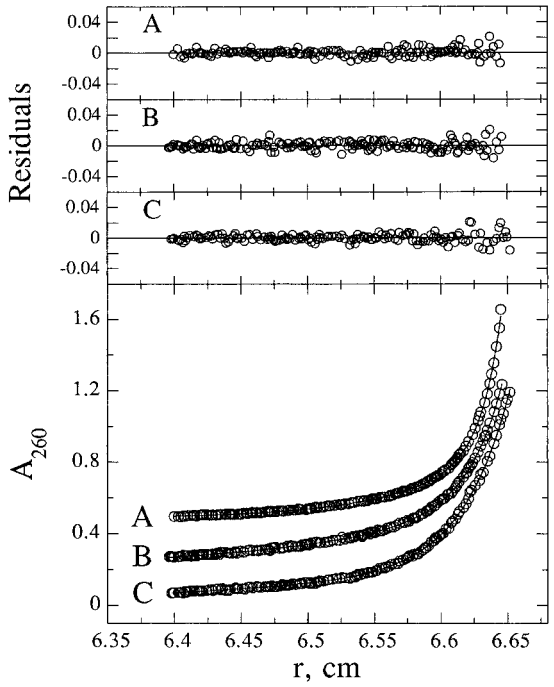

FIGURE 3: Sedimentation equilibrium of AGT−DNA mixtures at 20 ± 0.1 °C. Representative data taken at 20 000 rpm. All samples contained AGT (3.2 μM) in buffer consisting of 10 mM Tris (pH 7.6), 1 mM DTT, 1 mM EDTA, 100 mM NaCl. Samples A, B, and C contained, in addition, oligo 3 (0.1 μM), oligo 2 (0.5 μM), and oligo 1 (0.4 μM), respectively. The data were fit to eq 6 as described in Methods. The solid lines represent global fits to data sets data sets taken at 10 000, 15 000, and 20 000 rpm. Stoichiometric ratios for the AGT−DNA complexes were estimated from these fits. Curve A (80-mer), AGT:DNA = 8.9 ± 0.2; curve B (30-mer), AGT:DNA = 5.3 ± 0.2; curve C (16-mer), AGT:DNA = 3.8 ± 0.3.

assay (Figure 2). Prolonged film exposure did not reveal the presence of additional complex bands of either higher or lower mobility. Similar results were obtained with 30- and 80-residue DNAs (oligos 2 and 3; result not shown). Accordingly, sedimentation equilibrium data sets for AGT-DNA mixtures were analyzed in terms of the mechanism $nP + D \rightleftarrows P_nD$, in which complexes containing fewer than n monomers of AGT are not present to a significant degree. Shown in Figure 3 are representative data sets acquired at 20 000 rpm and 20 °C for mixtures of AGT with oligos 1, 2, and 3 (curves C, B, and A, respectively). Equilibrium data were taken at several different rotor speeds (see above) and the best fit of eq 6 to the combined data for each sample

