DNA recognition by quinoxaline antibiotics: use of base-modified DNA molecules to investigate determinants of sequence-specific binding of triostin A and TANDEM

Christian BAILLY* and Michael J. WARING†¹

*Laboratoire de Pharmacologie Moléculaire Antitumorale du Centre Oscar Lambret et INSERM U124, Place de Verdun, 59045 Lille, France, and †Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QJ, U.K.

The methodology of DNAase I footprinting has been adapted to investigate the sequence-specific binding of two quinoxaline drugs to DNA fragments containing natural and modified bases. In order to help comprehend the molecular origin of selectivity in the bis-intercalation of triostin A and TANDEM at CpG and TpA sites respectively, we have specifically examined the effect of the 2-amino group of guanine on their sequence specificity by using DNA in which that group has been either removed from guanine, added to adenine or both. Previous studies suggested that the recognition of particular nucleotide sequences by these drugs might be dependent upon the placement of the purine 2-amino group, serving as a positive or a negative effector for

triostin A and TANDEM respectively. However, the footprinting data reported here indicate that this is not entirely correct, since they show that the 2-amino group of guanine is important for the binding of triostin A to DNA but has absolutely no influence on the interaction of TANDEM with TpA steps. Apparently the binding of triostin A to CpG sites is primarily due to hydrogen bonding interaction between the cyclic peptide of the antibiotic and the 2-amino group of guanine residues, whereas the selective binding of TANDEM to TpA sites is not hydrogen-bond driven and probably originates mainly from steric and/or hydrophobic interactions, perhaps involving indirect recognition of a suitable minor groove structure.

INTRODUCTION

The right-handed B-form DNA double helix has distinct major and minor grooves which can be recognized by proteins and small molecules [1,2]. The minor groove, where most compounds of low molecular mass bind, is not only by definition narrower than the opposite major groove but also contains a different pattern of hydrogen bond donors and acceptors. Each A·T and G·C base pair presents two sets of acceptor groups in both the minor and major grooves, whereas hydrogen bond donor groups are not equally distributed. An exocyclic amino group protrudes toward the major groove on both A·T and G·C pairs but only exists in the minor groove of the G·C pair (Figure 1). In other words, the guanine 2-amino group is the only hydrogen bond donor available for recognition via the minor groove and as such it is expected, and has been shown, to be a key element for sequence-specific recognition of DNA by small molecules. In the last few years, we have demonstrated that the direct (digital) recognition of preferred binding sequences by AT- and GC-selective minor groove binders, such as distamycin and mithramycin respectively, and by various intercalating agents, such as actinomycin and daunomycin, depends crucially on the presence or absence of the 2-amino group of guanine in the minor groove [3–7]. Furthermore, we have shown that indirect (analogue) recognition of DNA sequences by certain DNA binding proteins (e.g. the factor for inversion stimulation and the high-mobility group protein HMG-D) is also controlled to a large extent by the location of that same exocyclic group in the minor groove [8,9]. The terms ‘digital’ and ‘analogue’ were coined to distinguish mechanisms of recognition based on direct interaction with cognate groups such as hydrogen bonding rather than local complementarity of secondary structure. In the present study, we have extended these investigations to two prominent members of

the well-characterized family of quinoxaline antibiotics, triostin A and des-*N*-tetramethyl-triostin A (TANDEM) [10] (Figure 1).

For many years it has been known that triostin A bis-intercalates into DNA and exhibits a distinct preference for GC-rich sequences, whereas its synthetic des-*N*-tetramethyl derivative TANDEM bis-intercalates selectively into alternating AT sequences [11–13]. Footprinting studies eventually revealed that triostin A binds specifically to sequences centred around a CpG step, whereas TANDEM selectively recognizes TpA sites [14–17]. Most binding studies have been performed with [*N*-MeCys³, *N*-MeCys⁷]TANDEM rather than with TANDEM itself, but the two synthetic drugs appear equivalent in terms of binding strength and specificity. Several structures of triostin A, TANDEM and [*N*-MeCys³, *N*-MeCys⁷]TANDEM bound to short duplex oligonucleotides have been elucidated by crystallography and NMR [18–25]. In general, the bis-intercalated complexes are comparable. For triostin A bound to NpCpGpN sites as well as for TANDEM bound to NpTpApN sites, the two quinoxaline rings bracket the central YpR dinucleotide step and the cyclic depsipeptide comes to lie in the helical minor groove. In each case, the two sandwiched base pairs become substantially underwound by ~10° to 26° and buckle inward by ~20°.

