DOE/ER/60588-3

THE IMPACT OF ENERGY RELATED POLLUTANTS ON CHROMOSOME STRUCTURE

DOE/ER/60588--3

DE90 002541

Progress Report

Grant Period: May 1, 1987 - April 30, 1990

Randolph L. Rill

Principal Investigator

Institute of Molecular Biophysics

The Florida State University

Tallahassee, FL. 32306-3006

Report Prepared: October 30, 1989

PREPARED FOR THE U.S. DEPARTMENT OF ENERGY

UNDER GRANT NUMBER DE-FG05-87-ER60588

PROGRESS REPORT

Summary

The sequence preferences of DNA binding by five intercalating chemicals were examined by techniques analogous to chemical DNA sequencing. The intercalators used were 7R,8S-dihydroxy-9S,10R-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene [anti-(+)-BPDE] and its enantiomer, anti-(-)-BPDE; 8-azido-ethidium, 7-azido-actinomycin D, and bis-(1,10-phenanthroline)Cu(I). The former four compounds form covalent adducts, some fraction of which are labile and cause DNA chain cleavage upon treatment with hot piperidine. The latter compound causes spontaneous chain cleavage via activated oxygen species generated in the presence of molecular oxygen and a reducing agent. Quantitative mapping data on the DNA cleavages caused by these chemicals were obtained on over 1200 DNA bases in restriction fragments and were statistically analyzed. The sequence preferences of all compounds were best described in terms of base triplets, but bases flanking the preferred triplets also influenced binding in many cases. Although the intercalators differed in degree of sequence selectivity and triplet sequences most preferred, YRY and YGG sequences were generally preferred, especially when the 5'-Y was T, while RRY and oligo-purine/oligo-pyrimidine tracts were generally avoided. Details of individual intercalator preferences and other features of the DNA binding of these compounds are described.

Overview

The following section provides a description of the research aims during this period, the general experimental approaches and specific commonly used protocols. Results are detailed in later sections and manuscripts enclosed.

Detection of Non-Random Binding of Intercalating Ligands. During the past grant period we addressed the question of whether various DNA intercalating agents exhibit "sequence specificity" in binding to DNA. The intercalators chosen included two expected to exhibit significant specificity (7-azido-actinomycin D, **azido-ActD** and bis-(1,10-phenanthroline) Cu(I), **$(Phen)_2Cu^+$**); one expected to show little specificity (8-azido-ethidium, **azido-Et**); and the anti-(+)-and anti-(-)-isomers of benzo(a)pyrene-7,8,-diol-9,10-epoxide (**BPDE**). Data on the former three compounds provide a comparison basis for studies of BPDE's and other activated polycyclic aromatic hydrocarbon carcinogens. All of these ligands were observed to interact non-randomly with DNA, both according to base type, and the sequence context of a given base.

The experimental approach taken utilized techniques analogous to DNA sequencing to both "map" sites of ligand binding, and to quantitate the binding. This approach depends on the fact that all of the above chemicals either spontaneously cleave DNA, or form base adducts that are alkali-labile and cause DNA chain cleavage when treated with hot piperidine (as in DNA sequencing chemistry). (Phen)2Cu(I) causes spontaneous DNA cleavage via an active oxygen species (probably •OH) generated via the Cu(I) <=> Cu(II) redox couple in the presence of molecular oxygen and a reducing agent. Azido-ActD and azido-Et are photoaffinity probes, producing a highly reactive nitrene when photolyzed by visible light. BPDE contains a ring epoxide that readily undergoes nucleophilic addition. The latter compounds were all found to produce alkali-labile adducts of all bases, but preferentially at guanines. DNA restriction fragments from clones of the sea urchin *S. purpuratus* early histone genes (of known sequence) in the plasmid pUC9 were labeled on a single end with ^{32}P or ^{35}S, incubated with varying amounts of ligand under appropriate conditions, repurified and piperidine treated (if necessary), then electrophoresed on standard DNA sequencing gels along with samples exposed to normal chemical sequencing procedures. The intensities of bands on autoradiograms of DNA sequencing gels are related to the relative reactivities of the bases. Under conditions often referred to as "single-hit" kinetics, when reaction is slight and no DNA chain is cleaved more than once, the relative intensities of bands are a direct measure of the relative rate constants for covalent addition. Simple inspection

of autoradiograms of such gels showed that for all ligands the reactions were dependent on the sequence context of the base (see below).

Identification of Preferred Sequences of Ligand Binding. Autoradiograms prepared as described above contain the information required to quantitatively establish the sequence preferences of any ligand that either spontaneously cleaves DNA or produces an alkali-labile product. Extraction of this information is technically difficult, however, since large data sets are required to achieve statistical significance. To date, most published studies of this type from other laboratories have been statistically deficient, often reporting data on only one or two sequences of 100 - 300 bases. Since there are 64 possible base triplets and 256 possible quartets, data on at least 1024 bases are required to represent each quartet four times in the data set (on average). Since quartets are not represented statistically in any limited collection of sequences, we have found that a larger data set is required to assure representation of each quartet at least twice. To date we have obtained data on at least 1200 bases for each of the ligands listed above. These data sets were sufficient to identify many major trends, but ideally should be enlarged to confirm certain findings about long range sequence effects on binding preferences.