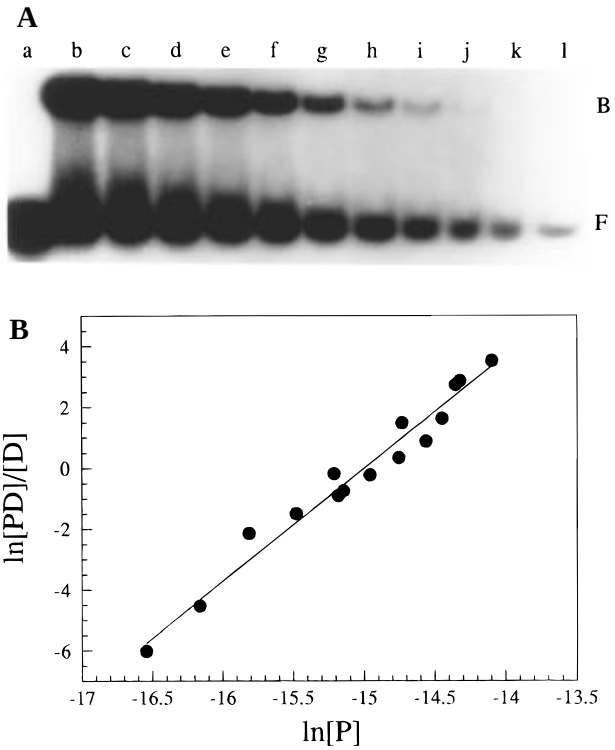

FIGURE 4: Serial dilution analysis of the AGT-oligo 1 interaction. A. Electrophoresis mobility shift assay. The initial mixture contained AGT (5.8 μM) and oligo 1 (1 μM) in 10 mM Tris (pH 7.6), 1 mM dithiothreitol, 1 mM EDTA, 10 μg/mL bovine serum albumin (lane b). Serial dilutions (0.66/step) were performed with the same buffer (lanes c−l) giving final DNA concentrations of 0.67, 0.45, 0.29, 0.19, and 0.13 μM and 88.2, 58.2, 39.2, 26.1, and 17.4 nM, respectively. Lane a contains DNA only (1 μM). Band designations: B, bound DNA; F, free DNA. B. Dependence of ln([PD]/[D]) on ln[P]. Data of A. The solid line is the linear least-squares fit to the data, giving $n = 3.8 \pm 0.2$ and $K_{PD} = 1.8\ (\pm 0.7) \times 10^{24}$ M^{-4}.

was obtained by least squares analysis. The data are consistent with stoichiometries of 3.8 ± 0.3, 5.3 ± 0.2, and 8.9 ± 0.2 in AGT complexes with oligos 1, 2, and 3, respectively. The small, uniformly distributed residuals indicate that protein stoichiometries of ∼4, ∼5, and ∼9 are consistent with the data for 16-, 30-, and 80-residue DNAs over the concentration ranges present and are a strong argument against the presence of AGT−DNA complexes with substantially different protein content in each reaction mixture.

Analysis of K_n and Complex Stoichiometry by Gel Mobility Shift Assay. Gel mobility shift assays were performed to explore the range of AGT and DNA concentrations below those accessible in the analytical ultracentrifuge. As shown in Figure 2, titration of the 16-mer oligo 1 with AGT produces a single complex at all AGT:DNA ratios tested, consistent with the simple binding mechanism $nP + D \rightleftarrows P_nD$. The ratio $[P_nD]/[D]$ varies with serial dilution (Figure 4A) allowing estimation of the complex stoichiometry n and association constant K_{PD} using eqs 1 and 2. As shown in Figure 4B, the stoichiometry value $n = 3.8 \pm 0.2$ is most consistent with the data. This agrees well with the value obtained by sedimentation equilibrium, in spite of a difference in the salt concentration of the buffer (serial dilution assays, ∼10 mM; sedimentation equilibrium assays, ∼110 mM). Our inability to detect higher stoichiometry complexes either with the highly sensitive mobility shift assay or at the

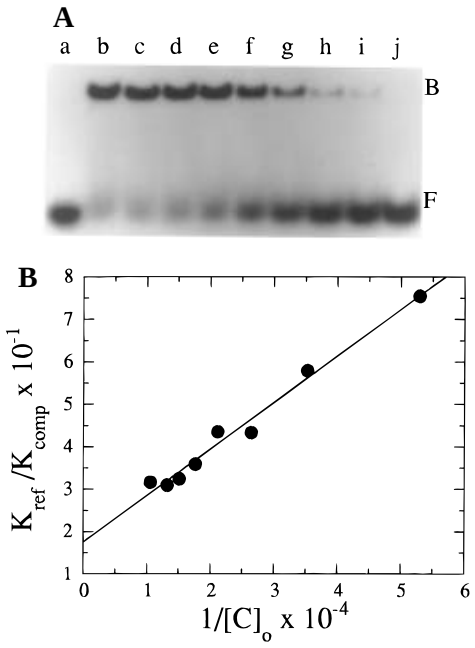

FIGURE 5: Binding competition assay. A. Electrophoresis mobility shift assay. Samples contained oligo 1 (0.4 μM) in 10 mM Tris (pH 7.6), 1 mM dithiothreitol, 1 mM EDTA, 10 μg/mL bovine serum albumin, and samples b−j contained, in addition, AGT (3.4 μM) and supercoiled pBR322 DNA at final base pair concentrations of 0, 1.9, 2.8, 3.8, 4.7, 5.7, 6.6, 7.6, and 9.4 μM, respectively. Incubation was at 20 ± 1 °C. Band designations: B, bound DNA; F, free DNA. B. Graph of K_{ref}/K_{comp} as a function of $1/[C_o]$ for the AGT−oligo 1 complex. Data of A.