The structural basis for the sequence-specific binding of triostin A and TANDEM to CpG and TpA sites respectively resides at least in part in the formation of two intermolecular hydrogen bonds between each NH group of the alanine residues of the drug and the nitrogen N-3 of the purine bases (guanine for triostin and adenine for TANDEM). Early studies showed that elimination of these particular hydrogen bonds as a consequence of replacing the alanine residues in TANDEM by lactic acid results in loss of the ability to bind tightly to DNA [26]. For triostin A, there exists an additional pair of specific hydrogen bonds between the alanine carbonyl group and the 2-amino

Abbreviations used: DAP, 2,6-diaminopurine; TANDEM, des-*N*-tetramethyl-triostin A.

¹ To whom correspondence should be addressed.

Figure 1 Structures of hydrogen-bonded purine-pyrimidine base pairs, drawn with the major groove side at the top and the minor groove side at the bottom; broken lines represent hydrogen bonds, and the purine 2-amino group is in bold (top); chemical structures of triostin A and TANDEM (bottom)

group of guanine. No such hydrogen bonds can form in the TANDEM-NpTpApN complex because the corresponding adenine C-2 position bears only a hydrogen atom rather than an exocyclic substituent; but the absence of this hydrogen bond is not sufficient to explain why TpA sites are strongly preferred over CpG sites by TANDEM. The width of the minor groove is nearly identical at TpA and CpG sites and is sufficient to accommodate the peptide rings of either drug [23]. Different hypotheses have been advanced. On the basis of NMR data, it has been proposed that it is the existence of a pair of intramolecular hydrogen bonds between the Ala CO and Val NH residues of TANDEM that prevents the drug from approaching the guanine 2-amino group [24]. For TANDEM to contact the exocyclic amino group which protrudes into the minor groove, the intramolecular hydrogen bond would have to be broken and that would be energetically unfavourable [24]. Subsequently it was suggested that better stacking interaction between the quinoxaline rings and the bases at TpA sites compared with CpG

sites helps to stabilize the complex [25]. Theoretical calculations showed that dipole-dipole interaction of the quinoxaline ring with an A-T base pair is indeed more favourable than with a G-C pair, also pointing to a prominent role for stacking interactions in the sequence specificity of TANDEM [27]. However, despite some elegant experiments, the reason why TANDEM binds to TpA sites, whereas triostin A prefers CpG sites, remains enigmatic and has prompted us to address the problem. The endeavour to elucidate the molecular rules that govern the recognition of different sequences by triostin A and TANDEM is not only crucial to understanding how the quinoxaline antibiotics recognize their preferred binding sites but is also important for determining how small molecules in general 'read' DNA sequences and, by extension, for the goal of synthesizing gene-targeted drugs.

Accordingly, we have compared the binding of triostin A and TANDEM to DNA molecules in which the purine 2-amino group has been either removed from guanine, added to adenine or both. A homologous series of 160 base pair fragments of DNA containing inosine and/or 2,6-diaminopurine residues (abbreviated DAP or D in a sequence for clarity) in place of guanosine and/or adenine residues respectively were synthesized by PCR and subjected to DNAase I cleavage in the absence and presence of the drug. For triostin A, we find that the purine 2-amino group is absolutely required and constitutes a key structural element which directs sequence-specific binding to CpG sites in DNA. By contrast, TANDEM is totally insensitive to the relocation of the exocyclic amino group; it binds to a TpD site just as well as it binds to a canonical TpA site. Unlike the NMR results reported for the related [*N*-MeCys³, *N*-MeCys⁷]-TANDEM [25], we find no evidence for binding of TANDEM to CpI steps. In our hands the binding of TANDEM to DNA is quite independent of the presence of the purine 2-amino group. Thus the molecular basis for the sequence specificity of triostin A and TANDEM must be subtly different.