Our first studies of this type were conducted with $(Phen)_2Cu(I)$, which spontaneously cleaves DNA. Autoradiograms were interpreted in this case by inspection and sequence preferences noted were put on a semi-quantitative base by a simple scoring system (1). This method is tedious and prone to errors. When a good computing densitometer and microcomputers were available we developed a series of programs written in Pascal to solve several problems associated with handling of large data sets and to facilitate identification of preferred sequence in a quantitative fashion. The intensities of about 100-150 bases typically can be integrated accurately on ^{32}P autoradiograms (somewhat more on ^{35}S autoradiograms). A data set of 1200 bases therefore requires data from up to a dozen sequences. These data must be obtained under conditions closely corresponding to single-hit kinetics, and must be merged accurately with normalization. To properly merge band intensities from a series of scans of a single lane, data were taken so that at least 3 bands from the end of one scan overlapped bands from the beginning of the next scan. The averages of the intensities of the overlapping bands from the two scans were then compared and used to normalize intensities of the second scan to that of the first. This process was repeated for successive scans on the same lane. A similar process was used to merge data from lanes corresponding to successive loadings on the DNA sequencing gel. Treatment with hot piperidine alone caused slight chain cleavage. To correct for background, the intensities measured from control lanes (samples treated with piperidine, but not ligand) were subtracted from the sample data prior to normalization and merging. This correction was usually small.

An additional normalization step was required to merge data from different sequences. Normalization was accomplished by comparing the average intensities of *all* bands from each sequence. The output of this program is then a single normalized sequence containing all of the bases and their normalized intensities, along with flags indicating where the individual sequences end. (These flags are important for further processing described below.) In principle this normalization procedure could introduce biases, e.g., if a particular restriction fragment was fortuitously rich or poor in preferred sequences. In practice, quantitative comparisons of highly preferred or avoided sequences indicated by analyses of the total, merged data sets with those indicated by examination of individual data sets showed that uncertainties of this nature did not affect the major trends. In general, uncertainties in the normalization procedure tend to *reduce* the distinctions between the reactivities of specific sequences (e.g. triplets), provided each sequence is represented several times in the data set. In the analyses below we have not attempted to draw conclusions from small differences in preferences or average intensities or in cases where a sequence was represented by one or two occurrences.

Two additional programs were written to extract statistical information from the complete data set. One sorts the intensity data for a chosen base (e.g. guanine) in order of increasing intensity, calculates and plots the number distribution of normalized intensities, and reports characteristics of the intensity distribution (see below). Distinctions according to base type are necessary because the *intrinsic* reactivities of the bases are different. For example, BPDE reactions preferentially at G's, and hardly at all at T's.

Another program searches for each occurrence of the selected base type, notes the sequence context of the base, and updates accumulators corresponding to each possible doublet, triplet, and quartet containing the

target base. These accumulators are used to calculate and report in tabular form the number of occurrences of the base in a particular sequence, the average intensity of bands corresponding to cleavage at the base in the specified sequence, and the propensities for high or low cleavage rates at the target base in the specified sequence (see below). The outputs of these programs provide several valuable insights into reaction preferences.

The concept of "sequence specificity" of binding is, in fact, not a very useful descriptor for small DNA ligands since virtually all sequences exhibit some reactivity, and the binding selectivity is far below that of DNA restriction endonucleases or other truly sequence specific DNA binding proteins. Sequence "preferences" is a more accurate descriptor, but still lacks quantitative meaning. We have used several approaches to provide a quantitative measure of preferences.

A plot of the number distribution of band intensities provides a visual representation of the range of reactivities of a particular base type and is useful for comparing the ranges of reactivities of different ligands. A comparison of these distributions for reactions of guanines illustrates the (expected) finding that the sequence "specificity" of DNA binding *decreases* in the order azido-ActD > anti-(-)-BPDE > azido-Et (Fig. 1). These distributions are generally broad, and moderately (azido-Et) to strongly (azido-ActD) skewed to the high side. Skewing of the distribution suggests that the intrinsic reactivity of a guanine is dependent on flanking sequences, and that some sequences are highly preferred.

A quantitative measure of the relative reactivities of ligands with a particular base is provided by comparing the *ratios* of the band intensities of the "high" and "low" reactivity groups for the ligands (Table I). Definition of the high and low groups is arbitrary, but the upper and lower 10 percentile ranges are commonly used in population statistics and were chosen here for comparison. For example, the most and least reactive guanines, defined in terms of the upper and lower 10 percentiles, differed by a factor of ≈ 17 in reactivity towards both BPDE isomers. By comparison, the guanines most and least reactive with azido-ActD differed by a factor of 40, while these most and least reactive with azido-Et differed by a factor of 5. If formation of the *non-covalent* complexes were rate limiting in each case, the band intensities could be interpreted in terms of site populations--hence non-covalent association constants--in which case these differences in reactivity correspond to approximately 2.2, 1.7, and 0.97 Kcal/mole differences in standard free energy of non-covalent between the most and least reactive sites of binding of azido-ActD, anti-(-)-BPDE, and azido-Et, respectively. (Non-covalent binding appears *not* to be the rate limiting step in BPDE adduct formation, however. As described by Geacintov et al. (2) the ratios of adducts formed at different sites are a complex function of reaction rates, non-covalent binding constants, and the efficiency of DNA addition, as opposed to hydrolysis, of the activated BPDE.

The sequence preferences of ligands were examined at the levels of all doublets, triplets and quartets including the target base (e.g., G). Sequences surrounding G's that were highly preferred were identified using two criteria: the propensity for occurrence of the sequence in the highly preferred group (defined as the upper 10 or 25 percentile of band intensities), and the band intensity of the target guanine of the sequence *averaged over all occurrences of the sequence*. The former criterion, termed the *"reaction propensity"* is defined as in the following example. For a triplet XGY with the target base (G) in the middle position, the propensity of occurrence of the sequence XGY within the high intensity range = P_{XGY}, where

$$P_{XGY} = [N_{XG^{\bullet}Y}/N_{G^{\bullet}}] \,/\, [N_{XGY}/N_G] = [N_{XG^{\bullet}Y}/N_{XGY}] \,/\, [N_{G^{\bullet}}/N_G]$$

$$= \frac{\text{Probability that a preferred G } (G^{*}) \text{ occurs in sequence XGY}}{\text{Probability that any G occurs in sequence XGY}}$$

$$= \frac{\text{Probability that sequence XGY is in the preferred group } (XG^{*}Y)}{\text{Probability that any G is in the preferred group } (G^{*})}$$

The limiting values of a propensity are:

$P_{XGY} = 0$	if no XGY sequences are in the preferred group.
$P_{XGY} = 1.0$	if XGY sequences are in the preferred group with statistical frequency.
$P_{XGY} = N_G/N_{G^\bullet} \approx 4$	if all XGY sequences are in the preferred group. if the preferred group = upper 25 percentile.