high protein concentrations obtained at sedimentation equilibrium is consistent with the notion that oligo 1 is protein-saturated in the 4:1 complex. Since this DNA contains 16 residues, this stoichiometry implies a ratio of ~4 nucleotides/AGT monomer, a value significantly smaller than the optimum site size found for the rate of methyl-group transfer from oligonucleotides (7 nt; Liem *et al*., 1993) or that estimated for binding to high molecular weight duplex DNA as ligand (8 bp; Chan *et al*., 1993). Since AGT sediments as a monomer, the absence of electrophoretically resolved complexes with protein:DNA stoichiometries <4 at low protein:DNA ratios strongly suggests that the protein binds cooperatively to single-stranded DNA. This contrasts with an earlier observation that the homologous *E. coli ada* protein binds DNA without cooperativity (Takahashi *et al*., 1990). Serial dilution analysis also yields an estimate of the formation constant of the complex, $K_{PD} = 1.8\ (\pm 0.7) \times 10^{24}$ M^{-4}. Assuming equipartition of the binding free energies among four AGT monomers, this formation constant corresponds to a monomer association constant of $\sim 1.2 \times 10^{6}$ M^{-1}, in reasonable agreement with an estimate obtained for high molecular weight duplex DNA (Chan *et al*., 1993).

AGT Preferentially Binds Single-Stranded DNA. Binding competition assays provide a measure of the relative affinities of a protein for two DNAs (Fried & Crothers, 1981; Fried, 1989). If the 16-mer oligodeoxyribonucleotide is taken as a reference ligand, the relative affinities of the protein for different competitors can be compared. As a first step, we chose superhelical, linear, and single-stranded forms of pBR322 DNA as competitors. A representative competition assay is shown in Figure 5A. Relatively small molar excesses of competitor ($<10^3$) are sufficient for dissociation of AGT from the reference DNA. Since competitor saturation is significant under these conditions, neighbor exclusion should reduce the effective affinity for competitor, resulting in [competitor]-dependent values of K_{ref}/K_{comp} that are greater than those that would occur in the absence of neighbor exclusion. Evidence for this effect can be seen in Figure 5B. To minimize this complication, we have extrapolated K_{ref}/K_{comp} to $1/[C]_o = 0$, equivalent to infinite [competitor] (Hudson & Fried, 1990; Vossen *et al*., 1996), obtaining the limiting values of K_{ref}/K_{comp} given in Table 2. Although values of K_{ref}/K_{comp} differ for linear and supercoiled plasmid DNAs, both are bound significantly less tightly than the single-stranded 16-mer. In addition, single-stranded DNA is a more effective competitor than the duplex form. Finally, AGT binds more tightly to superhelical competitor than to the linear form. As discussed below, these results indicate that AGT binds single-stranded DNA more tightly than the duplex form. Such preferential binding is a characteristic of DNA melting proteins (Jensen & von Hippel, 1976; von Hippel *et al*., 1982).

Table 2: Relative Affinities of AGT for the 16-Mer Oligodeoxyribonucleotide with Respect to Various Competitors[a]

competitor	K_{ref}/K_{comp}
pBR322 (sc)[b]	17.5 ± 2.8
pBR322 (li)[c]	35.7 ± 4.6
pBR322 (ss)[d]	6.4 ± 2.2
calf thymus	312.1 ± 105.3
poly d(A-T)	105.0 ± 10.2
poly dA - poly dT	128.5 ± 41.0
poly d(G-C)	12.3 ± 2.1
poly dG - poly dC	15.0 ± 7.8

[a] Reactions performed at 21 (±1) °C in 10 mM Tris (pH 7.6), 1 mM dithiothreitol, 1 mM EDTA, 10 μg/mL bovine serum albumin. [b] Supercoiled form. [c] Linear form. [d] Single-stranded form.

Base-Composition and Sequence-Dependence of AGT Binding to DNA. In a second series of competition assays we used poly d(A-T), poly dA-poly dT, poly d(G-C), poly dG-poly dC, and sonicated calf thymus (CT) DNA as competitors, obtaining the hierarchy $K_{polyd(G-C)} \approx K_{polydG-polydC} > K_{polyd(A-T)} \approx K_{polydA-polydT} > K_{CT}$ (Table 2). Although AGT can remove O^4-alkyl adducts from thymine residues, removal of O^6-adducts from guanine residues has been postulated to be its principal role (Lindahl *et al*., 1988; Pegg 1990; Zak *et al*., 1994). The preferential binding to G+C-rich DNA detected here is consistent with that suggestion. Intriguingly, sonicated calf thymus DNA was the least effective competitor tested. This result suggests the existence of sequences in mammalian genomic DNA that are bound even less tightly than the sequence populations present in the other competitors that were tested. In spite of these differences, the small range in relative binding affinity detected in these assays (~50-fold) argues that AGT is capable of only limited sequence discrimination. The rate of guanine dealkylation exhibits a similarly weak dependence on the DNA sequence flanking the affected base (Dolan *et al*., 1988; Liem *et al*., 1993). Such low sequence specificity is appropriate for a protein that must repair DNA lesions in a wide variety of sequence contexts.