MATERIALS AND METHODS

Quinoxalines

Triostin A was a gift of Drs. H. Otsuka and T. Yoshida of Shionogi and Co., Osaka, Japan. TANDEM, kindly supplied by Dr. R. K. Olsen, was synthesized as previously described [28]. In both cases, the drug was dissolved to a concentration of 100 μ M in 10 mM Tris/HCl, pH 7.0, 10 mM NaCl containing 40% (v/v) methanol. The stock solution was diluted to working concentrations with appropriate volumes of 10 mM Tris/HCl, pH 7.0, 10 mM NaCl and methanol so as to yield a final methanol concentration not exceeding 10% (v/v) in the footprinting reactions. Under these conditions methanol is known not to affect the nuclease activity [29]. Extinction coefficients of 10900 $M^{-1} \cdot cm^{-1}$ and 12130 $M^{-1} \cdot cm^{-1}$ were used to determine triostin A and TANDEM concentrations from absorbance measurements at 325 nm [14].

Chemicals and biochemicals

Ammonium persulphate, Tris base, acrylamide, bis-acrylamide, ultrapure urea, boric acid, tetramethylethylenediamine and dimethyl sulphate were from BDH. Formic acid, piperidine and formamide were from Aldrich. Photographic requisites were from Kodak. Bromophenol Blue and Xylene Cyanol were from Serva. The nucleoside triphosphate labelled with [³²P] (γ -ATP) was obtained from NEN Dupont. 2,6-Diaminopurine deoxyribonucleoside-5'-triphosphate was prepared by Dr. M. Guo via phosphorylation of the requisite nucleoside kindly provided by

Drs. Otto Dahl and Peter Nielsen of the University of Copenhagen. Restriction endonucleases *EcoRI* and *AvaI* (Boehringer), *Taq* polymerase (Promega), DNase I (Sigma) and T4 polynucleotide kinase (Pharmacia) were used according to the supplier's recommended protocol in the activity buffer provided. The primers, 5'-AATTCCGGTTACCTTTAATC and 5'-TCGGGA-ACCCCCACCACGGG having a 5'-OH or 5'-NH₂ terminal group, were obtained from the Laboratory of Molecular Biology, Medical Research Council, Cambridge. Checks were carried out to ensure that the primers blocked with a 5'-NH₂ group were free from contamination and did not serve as substrates for labelling by the kinase. All other chemicals were analytical grade reagents, and all solutions were prepared using doubly deionized, Millipore-filtered water.

Preparation, purification and labelling of DNA fragments containing natural and modified nucleotides

Plasmid pKMp27 [30] was isolated from *Escherichia coli* by a standard SDS/NaOH lysis procedure and was purified by banding in CsCl-ethidium bromide gradients. Ethidium was removed by several isopropanol extractions followed by exhaustive dialysis against Tris-EDTA buffer. The purified plasmid was then precipitated and resuspended in appropriate buffer before digestion by the restriction enzymes. The 160 base pair *tyrT*(A93) fragment used as a template was isolated from the plasmid by digestion with restriction enzymes *EcoRI* and *AvaI*. It is worth mentioning that this template DNA bore a 5'-phosphate due to the action of *EcoRI* and thus only the newly synthesized DNA (with normal or modified nucleotides) can be labelled by the kinase.

PCR

The protocol used to incorporate inosine and/or DAP residues into DNA is comparable with those previously used to incorporate 7-deazapurine or inosine, with only a few minor modifications [3,4,31]. PCR reaction mixtures contained 10 ng of *tyrT*(A93) template, 1 μ M each of the appropriate pair of primers (one with a 5'-OH and one with a 5'-NH₂ terminal group) required to allow 5'-phosphorylation of the desired strand, 250 μ M of each appropriate dNTP (dTTP, dCTP plus dATP or dTTP and dGTP or dTTP according to the desired DNA) and 5 units of *Taq* polymerase in a volume of 50 μ l containing 50 mM KCl, 10 mM Tris/HCl, pH 8.3, 0.1 % Triton X-100 and 1.5 mM MgCl₂. To prevent unwanted primer-template annealing before the cycles began, the reactions were heated to 60 °C before adding the *Taq* polymerase [32]. Finally, paraffin oil was added to each reaction to prevent evaporation. After an initial denaturing step of 3 min at 94 °C, 20 amplification cycles were performed, with each cycle consisting of the following segments: 94 °C for 1 min, 37 °C for 2 min, and 72 °C for 10 min. After the last cycle, the extension segment was continued for an additional 10 min at 72 °C, followed by a 5 min segment at 55 °C and a 5 min segment at 37 °C. The purpose of these final segments was to maximize annealing of full-length product and to minimize annealing of unused primer to full-length product. The reaction mixtures were then extracted with chloroform to remove the paraffin oil, and parallel reactions were pooled. Several extractions with water-saturated n-butanol were performed to reduce the volume before loading the samples on to a 6% non-denaturing polyacrylamide gel. After electrophoresis for about 1 h, a thin section of the gel was stained with ethidium bromide so as to locate the band of DNA under UV light. The same band of DNA free of ethidium was excised, crushed and soaked in

elution buffer (500 mM ammonium acetate, 10 mM magnesium acetate) overnight at 37 °C. This suspension was filtered through a Millipore 0.22 μ m filter and the DNA was precipitated with ethanol. After washing with 70 % ethanol and vacuum drying of the precipitate, the purified DNA was resuspended in the kinase buffer.