These *two* criteria are useful because they measure different aspects of the sequence preferences. The average band intensity for cleavage at a guanine in a particular sequence is a quantitative measure of the degree of preference for reaction (to form a piperidine-labile adduct) at that sequence, relative to other sequences. If the reaction rate reflects the equilibrium population of non-covalently bound ligand at that sequence, the ratio of the average intensity for a specific sequence to the average intensity for all sequences, e.g. $<I_{XGY}>/<I>$, measures the binding affinity of an XGY site relative to an "average" site, and can be related to the difference in free energy for binding at the preferred site relative to an "average" site. Ideally, $\Delta G_{0,Preferred} - \Delta G_{0,Average} = RT \bullet \ln[<I_{Pref.}>/<I>]$. As noted above, this ideal probably does not apply to the case of BPDE reactions and the intensities cannot be simply related to a free energy for any single process.

The propensity for occurrence of a specific sequence in the preferred group is a measure of the *sufficiency* of that sequence for determining a high reactivity. For example, below we show that the G's most reactive with both BPDE isomers are those followed by a second G. However, a GG sequence is not *sufficient* to guarantee high reactivity of the 5'-G, as indicated by the failure of the propensity for reaction at GG doublets to approach the limit expected if the 5'-G's of *all* GG doublets were highly reactive. In fact, the specificities of the two BPDE isomers appear to be determined most stringently in terms of quartet to pentamer sequences.

In general terms we have found that at least triplet sequences are required to define the sequence preferences of *all* ligands studied. In addition, for all ligands we have identified several "surprise" sequences that are strongly preferred; examples are: TGGG for azido-ActD, TGG for azido-Et, and AGG for BPDE. In many cases there are clearly influences of the next-nearest neighbors on the reactivity of a base, hence at least quartets or pentamers are required to fully characterize the preferences. These effects largely cannot be explained in terms of simple stereo/electronic factors or current knowledge of DNA conformation.

The Sequence Preferences of Piperidine-Labile Adduct Formation by Anti-BPDE Isomers.

Sequence preferences for guanine addition in restriction fragments. Two methods are available for cleaving DNA at BPDE adduct sites. Piperidine causes cleavage at adducts on guanine N7, adenine N3, and perhaps cytosine N^4 or O^2 ((3), see also Proposed Research). The more prevalent guanine N^2 and adenine N^6 adducts are stable to piperidine, but quantitative cleavage of the DNA at these adduct sites can be obtained with high power 355 nm laser light (4,5). An appropriate laser was not available to us, so we initially studied the products of piperidine cleavage. A high power, variable wavelength pulsed laser recently became available. We have established conditions for causing scission at BPDE adducts, but have no significant preference data as yet.

Data on piperidine labile adducts were obtained on over 1200 bases, including ≈ G's, using the anti-(+)- and anti-(-)-BPDE isomers. Both isomers reacted non-randomly with guanines, as indicated visually by the distributions of band intensities (Fig. 1). Both distributions peaked at a relatively low intensity and were skewed to the high intensity side, suggesting that some sequences are preferred significantly over most. As noted above, a comparison of the intensities of bands corresponding to the most and least reactive ten percentiles of G's provides a measure of the degree of sequence selectivity (Table I). By this measure BPDE isomers were similar in selectivity, and were about half as selective as azido-actinomycin D, but about three times as selective as azido-ethidium.

Bands corresponding to cleavage products of DNA reacted with anti-(-)-BPDE were sharp and migrated with those obtained from DNA sequencing reactions except at high BPDE doses (Fig. 2), but products resulting from anti-(+)-BPDE reactions were usually broad and appeared to lag behind corresponding bands in sequencing lanes. Based on studies of reactions of oligonucleotides (see below) we attribute this difference to the higher overall reactivity of the anti-(+)-isomer and its much greater propensity to form stable guanine N^2 adducts which reduce the DNA fragment mobility. This phenomenon was reduced at low anti-(+)-BPDE doses, but with a significant loss of sensitivity. As a consequence the data obtained with anti-(+)-BPDE were less extensive and precise preferences could not be unequivocally established. Nonetheless, clear differences in the preferences of the two isomers were noted. A more complete description of our data analysis is given in an enclosed manuscript draft.

The wide range of guanine reactivities with anti-(-)-BPDE was clearly correlated with sequence preferences. The 5'-guanines in GG doublets *not preceded by a guanine* were most reactive on average. The second and successive G's in oligo-G tracts were not unusually reactive. Our results generally agreed with the limited data of Boles and Hogan, except that they noted strong and approximately equal cleavage at all G's in guanine tracts upon photolysis of plasmid DNA modified by racemic anti-(±)-BPDE with 355 nm laser light (5). Since racemic anti-BPDE predominantly forms piperidine-stable guanine N^2 adducts, this difference between the two results may reflect differences in the sequence requirements for BPDE addition at alternative sites. Alternatively, since Boles and Hogan utilized supercoiled DNA, the high reactivity of G tracts may be a consequence of an unusual conformation caused by torsional strain.