DISCUSSION

O^6-Alkylguanine-DNA alkyltransferase protects DNA genomes from the mutagenic effects of O^6-guanine and O^4-thymine alkylation. In addition, it has been implicated as

an important source of tumor resistance to chemotherapeutic alkylating agents (Dolan *et al*., 1990; Pegg, 1990; Dumenco *et al*., 1993; Nakatsuru *et al*., 1993). In spite of these important roles, remarkably little is known about its mechanisms of interaction with DNA. Here, we have shown that AGT protein is rigorously monomeric in free solution and that it forms complexes with 16-, 30-, and 80-residue single-stranded oligodeoxyribonucleotides that contain ∼4, ∼5, and ∼9 molecules of protein, respectively. The apparent absence of complexes with higher stoichiometries, under conditions of high protein excess, even at the high protein concentrations obtained at sedimentation equilibrium, supports the notion that the DNA is protein-saturated in these complexes. The presence of high-stoichiometry complexes at equilibrium with unbound DNA and the absence of species with intermediate stoichiometry argue strongly that the binding of AGT to single-stranded DNAs is cooperative. Although this result contrasts with an earlier finding for the *E. coli ada* protein (Takahashi *et al*., 1990), cooperative DNA binding is a characteristic of several other single-stranded DNA binding proteins including *E. coli* SSB and bacteriophage T4 gene 32 protein (Lohman & Ferrari, 1994; Casas-Finet & Karpel, 1993). As outlined below, cooperative interactions may provide a mechanism for the the targeting of AGT to specific DNA regions.

The binding density of AGT complexes decreases markedly with increasing DNA length, from ∼4 nt/protein with the 16-mer (oligo 1) to ∼9 nt/protein with the 80-mer (oligo 3). These values are in accord with the size of the binding site predicted from the crystal structure of the *E. coli ada* DNA alkyltransferase (Moore *et al*., 1993) and are consistent with the length-dependence of DNA dealkylation activity observed for the human enzyme (Liem *et al*., 1993). The decrease in binding density that we have observed with increasing DNA length may reflect the formation of unoccupied gaps on the DNA lattice when the number of potential binding sites in the lattice exceeds that typically occupied by a cooperatively-bound cluster of proteins (McGhee & von Hippel, 1974). Alternatively the binding mechanism may differ for DNA sites located in the center of the molecule and for those near DNA ends. Since the proportion of binding sites at ends decreases with increasing DNA length, the observed decrease in binding density would occur if sites at ends were smaller than those in the center of the molecule. Experiments designed to test these notions are currently underway.

Although supercoiled, linear, and single-stranded forms of pBR322 DNA can act as competitors for the AGT-oligo 1 complex, the single-stranded form is significantly more effective than the others (Table 2). This indicates that AGT binds preferentially to single-stranded DNA. Consistent with this inference, superhelical pBR322 is found to be a more effective competitor than the linear form. Supercoiled *E. coli* plasmids, isolated in the absence of perturbing agents, are underwound with respect to linear duplexes (Bauer, 1978). This underwinding facilitates the formation of single-stranded regions (Wang, 1974) which could account for the observed binding preference of AGT. AGT may act on a target base that has been rotated out of the DNA duplex (Spratt & Campbell, 1994). Similar mechanisms have been proposed for human and herpes simplex virus uracil-DNA glycosylases [reviewed by Demple (1995)] and for the cytosine 5-methyltransferase of the *Hha*I restriction-modification system (Klimasauskas *et al*., 1994). For AGT, exposure of the O^6-alkylguanine would be facilitated by localized DNA melting, driven by the differential affinity of the protein for single-stranded DNA. On the other hand, AGT dealkylates duplexes more rapidly than single-stranded forms (Yarosh *et al*., 1986; Dolan *et al*., 1985; Spratt & de los Santos, 1992). Thus, if DNA melting is part of the mechanism, strand separation is not likely to be the rate-limiting step.