DNA labelling and purification

The purified PCR products were 5'-end labelled with [γ -³²P]ATP in the presence of T4 polynucleotide kinase according to a standard procedure for labelling blunt-ended DNA fragments [33]. After completion the labelled DNA was again purified by 6% polyacrylamide gel electrophoresis and extracted from the gel as described above. Finally, the radioactive polynucleotide was resuspended in 10 mM Tris/HCl, pH 7.0 buffer containing 10 mM NaCl.

DNase I footprinting

DNase I experiments were performed essentially according to the original protocol [29]. The digestion of the samples (6 μ l) of the labelled DNA fragment dissolved in 10 mM Tris buffer, pH 7.0, containing 10 mM NaCl was initiated by the addition of 2 μ l of a DNase I solution whose concentration was adjusted to yield a final enzyme concentration of about 0.01 unit/ml in the reaction mixture. The extent of digestion was limited to less than 30 % of the starting material so as to minimize the incidence of multiple cuts in any strand ('single-hit' kinetic conditions). Optimal enzyme dilutions were established in preliminary calibration experiments. After 3 min, the digestion was stopped by freeze-drying, samples were lyophilized, washed once with 50 μ l of water, lyophilized again and then resuspended in 4 μ l of an 80 % formamide solution containing tracking dyes. Samples were heated at 90 °C for 4 min and chilled in ice for 4 min before electrophoresis.

Electrophoresis and autoradiography

DNA cleavage products were resolved by PAGE under denaturing conditions (0.3 mm thick, 8 % acrylamide gels containing 8 M urea) capable of resolving DNA fragments differing in length by one nucleotide. Electrophoresis was continued until the Bromophenol Blue marker had run out of the gel (about 2.5 h at 60 W, 1600 V in TBE buffer, BRL sequencer model S2). Gels were soaked in 10% acetic acid for 15 min, transferred to Whatman 3MM paper, dried under vacuum at 80 °C, and subjected to autoradiography at -70 °C with an intensifying screen. Exposure times of the X-ray films (Fuji R-X) were adjusted according to the number of counts per lane loaded on each individual gel (usually 24 h).

Quantification by storage phosphor imaging

A Molecular Dynamics 425E PhosphorImager was used to collect data from storage screens exposed to the dried gels overnight at room temperature [34]. Baseline-corrected scans were analysed by integrating all the densities between two selected boundaries using ImageQuant version 3.3 software. Each resolved band was assigned to a particular bond within the *tyrT*(A93) fragment by comparison of its position relative to sequencing standards generated by treatment of the DNA with formic acid followed by piperidine-induced cleavage at the purine residues (G + A track).

Purine 2-amino group strongly influences the sequence-specific binding of triostin A to DNA

sequencing gels. A typical phosphorimage of a gel obtained with the normal and modified PCR products is shown in Figure 2. It is immediately obvious that the cleavage profile varies substantially from one DNA to another, depending on the ability of the drug to find its preferred target sites. With the inosine DNA (G $\rightarrow$ I substitution) the cleavage patterns show absolutely no footprints whatsoever, indicating that the removal of the 2-amino group is detrimental to the binding of triostin A to DNA. This is in accordance with previous results obtained with the related antibiotic echinomycin which also fails to bind DNA lacking the purine 2-amino group [3]. In contrast, it can be seen at glance that the DAP-containing DNA (A $\rightarrow$ D substitution) provides a good substrate for triostin A. Particularly strong footprints can be identified around nucleotide positions 32 and 88. The latter site is flanked by regions where the cleavage by the enzyme has been massively enhanced in the presence of the antibiotic. Clearly the DAP-DNA contains some excellent binding sites for triostin A as well as sequences to which the drug refuses to bind. It is also apparent from the gel that (i) the footprinting pattern obtained with the doubly substituted I+DAP-DNA resembles that obtained with the DAP-DNA and (ii) the minimal drug concentration required to detect footprints on both DAP-containing DNA species is considerably lower than is needed with DNA containing just the natural bases. As shown in Figure 2 and in previous studies [14], a concentration in the 10–20 μ M range is needed in order to evidence binding of triostin A to CpG sites. With both the DAP-DNA and I+DAP-DNA, the footprint around position 88 is already very pronounced at only 0.5 μ M triostin A (Figure 2A). A full titration of the DAP-substituted DNAs with a large range of concentrations of antibiotic (Figure 2B) reveals that the binding to the newly created site can be unambiguously detected at a concentration as low as 10–20 nM, i.e. about a thousand times lower than that required to detect binding to the best sites in normal DNA. There is no doubt that the A $\rightarrow$ D substitution potentiates the interaction of triostin A with DNA considerably. Even in the presence of inosine residues, the binding of triostin A to DAP-containing sites remains considerably tighter than its binding to CpG sites.