Guanines flanked on the 5'-side by T were also generally more reactive than average with anti-(-)-BPDE, with the triplet preferences being in the order TGG ≥ TGT > TGA ≈ TGC. By contrast, anti-(+)-BPDE strongly preferred AGG and CGG triplets, but TGG triplets were only slightly more reactive than average, and other TGN triplets exhibited average reactivity (Table II). In most cases the bands corresponding to strong cleavage at AGG or CGG sequences modified by anti-(+)-BPDE were too broad to make specific base assignments; however, one particularly clear data set obtained with anti-(+)-BPDE suggested another intriguing difference between anti-(-)- and anti-(+)-isomers. In two of two occurrences the strongest cleavage at a CGG triplet occurred at the C instead of the 3'-flanking G. The same sequences reacted with anti-(-)-BPDE were cleaved most strongly at the 5'-G (Fig. 3). Cleavage at one of these C's was extremely strong--greater than that of any G in the sequence! This result needs to be confirmed with other sequences, but is very intriguing and could represent another major difference between anti-(+)- and anti-(-)-BPDE isomer adducts.

More detailed examination of the data for anti-(-)-BPDE preferences showed that bases 5'- and 3'-flanking the preferred triplets affect reactivities. For example, the underlined G's in the following quartets containing triplets that were highly preferred on average exhibited near average or lower than average reactivities: GA<u>G</u>G, A<u>G</u>GG, TT<u>G</u>G, T<u>G</u>GG, TC<u>G</u>G, GC<u>G</u>G, C<u>G</u>GA (Table III, IV). By contrast, T<u>G</u>T triplets reactivities appeared relatively insensitive to flanking bases, with the possible exception of the extra high reactivities of two CT<u>G</u>T sequences. T<u>G</u>A and T<u>G</u>C sequence reactivities were sensitive to flanking bases (Table III,IV). Generally similar reactivity trends were noted for anti-(+)-BPDE, although we have less confidence in these data. The quartets CC<u>G</u>G and C<u>G</u>GA appeared to be particularly reactive with anti-(+)-BPDE, and less so with anti-(-)-BPDE. These observations provided the rationale for synthesis of the self-complementary decamer (5'-AATCCGGATT-3') for further studies.

We have found that the sequences preferences of all chemicals examined are expressed better at the level of triplets, rather than doublets. A more comprehensive analysis of the reactivities of *guanines* with anti-(-)-BPDE has suggested the following trends, although a larger data set is required to achieve statistical significance in some cases. (The following abbreviations are used: R = purine, Y = pyrimidine, the target G is underlined.)

I. R<u>G</u>R sequences are less reactive than average, *excepting* A<u>G</u>G.

   (a) A<u>G</u>G sequences are highly reactive unless flanked on the 5'- or 3'-side by a G.
   (b) Of the *less* reactive R<u>G</u>R sequences, those flanked on the 3'-side by a pyrimidine (e.g. R<u>G</u>RY) are the least reactive.

II. RGY sequences are usually less reactive than average; none are more reactive than average.

III. YGR sequences are generally more reactive than average, but with several exceptions.

(a) A 5'-flanking purine (RYGR) generally enhances reactivity excepting RCGA sequences, which are lower than average.
(b) A 5'-flanking pyrimidine (YYGR) generally enhances reactivity excepting at YTGA.
(c) Effects of 3'-flanking bases are variable.

i. TGRR reactivities are *higher* than average excepting TGAA.
ii. CGRR reactivities are *lower* than average excepting CGGG.
iii. YGAY reactivities are near average.
iv. YGGY reactivities are significantly above average.

IV. YGY sequences are generally slightly more reactive than average with a few exceptions.

(a) YYGT and YGTY (excepting CGTT) reactivities are significantly above average.
(b) CGCT reactivities are below average.

The determinants of sequence preferences by anti-(-)-BPDE are obviously complex, and the same degree of complexity has been found for every chemical we have examined (see below).

It should be noted that the sequence preferences determined for BPDE reactions with guanines do not straightforwardly indicate the actual sites of non-covalent binding or the precise mode of binding in the productive (transition state) complex. For example, we cannot determine from these data alone whether BPDE binds in the CG or GG step in highly preferred CGG sequences. Furthermore, intercalation may not in some cases be the mode of non-covalent or transition state binding. Examination of the reactivities of flanking bases potentially provides some insights into the primary binding site and orientation of the ligand within the site.

Sequence preferences for anti-(-)-BPDE addition at other bases. Bands were also observed on sequencing gels corresponding to cleavage at labile adducts of adenine and cytosine. On average the intensities of these bands were about 1/6 those corresponding to cleavages at guanines. The most reactive *adenines* occurred in the following quartets (N = any base): (T/G)AGT, CGA(A/T), (T/C)ACC, CCAN, and NAGG. The latter two quartets are obviously related to the preference for the first guanine in AGG and TGG triplets. We do not have high confidence in the apparent preference for the A in AGG triplets because of uncertain overlap from the strong adjacent G band, but the high reactivity of the A in CCAN is consistent with binding of anti-(-)-BPDE in the T-3',5'-G step of 5'-TGG-3'and reacting with either the G or the A on the opposing strand. The first three sequences containing preferred A's (and their complements) are not strongly preferred for formation of labile G adducts.