The affinity of AGT for duplex DNAs appears to depend on base composition. Poly d(G-C) and poly dG-poly dC (100% G+C) are bound more tightly than is linear duplex pBR322 DNA (54% G+C; Sutcliffe, 1978a), which is bound more tightly than poly d(A-T) and poly dA-poly dT (0% G+C). While the human enzyme dealkylates both O^4-alkylthymine and O^6-alkylguanine residues, the rate of dealkylation is significantly greater with O^6-alkylguanine-containing substrates (Zak *et al*., 1994). On this basis, differences in binding affinity may contribute to the observed rate differences. An exception to this general trend was unfractionated bovine genomic DNA (40% G+C; Adams *et al*, 1981). This DNA was an bound less tightly than any other competitor tested, implying the existence of a class of sites that is bound with substantially lower affinity than those available in the repeating homopolymers or in pBR322. Although this is an indication of sequence specificity, the maximum difference in relative affinity among all competitor DNAs is less than 50-fold. Such weak sequence preference may be necessary if AGT is to act efficiently at alkylguanine sites in different genomic sequence contexts.

A number of important questions remain for future investigation. First, what is the mechanism of DNA-binding cooperativity? Protein−protein interactions seem likely on the basis of the small size of the DNA site occupied by each protein. Curiously, in the absence of DNA, the protein shows no tendency to oligomerize. However, both structural considerations and spectroscopic results argue that AGT is likely to undergo a significant conformational change on binding DNA (Moore *et al*., 1994; Takahashi *et al*., 1990; Chan *et al*., 1993). Such a change could expose new surfaces for protein−protein interaction, resulting in cooperative binding. Second, how does AGT discriminate between lesion sites and the vast excess of competing DNA in which they are embedded? It is striking that the relative affinities for our population of competitors, including single-stranded, duplex, and supercoiled templates, differ by less than 50-fold. Single-stranded and duplex DNAs differ by a similarly small factor in their effectiveness as competitive inhibitors of demethylation of high molecular weight DNA substrates (Bhattacharyya *et al*., 1990). These results argue that AGT has a poor ability to discriminate between DNA sites on the basis of their secondary structure, topological state, or base composition. They also suggest that features not presented by our current population of competitor DNAs are necessary to guide AGT to its sites of action. Such features could include localized distortions of the DNA helix caused by alkylation and/or the presence of additional protein factor(s) that possess the discriminatory activity. Finally, what is the site of AGT action? The binding preference for single-stranded regions over duplex suggests one possibility. Although single-stranded regions do not represent a large fraction of the genome at any moment, such regions are found in advance of DNA-replication complexes. Coopera-

tive interactions between AGT monomers may be sufficient to ensure at least transient occupancy of these regions. High-density (cooperative) binding of such sites would increase the chance of a productive encounter of AGT with a modified base while the movement of replication forks would ensure a systematic scanning of the genome. AGT function in this context would represent a last chance to effect DNA repair before the mutagenic potentials of O^6-alkylguanine and O^4-alkylthymine residues can be expressed.

REFERENCES

Adams, R. L. P., Burton, R. H., Campbell, A. M., Leader, D. P., & Smellie, R. M. S. (1981) *The Biochemistry of the Nucleic Acids*, 9th ed., pp 517, Chapman and Hall, London.

Bauer, W. R. (1978) *Annu. Rev. Biophys. Bioeng. 7*, 287−313.

Bhattacharyya, D., Foote, R. S., Boulden, A. M., & Mitra, S. (1990) *Eur. J. Biochem. 193*, 337−343.

Casas-Finet, J. R., & Karpel, R. L. (1993) *Biochemistry 32*, 9735−9744.

Chan, C. L., Wu, Z., Ciardelli, T., Eastman, A., & Bresnick, E. (1993) *Arch. Biochem. Biophys. 300*, 193−200.

Cohen, G., & Eisenberg, H. (1968) *Biopolymers 6*, 1077−1100.

Cohn, E. J., & Edsall, J. T. (1943) *Proteins, Amino Acids and Peptides as Ions and Dipolar Ions*, pp 391, Reinhold, New York.

Demple, B. (1995) *Curr. Biol. 5*, 719−721.

Dolan, M. E., Morimoto, K., & Pegg, A. E. (1985) *Cancer Res. 45*, 6413−6417.

Dolan, M. E., Oplinger, M., & Pegg, A. E. (1988) *Carcinogenesis 9*, 2139−2143.

Dolan, M. E., Moschell, R. C., & Pegg, A. E. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 5368−5372.