The differences between normal DNA and its counterparts containing DAP with or without inosine nucleotides are well illustrated by the quantified cleavage plots shown in Figure 3. These plots were constructed from densitometric analyses of phosphorimages of several gels including those shown in Figure 2. As expected, the cleavage of normal DNA is most affected at GC-rich sequences and the positions of the footprints (negative values) coincide rather well with the location of CpG that is known to bind triostin A. The cleavage patterns observed with the *tyrT*(A93) fragment are totally consistent with those previously reported for triostin A and this particular DNA as well as for other natural DNA fragments under similar experimental conditions [14]. By contrast, with the I + DAP-DNA the cleavage at IC-rich sequences (e.g. around position 75) is considerably enhanced, whereas that at DT-rich tracts is markedly impaired. The differential cleavage plot for the I + DAP-DNA appears, to a certain extent, like a mirror image of that seen with normal DNA (Figure 3A), just as if the binding sites had been shifted from GC sequences to DT sequences. The quantification confirms that shifting the 2-amino group from guanine to adenine residues has a considerable influence on the sequence-specific binding of triostin A.

The position of the footprints produced by triostin A on the doubly substituted DNA corresponds nicely to the location of the TpA dinucleotides which have now become TpD by virtue of the transferred purine 2-amino group. In normal DNA, triostin

Figure 3 Quantitative analysis of triostin A–DNA recognition

(A) Differential cleavage plots comparing the susceptibility of the normal and the doubly-substituted I+DAP-DNA to DNase I attack in the presence of triostin. The sequence shown on the x-axis corresponds to that of the Watson strand of the *tyrT*(A93) fragment containing natural bases. In the modified DNA, adenine and guanosine residues are replaced by DAP and inosine residues. Positive and negative values correspond, respectively, to enhanced or diminished DNAase I cutting at each internucleotide bond. The vertical scale is in units of $\ln(f_a) - \ln(f_c)$, where f_a is the fractional cleavage at any bond in the presence of the drug and f_c is the fractional cleavage of the same bond in the control. The results are displayed on a logarithmic scale for the sake of convenience. The filled and open rectangles show the positions of the CpG and TpA binding sites, respectively. The two lower panels show the differential cleavage plots determined with I+DAP-DNA (B) and DAP-DNA (C) in the presence of increasing concentrations of triostin A from 2.5 nM to 50 μ M as indicated.

A does not bind to the ATAT box located at positions 87–90 but recognizes the flanking GC-rich sequences around positions 73–79 and 95–107 (Figure 3A). The situation is completely reversed with the DNA containing DAP residues instead of adenines. As shown in Figures 3(B) and 3(C), triostin A fails to bind to the IC clusters (where the DNAase I cleavage is enhanced)

but now binds to the regions surrounding the TpD dinucleotides. As mentioned above, with both the I+DAP-DNA (Figure 3B) and the DAP-DNA (Figure 3C) the interaction of triostin A at the TTDC and DTDT sites is considerably strengthened as judged from the appearance of the footprints at very low drug concentrations.

Figure 4 DNAase I footprinting of 50 μ M TANDEM on the Crick strand-labelled *tyrT*(A93) DNA containing natural or modified bases

Other details as for Figure 2.