The most reactive *cytosines* occurred in the following quartets: (A/T)TCGN, (A/C)CC(G/C), GCCT, ACC(C>A), and CCCC. (The NCGG sequence also appeared highly reactive, but these data are suspect.) The quartets represented in (A/T)TCG(A/T/C) are not preferred sites of guanine addition. It is interesting to note that the 5'-ATCG-3' sequence is complementary to the 5'-CGAT-3' sequence containing a preferred adenine noted above, suggesting that preferred binding may occur in the G-3',5'-A step. (We have also noted high reactivity of the A or T in a GAT triplet occurring in an oligonucleotide--see below.) The remaining sequences are related to the high preference for G's in G doublets or oligo-G tracts. The preferred reactivity of the ACC(G/C) quartets is noteworthy because high reactivity of the A in (T/C)ACC sequences was also observed, although the last G in GGT sequences is not highly preferred. Studies with azido-actinomycin D and azido-ethidium have also indicated that there is something unique about TGGT and TGGGT sequences (see below).

Reactions of anti-BPDE isomers with oligonucleotide duplexes. We have begun to examine the reactions of both isomers with two duplexes containing central CCGG quartets:

(5'-AATCCGGATT-3')$_2$ and
(5'-CTACCGGATAC-3')•(3'-GATGGCCTATG-5').

Oligomers were labeled on the 5'-end with ^{32}P, annealed to form duplexes, then exposed to increasing doses of anti-BPDE isomers. Samples were then treated or not treated with piperidine and electrophoresed on 20% polyacrylamide DNA sequencing gels. With both isomers we observed the appearance of bands migrating slower than the parent band which we take to be oligomers containing BPDE adducts at different sites since the conformation, hence the frictional coefficient and electrophoretic mobility, of the oligomer is expected to be sensitive to adduct site. Usually we observed one or two predominant bands, which probably represent products of addition at the two guanines. We noted higher yields of slowly migrating material with anti-(+)-BPDE than with anti-(-)-BPDE at identical doses. A high fraction ($\geq$ 70%) of the slowly migrating material from anti-(-)-BPDE reactions was piperidine labile, and a much smaller fraction of the corresponding products of anti-(+)-BPDE reaction was labile. Both of these observations were consistent with the known relative reactivities of the two isomers.

The 5'-guanine in the 5'-CCGG-3' sequences in both oligonucleotides yielded more alkali-labile adducts than the 3'-guanine, in agreement with studies of DNA restriction fragments. Interestingly, the adenine in the GAT sequence of 5'-CTACCGGATAC-3' also produced high yields of labile anti-(-)-BPDE adduct, and lower yields of anti-(+)-BPDE adduct (Fig. xx). These results are consistent with our observations of certain GAT sequences in restriction fragments, and the reported greater reactivity of A's with anti-(-)-BPDE (6).

We have encountered some difficulties with what appears to be background formation of low amounts of piperidine-labile adducts of all bases, even at very low BPDE doses. We believe this is due to trace amounts of single stranded oligonucleotide arising because of the very low oligonucleotide concentrations ($\approx$ 1 μM) used in these experiments. This hypothesis is supported by observations that increasing the salt concentration reduced (but did not eliminate) this background. Increasing the oligonucleotide concentration should greatly reduce this problem.

Sequence Preferences of DNA Binding by Azido-Ethidium.

We have used 8-azido-ethidium (azido-Et), in which the 8-amino group of the parent ethidium is replaced by a photo-labile azide moiety, as a photoaffinity probe of the preferred sites of ethidium binding. Previous studies by Graves and co-workers (7,8) have shown that 8-azido ethidium binds to DNA in a manner virtually indistinguishable from the parent ethidium in terms of the binding constant and neighbor exclusion parameter, and certain spectroscopic properties. Azido-ethidium and other aromatic azido compounds are photo-labile, forming highly reactive nitrenes that add to DNA bases upon illumination with visible light. Some fraction of these base adducts are labile to hot piperidine, allowing mapping of adduct sites on radioactively end-labeled DNA restriction fragments by techniques described above.

We observed cleavage of azido-Et treated DNA at all bases, with yields of labile adducts being in the order G>C$\approx$T>A. The degree of sequence preference observed with azido-Et was less than that of anti-BPDE, (Phen)$_2$Cu$^+$, or azido-ActD. Nonetheless, definite preferences were noted, and the least and most preferred sequences differed in reactivity by factors of 4-5. The nitrenes created by photolysis of aromatic azides are highly reactive and are expected to react rapidly with solvent or DNA bases. The *off* rate constant for ethidium bound to DNA is on the order of milliseconds, hence we expect that the nitrene will react at the binding site or with solvent prior to dissociation. In this case the relative band intensities obtained on DNA sequencing gels under "single hit" conditions reflect the site populations--hence the relative association constants for binding to a particular sequence. A factor of 5 difference in association constant corresponds to a significant difference in ΔG^o of about 1 Kcal/mole.

As observed with anti-BPDE isomers, the most reactive G's on average occurred in GG doublets. TGG, GG(T>G>C), and CGG were the only triplets with central G's that were significantly more reactive than

average. Bases next-nearest the adducted guanine exerted significant effects. The most preferred quartets are represented by: YGGY, TGGG, GTGG, ACGC, TCGT and TGGT, GGT(A,G>T), GGG(G>A,T), GGCA, CGTA, TGAG. The strongest *pentamer* preferences suggested are therefore YGGYA, TGGGG, GTGGT, and TCGTA.

Similar analyses of data on cleavages at other bases indicated the following preferences:

Cytosines:
(N)CCG(C,T)
(T)GCG(C,T)
(A)ACG(G)
(A)CCA(A)

Thymines:
(N)ATG(G,T)
(T>C)GTG(T,C)
(A>C)GTA(T>C)
(A>T)CTA(T>G)

Adenines:
(T)TAG(C>A)
(C)TAT(G)
(N=C)CAT(T,G)

Considered together, these data for all bases suggest that YGG and YRY sequences are generally strong ethidium binding sites. Comparison of the reactivities of the three bases in these triplets indicated that intercalation occured in the pyrimidine-3',5'-purine step of YRY triplets, as expected. By contrast, the high reactivities of both G's in YGGY sequences, particularly TGGT, suggests that intercalation occurs between the two G's. Reactivities of YGG and YRY triplets clearly were not uniform, however, and depended significantly on 5'- and 3'-flanking bases. A manuscript in preparation describes this work in detail. Copies will be sent as soon as possible.