Dolan, M. E., Mitchell, R. B., Mummert, C., Moschel, R. C., & Pegg, A. E. (1991) *Cancer Res. 51*, 3367−3372.

Dumenco, L. L., Allay, E., Norton, K., & Gerson, S. L. (1993) *Science 259*, 219−222.

Fried, M. G. (1989) *Electrophoresis 10*, 366−376.

Fried, M. G., & Crothers, D. M. (1981) *Nucleic Acids Res. 9*, 6505−6525.

Fried, M. G., & Bloomfield, V. A. (1984) *Biopolymers 23*, 2141−2155.

Hudson, J. M., & Fried, M. G. (1990) *J. Mol. Biol. 214*, 381−396.

Jensen, D. E., & von Hippel, P. H. (1976) *J. Biol. Chem. 251*, 7198−7214.

Johnson, M. L., Correia, J. J., Yphantis, D. A., & Halvorson, H. R. (1981) *Biophys. J. 36*, 575−588.

Klimasauskas, S., Kumar, S., Roberts, R. J., & Cheng, X. (1994) *Cell 76*, 357−369.

Laue, T. M., Shah, B. D., Ridgeway, T. M. & Pelletier, S. L. (1992) in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J., & Harding, J. C., Eds.) pp 90−125, The Royal Society of Chemistry, Cambridge, England.

Liem, L. K., Wong, C. W., Lim, A., & Li, B. F. L. (1993) *J. Mol. Biol. 231*, 950−959.

Lindahl, T., Sedgwick, B., Sekiguchi, M., & Nakabeppu, Y. (1988) *Annu. Rev. Biochem. 57*, 133−157.

Lohman, T. M., & Ferrari, M. E. (1994) *Annu. Rev. Biochem. 63*, 527−570.

Loveless, A. (1969) *Nature 233*, 206−207.

Maxam, A., & Gilbert, W. S. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 560−565.

McGhee, J., & von Hippel, P. H. (1974) *J. Mol. Biol. 86*, 469−489.

Mitra, S., & Kaina, B. (1993) *Nucleic Acids Res. 44*, 109−142.

Moore, M. H., Gulbis, J. M., Dodson, E. J., Demple, B., & Moody, P. C. E. (1994) *EMBO J. 13*, 1495−1501.

Nakatsuru, Y., Matsukuma, S., Nemoto, H., Sugano, M., Sekiguchi, M., & Ishikawa, T. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 6468−6472.

Pegg, A. E. (1990) *Cancer Res. 50*, 6119−6129.

Pegg, A. E., & Singer, B. (1984) *Cancer Invest. 2*, 223−237.

Pegg, A. E., Boosalis, M., Samson, L., Moschel, R. C., Byers, T. L., Swenn, K., & Dolan, M. E. (1993) *Biochemistry 32*, 11998−12006.

Pegg, A. E., Dolan, M. E., & Moschell, R. C. (1995) *Progress Nucleic Acid Res. Mol. Biol. 51*, 167−223.

Roy, R., Shiota, S., Kennel, S. J., Raha, R., von Wronski, M., Brent, T. P., & Mitra, S. (1995) *Carcinogenesis 16*, 405−411.

Spratt, T. E., & de los Santos, H. (1992) *Biochemistry 31*, 3688−3694.

Spratt, T. E., & Campbell, C. R. (1994) *Biochemistry 33*, 11364−11371.

Sutcliffe, J. G. (1978a) *Cold Spring Harbor Symp. Quant. Biol. 43*, 77−90.

Sutcliffe, J. G. (1978b) *Nucleic Acids Res. 5*, 2721−2728.

Takahashi, M., Sakumi, K., & Sekiguchi, M. (1990) *Biochemistry 29*, 3431−3436.

von Hippel, P. H., Kowalczykowski, S. C., Lonberg, N., Newport, J. W., Paul, L. S., Stormo, G. D., & Gold, L. (1982) *J. Mol. Biol. 162*, 795−818.

Vossen, K. M., Stickle, D. F., & Fried, M. G. (1996) *J. Mol. Biol. 255*, 44−54.

Wang, J. C. (1974) *J. Mol. Biol. 89*, 783−801.

Yarosh, D. B., Hurst-Calderone, S., Babich, M. A., & Day, R. S., 3rd. (1986) *Cancer Research 46*, 1663−1668.

Zak, P., Kleib, K., & Laval, F. (1994) *J. Biol. Chem. 269*, 730−733.

BI960971K