Binding of TANDEM to TpA sites is independent of the position of purine 2-amino groups

The same footprinting experiments with normal and modified DNA species were repeated in the presence of TANDEM. One example of a gel run with the *tyrT* fragment labelled on the Crick strand is shown in Figure 4. The results are completely different

from those obtained with triostin A. Indeed, it can easily be seen that the binding of TANDEM to DNA is totally unaffected by the different base substitutions. The footprint around position 92 remains unchanged whether the DNA contains natural bases or inosine and/or DAP (vertical bars in Figure 4). Neither the position of the binding site nor the intensity of the footprint seem to vary (although, of course, the banding patterns in the four control tracks differ considerably as we have noted elsewhere [6]). The differential cleavage plots in Figure 5 reveal two strong footprints at positions 91 and 109, each of which coincides with the position of a TpA dinucleotide, as expected. The stronger protection at position 91 compared with 109 agrees with the tighter binding of TANDEM to an ATAT site compared with a TTAT site [17]. Neither the removal of the purine 2-amino group from guanines nor its ectopic addition to adenines affects the binding to these TpA sites. In other words, the introduction of a moderately large substituent between the drug and its presumptive seat against the minor groove side of the adenine residue does not perturb the recognition process. Moreover, the formation of numerous TpI or CpI sites (by virtue of the G $\rightarrow$ I substitution) at different positions along the *tyrT* fragment is not sufficient to create TANDEM binding sites.

DISCUSSION

In earlier reports [3,4] we showed that the purine 2-amino group is a crucial requirement for binding of echinomycin into the minor groove of the double helix at CpG sites. The same conclusion applies for triostin A, which is structurally close to echinomycin and binds equally well to CpG sites. The bio-synthetic bis-quinoline analogue of echinomycin called 2QN [35] behaves similarly (C. Bailly and M. J. Waring, unpublished work). Therefore there is no doubt that the 2-amino group of guanine constitutes a key structural element in the mechanism of DNA recognition by the quinoxaline family of antibiotics. However, it is no less obvious that other aspects of DNA structure must also contribute to sequence recognition. From a minor groove point of view, CpG and TpD share the same hydrogen bonding capabilities and therefore should bind triostin A equally well. This is clearly not the case. As shown in this study, the binding of triostin A to TpD sites is enormously preferred over binding to the canonical CpG sites. There is something special about regions of alternating T·D base pairs which generates unusually good triostin A binding sites. The

Figure 5 Differential cleavage plots comparing the susceptibility of the normal and modified DNA species to DNase I cutting in the presence of 50 μ M TANDEM

The two rectangles show the positions of the TpA steps (TpD steps in DAP-DNA and I+DAP-DNA) present at the binding sites. Other details as for Figure 3.

local structure and/or the rigidity of the TpD sites could be exploited by triostin A to fit particularly neatly within the minor groove. Alternatively, the stacking of its quinoxaline rings upon DAP·T base pairs could be especially propitious, as has been suggested for echinomycin [27]. It appears that in common with many GC-selective DNA-binding antibiotics, such as actinomycin and mithramycin, both DNA structure and interaction with guanine are involved in determining the specific binding of quinoxalines to their target sites.

The results with TANDEM are much more puzzling. Before the experiments were performed, we anticipated that (i) the addition of the 2-amino group on adenine residues would hinder the proper fitting of the bulky cyclic peptide moiety of TANDEM into the minor groove at TpA sites and (ii) the removal of that exocyclic substituent would permit the drug to bind to CpI sites. NMR experiments have suggested that the related drug [*N*-MeCys³,*N*-MeCys⁷]TANDEM can bind sequence specifically to CpI much as it binds to TpA [25]. The results reported here show that our prediction was not correct and that the exocyclic guanine 2-amino group apparently plays no part in the recognition of TpA sites by TANDEM. In that respect, for reasons which remain obscure the footprinting data are difficult to reconcile with the NMR studies conducted with a short oligonucleotide d(GGACITCC)₂ complexed with [*N*-MeCys³,*N*-MeCys⁷]TANDEM. The central tetranucleotide of that sequence occurs between positions 57 and 60 of inosine-substituted *tyr*T DNA but there is little sign of footprinting on inosine DNA at that site compared with the other three species (Figure 5). We note that the NMR experiments were of necessity conducted at much higher concentrations of the TANDEM analogue, indeed of both interacting species, under which circumstances it might be more difficult to detect differences in the thermodynamics of binding. Whether this can explain the discrepancy we cannot tell. Even so, it seems unlikely that the detailed nature of binding will be established without direct structural studies e.g. by X-ray crystallography.