Sequence Preferences of DNA Binding by Azido-Actinomycin D.

The sequence preferences of 7-azido-actinomycin D were examined as a test of our procedures because this compound is one of the most "sequence-specific" DNA binding drugs known, and the binding has been well characterized by other means. Actinomycin D (ActD) is an antitumor antibiotic that appears to act *in vivo* by inhibiting RNA polymerase. The mechanism of non-covalent DNA binding of ActD has attracted much attention because of its strong preference for 5'-dGpdC-3' sequences. Binding is thought to involve intercalation of the tricyclic aromatic phenoxazone ring into a GC step, with the two equivalent cyclic pentapeptide lactone substituents lying in the minor groove and making hydrogen bond contacts with the N^2-amino groups of the nearest neighbor guanines. Recent studies of the non-covalent binding of ActD to oligonucleotide duplexes have indicated, however, that binding is also influenced by next-nearest neighboring bases (9). We examined this higher order specificity using 7-azido-actinomycin D as a photoaffinity probe as described for azido-ethidium. Data accumulated on about 800 base pairs showed that GC doublets were strongly preferred addition sites, as expected, but only if the 5'-flanking base was a pyrimidine and the 3'-flanking base was ***not*** cytosine. In addition we observed a strong preference for binding at a GG doublet in 5'-TGGG-3', and exceptionally strong reaction with TGGGT. Strong non-covalent binding of ActD to these sequences has not been reported. These experiments are described in more detail in a manuscript in press (10) (copy enclosed).

We have now obtained data on an additional ≈ 400 bases, and are in the process of reanalyzing G, as well as other base preferences. Our collaborator, Dr. David Graves (U. Miss.) is beginning non-covalent binding studies with oligonucleotides to determine if the sequence preferences we noted are paralleled by non-covalent association constants of the parent ActD.

DNA Binding and Cleavage by Bis-(1,10-phenanthroline) Cu(I).

The sequence preferences of DNA cleavage in restriction fragments. The bis-(1,10-phenanthroline) Cu(I) complex, $(Phen)_2Cu(I)$, was chosen as the first compound for study by these techniques for two reasons. DNA cleavage by this ligand is spontaneous and not base-selective, thus in some respects the acquisition and analysis of data was facilitated. In the presence of molecular oxygen and a reducing agent a Cu(I) < = > Cu(II) redox cycle is generated at the DNA binding site, producing an active oxygen species commonly thought to be hydroxyl radical (11,12). The hydroxyl radical is short-lived and reacts with solvent or diffuses to a neighboring nucleotide where it abstracts a proton from C1' or C4', causing chain cleavage in the immediate vicinity of the binding site (13,14). Of more fundamental significance in the choice of this reagent were the observations, based on low resolution mapping experiments, that $(Phen)_2Cu(I)$ shows a marked preference for sequences 5'-flanking genes, and tends to avoid protein coding sequences (15,16).

The sequence preferences of DNA cleavage by $(Phen)_2Cu^+$ were determined by mapping sites on >2000 base pairs (bp). As described in the previous progress report and in detail in a recent publication ((1) copy enclosed), we have found that $(Phen)_2Cu^+$ has a predominant preference for TAT triplets, and a moderate preference for TGT triplets; and also a strong preference for TAAT, TAGPy, and CAGT quartets. Oligopurine and oligopyrimidine tracts are poorly reactive. We have also shown that $(Phen)_2Cu^+$ binds in the minor groove at preferred sequences. This strong TAT preference is unusual for a potentially intercalating DNA ligand, and can be explained mainly in terms of a requirement for binding at the minor groove side of a pyrimidine-3',5'-purine step and the stereo/electronic effects of 2-amino groups of guanine, which inhibit minor groove binding.

Effects of single base substitutions on cleavages at pyrimidine-purine-pyrimidine sequences in oligonucleotides. To test the influence of the guanine 2-amino group on specificity we synthesized the undecamers of the general formula: CCCT(PyPuPy)CCCC, where PyPuPy = TAT, CAT, TGT, TIT (I = inosine) and TAC; and their complements: GGGA(PuPyPu)GGGG, where PuPyPu = ATA, GTA, ITA, ACA, ATG, and ATI. Minihelices of complementary oligonucleotides containing TAT, TGT and TIT were cleaved strongly at the central purine and weakly at other positions, as expected from results obtained with long DNA fragments, demonstrating that local helix geometry is sufficient to define the cleavage specificity. Substitution of I for G in the oligomers containing CAT:ITA and TIT:ACA significantly enhanced cleavage at the central purine in PyPuPy sequences, confirming that a guanine 2-amino group in the minor groove inhibits binding. The oligomers containing central TAC:ATG and TAC:ATI sequences, however, exhibited similar reactivity that was reduced relative to TAT, indicating that more subtle factors influence the base preference at the third position in PyPuPy triplets. This work is described in detail in a Biochemistry article (17) (copy enclosed).

The non-covalent binding of $(Phen)_2Cu^{\pm}$ and related compounds to DNA. More recently we developed techniques to perform spectral and viscometric titrations of DNA with $(Phen)_2Cu^+$ under oxygen-free conditions (required to prevent oxidation of Cu^+ to Cu^{++}, and examined $(Phen)_2Cu^+$ binding to random sequence DNA. The spectral properties of the $(Phen)_2Cu+$/DNA complex and the effect of $(Phen)_2Cu^+$ on the viscosity of DNA solutions are consistent with an intercalation binding mechanism. The spectrum of the complex exhibited strong hypochromism and a strong induced circular dichroism spectrum consistent with exciton coupling, and addition of $(Phen)_2Cu^+$ to DNA caused a large increase in DNA solution viscosity. The binding isotherm was complex, suggesting positive cooperativity, and was dependent on the concentration of excess phenanthroline. Excess phenanthroline also had subtle effects on the absorption properties of the $(Phen)_2Cu^+$:DNA complex. (Phenanthroline is required in excess to assure that $(Phen)_2Cu^+$ is the predominant Cu^+ species.) In all cases binding was well described by McGhee-Von Hippel theory with a neighbor exclusion parameter $n \approx 2$, as found for many intercalators, and a positive cooperativity parameter ω = 3.0 to 4.6, depending on the excess phenanthroline concentration. Binding measurements were also performed with the redox-inactive bis-(2,9-dimethyl-1,10-phenanthroline)Cu(I) complex, and with phenanthroline alone. No DNA binding of these two compounds could be detected by spectroscopic or viscosity criteria, although phenanthroline appeared to be a competitive inhibitor of $(Phen)_2Cu^+$ binding. The mono-phenanthroline:Cu(I) complex appeared to bind DNA by intercalation, however, as indicated by the large increase in DNA solution viscosity when titrated with 1:1 Cu(I):phenanthroline. Additional aspects of the complex equilibria in this system are described in a nearly completed manuscript. Copies will be sent as soon as possible.

Effects of reducing agent and phenanthroline concentration on the DNA by copper:phenanthroline. During the course of the above studies we noted that the characteristic visible spectrum of $(Phen)_2Cu^+$ was not obtained in the presence of the reducing agent mercaptopropionic acid (MPA), but was formed in the presence of ascorbate, an alternative reducing agent. Additional visible and NMR spectroscopic studies showed that the $(Phen)_2Cu^+$ complex was disrupted by MPA, perhaps due to formation of an (MPA)(Phen)Cu(I) complex. This finding raised the possibility that the cleavage preferences observed previously were due to this uncharacterized species, not $(Phen)_2Cu^+$. We subsequently showed, however, that the preference for TAT was still observed when ascorbate was substituted for MPA as the reducing agent, and that the DNA cleavage reaction proceeded much faster with ascorbate as reductant. Excess phenanthroline accelerated cleavage in the presence of MPA, but decreased cleavage in the presence of ascorbate. We conclude that the $(Phen)_2Cu^+$ complex is the major DNA cleaving agent even in the presence of MPA, but there is competition between MPA and phenanthroline for Cu^+. The decrease in DNA cleavage by phenanthroline when ascorbate was the reductant is consistent with the inhibitory effect of phenanthroline on $(Phen)_2Cu^+$ binding noted above. In addition, when the Cu:Phen ratio was reduced to 1:1 we observed a strong preference for cleavage at the central CG of a 5'-CCGGG-3' sequence, and very reduced cleavage at a nearby TAT, which was highly preferred at a Cu:Phen ratio of 1:2. This result is consistent with observations that N-alkyl phenanthrolinium ions prefer GC-rich DNA over AT-rich DNA (16), and directly demonstrates the strong influence of the second phenanthroline ring in directing the sequence preference of the $(Phen)_2Cu^+$ complex to TA rather than CG steps.

A model of $(Phen)_2Cu^{\pm}$ binding to pyrimidine-purine-pyrimidine triplets. All of the spectroscopic and oligonucleotide cleavage data obtained are consistent with a model we have proposed for $(Phen)_2Cu^+$ binding to preferred TAT sequences in which one phenanthroline ring intercalates in the T-3',5'-A step from the minor groove side, and the second phenanthroline ring of the complex lies in the minor groove, nearly parallel with the helix axis and approximately in the positions occupied by the guanine N^2-amino groups in a C-3',5'-G step (17). This positioning is consistent with the low degree of DNA cleavage observed at CG steps and CAT sequences, and the relative rates of cleavage at different nucleotide sugars in the binding site.

The student who performed these experiments (J. Veal) has now graduated and work with $(Phen)_2Cu^+$ has been terminated. I believe these results have implications for design of a new class of DNA binding drugs, and will be preparing grant applications to continue work on this complex and structurally related compounds.

Overall Conclusions.

All of the compounds studied are intercalators, but they differ significantly in the degree of side chain complexity. The degree of sequence preference for ligand binding generally increased with increasing complexity, a result that is not surprising, but lends credence to the methods used. Despite the obvious influences of the side chain several common themes were noted that can be most simply interpreted in terms of fundamental preferences for binding by intercalation. Several general conclusions can be drawn from the collective analyses.

1. There are no "simple" rules, and specification of preferences in terms of doublets is often a misleading oversimplification. Classic descriptions in terms of GC, AT, or "pyrimidine-purine step" specificities are relatively meaningless and can be misleading.

2. There are strong influences of *both* nearest-neighbors of a base on reactivity.

3. Next nearest-neighbors also significantly influence reactivities.

4. YRY and YGG sequences are generally good binding sites for intercalators, particularly if the 5'-pyrimidine is T. Oligopurine/oligopyrimidine tracts are generally poor sites of intercalator binding, but are not totally avoided.

These conclusions conform only in part to currently accepted concepts of intercalator binding. A strong dependence of reactivity on one of the bases flanking the modified base is expected to be related to the

energetics of base unstacking and the stereo/electronic "fit" of the ligand within the intercalation site. These effects can be investigated in relatively straightforward (though not necessarily unambiguous) ways by standard molecular modeling techniques. The influences of next-nearest-neighbors of the intercalation site are hard to comprehend in these terms for simple intercalators lacking extended side chain substituents such as ethidium and BPDE, and are most likely to reflect subtle influences of the local, sequence-dependent DNA conformation in the ground (unbound) or perturbed (bound) states.

NMR Studies of the d(AATCCGGATT)$_2$ Duplex.

During the Spring 1989 semester I visited the laboratory of Dr. David Kearns, Chemistry Department, U. Cal. San Diego, to gain experience with 2D NMR methods. I examined the thermal denaturation of the above duplex, and assigned nearly all non-exchangeable proton resonances by 2D TOCSY and ROESY NMR experiments (all base, H1', H2', H2", H3' and several H4' resonances). This duplex will be used for studies of B[a]P-derivative binding (see Proposed Research).

Publications and Manuscripts Under This Grant Period 5/1/87-Present.

Rowland, R. D., and Rill, R. L., "Atypical Changes in Chromatin Structure During Development in the Sea Urchin Lytechinus variegatus", Biochim. Biophys. Acta 980, 169-178 (1987)

Strzelecka, T. S. and Rill, R. L., "Solid State ^{31}P NMR Studies of DNA Liquid Crystalline Phases. The Isotropic to Cholesteric Transition", J. Am. Chem. Soc. 109, 5413-4518 (1987).

Veal, J. and Rill, R. L., "The Sequence Specificity of DNA Cleavage by $(Phen)_2Cu^+$", Biochemistry 27, 1822-1827 (1988).

Veal, J. M. and Rill, R. L., "Sequence Specificity of DNA Cleavage by Bis-(1,10)-phenanthroline Cu(I): Effects of Single Base Pair Transitions on the Cleavage of Preferred Pyrimidine-Purine-Pyrimidine Triplets", Biochemistry 28, 3243-3250 (1989).

Smith, R. M. and Rill, R. L., "Carbon-13 NMR Studies of Nucleosome Cores. Assignment of Mobile Histone Chain Segments", J. Biol. Chem. 264, 10574-10581 (1989)

Rill, R. L., Marsch, G., and Graves, D., "7-Azido-Actinomycin D: A Probe of the Sequence Specificity of Actinomycin D Binding to DNA", J. Biomolecular Structure & Dynamics, in press.

Rill, R. L. and Marsch, G., "The Sequence Preferences of Covalent DNA Binding by the Anti-(+)- and Anti-(-)-benzo[a]pyrene Diol Epoxides", submitted.

Rill, R. L., Marsch, G., and Graves, D., "The Sequence Specificity of Covalent DNA Binding of 8-Azido-ethidium", in preparation.

Veal, J. M. and Rill, R. L., "The Non-covalent DNA Binding of Bis-(1,10-Phenanthroline)Cu(I) and Related Compounds", in preparation.

Veal, J. M. and Rill, R. L., "The Influence of Reducing Agent and Phenanthroline Concentration on DNA Cleavage by Phenanthroline/Copper", in preparation.

Predoctoral Associates Trained.

Carolyn Waldron, M.S. in Biology received, 1988.
Teresa Strzelecka, Ph.D. in Molecular Biophysics received, 1988.
James M. Veal, Ph.D. in Molecular Biophysics received, 1989.
Glenn Marsch, Ph.D. candidate, Molecular Biophysics (expected graduation 6/90).
Karl Hecker, Ph.D. candidate, Chemistry (third year student).
Kunal Merchant, Ph.D. candidate, Molecular Biophysics (second year student).
Qi Chen, Ph.D. candidate, Molecular Biophysics (second year student).

Bibliography

1. Veal, J.M., and Rill, R.L. (1988) *Biochemistry.* **27**, 1822-1827

2. Geacintov, N.E., Shahbaz, M., Ibanez, V., Moussaoui, K., and Harvey, R.G. (1988) *Biochemistry.* **27**, 8380-8387

3. Sage, E., and Haseltine, W.A. (1984) *J. Biol. Chem.* **259**, 11098-11102

4. Boles, T.C., and Hogan, M.E. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81**, 5623-5627

5. Boles, T.C., and Hogan, M.E. (1986) *Biochemistry.* **25**, 3039-3043

6. Osborne, M.R., Jacobs, S., Harvey, R.G., and Brookes, P. (1981) *Carcinogenesis.* **2**, 553-558

7. Garland, F., Graves, D.E., Yielding, L.W., and Cheung, H.C. (1980) *Biochemistry.* **19**, 3221-3226

8. Graves, D.E., Watkins, C.L., and Yielding, L.W. (1981) *Biochemistry.* **20**, 1887-1892

9. Chen, F.M. (1988) *Biochemistry.* **27**, 6393-6397

10. Rill, R.L., Marsch, G.A., and Graves, D.E. (1989) *J. Biomol. Struct. Dyn.* **in press,**

11. Sigman, D.S., Graham, D.R., D'Aurora, V., and Stern, A.M. (1979) *J. Biol. Chem.* **254**, 12269-12272

12. D'Aurora, V., Stern, A.M., and Sigman, D.S. (1978) *Biochem. Biophys. Res. Commun.* **80**, 1025-1032

13. Pope, L.M., Reich, K.A., Graham, D.R., and Sigman, D.S. (1982) *J. Biol. Chem.* **257**, 12121-12128

14. Spassky, A., and Sigman, D.S. (1985) *Biochemistry.* **24**, 8050-8056

15. Cartwright, I.A., and Elgin, S. (1982) *Nucleic Acids Res.* **10**, 5835-5852

16. Gabbay, E.J., Scofield, R.E., and Baxter, C.S. (1973) *J. Am. Chem. Soc.* **95**, 7850-7857

17. Veal, J.M., and Rill, R.L. (1989) *Biochemistry* **28**, 3243-3250