The question as to why TANDEM binds specifically to TpA sites remains an unresolved issue at this point. Our study indicates that the specific recognition is not based upon an obstructive effect of the guanine 2-amino group as was originally thought. Nor is the manifest preference of TANDEM for binding to TpA sites flanked by A·T or T·A base pairs explicable on a similar basis: our findings agree with those of others that G·C base pairs flanking the TpA dinucleotide are not acceptable, and the situation is unchanged if the flanking pairs are altered to IC (cf. Figures 4 and 5). Indirectly, these observations reinforce the idea that the binding specificity arises largely from stacking and hydrophobic interactions. Several recent investigations on drug–DNA interactions, such as a comprehensive spectroscopic and thermodynamic study of the interaction of the minor groove binder Hoechst 33258 with the d(CGCAAATTTGCG)₂ duplex, have suggested that the hydrophobic transfer of the drug from solution into the duplex binding site provides the major driving force for the recognition process [36]. More specific molecular interactions, including hydrogen bonding and van der Waals contacts, were found to play only a minor role in stabilizing the drug–DNA complex. By analogy, we could hypothesize that hydrophobic interactions between TANDEM and favoured TpA sites might play an important part in determining the molecular basis of the sequence recognition process. It is very likely that the local structure of the DNA helix at TpA sites, perhaps related to the width of the minor groove, as well as the hydrophobic potential of the drug are involved in determining sequence-

specific binding of quinoxalines like TANDEM to DNA. In terms of drug design, this means that further development of DNA-targeted compounds could potentially be achieved by judiciously introducing non-polar groups into the drug structure. The present study is therefore a useful guide to the rational development of DNA reading molecules.

This work was done with the support of research grants (to C.B.) from the Association pour la Recherche sur le Cancer and (to M.J.W.) from the Wellcome Trust, Cancer Research Campaign, Association for International Cancer Research and the European Union. We thank the Sir Halley Stewart Trust for a grant to assist cooperation, and we are grateful to Dean Gentle and Julie Morgan for technical assistance.

REFERENCES

- Neidle, S. and Waring, M. J. (1993) *Molecular Aspects of Anticancer Drug–DNA Interactions*, vols. 1 and 2, Macmillan, London
- Travers, A. A. (1993) *DNA–Protein Interactions*, Chapman and Hall, London
- Marchand, C., Bailly, C., McLean, M. J., Moroney, S. E. and Waring, M. J. (1992) *Nucleic Acids Res.* **20**, 5601–5606
- Bailly, C. and Waring, M. J. (1995) *Nucleic Acids Res.* **23**, 885–892
- Bailly, C. and Waring, M. J. (1995) *J. Am. Chem. Soc.* **117**, 7311–7316
- Bailly, C., Møllegaard, N. E., Nielsen, P. E. and Waring, M. J. (1995) *EMBO J.* **14**, 2121–2131
- Jennewein, S. and Waring, M. J. (1997) *Nucleic Acids Res.* **25**, 1502–1509
- Bailly, C., Waring, M. J. and Travers, A. A. (1995) *J. Mol. Biol.* **253**, 1–7
- Bailly, C., Payet, D., Travers, A. A. and Waring, M. J. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13623–13628
- Waring, M. J. (1993) in *Molecular Aspects of Anticancer Drug–DNA Interactions*, Vol. 1, (Neidle, S. and Waring, M. J., eds.), pp. 213–242, Macmillan, London
- Lee, J. S. and Waring, M. J. (1978) *Biochem. J.* **173**, 129–144
- Fox, K. R., Olsen, R. K. and Waring, M. J. (1982) *Biochim. Biophys. Acta* **696**, 315–322
- Fox, K. R., Olsen, R. K. and Waring, M. J. (1982) *Br. J. Pharmacol.* **70**, 25–40
- Low, C. M. L., Olsen, R. K. and Waring, M. J. (1984) *FEBS Lett.* **176**, 414–420
- Low, C. M. L., Fox, K. R., Olsen, R. K. and Waring, M. J. (1986) *Nucleic Acids Res.* **14**, 2015–2033
- Lavesa, M., Olsen, R. K. and Fox, K. R. (1993) *Biochem. J.* **289**, 605–607
- Fletcher, M., Olsen, R. K. and Fox, K. R. (1995) *Biochem. J.* **306**, 15–19
- Viswamitra, M. A., Kennard, O., Cruse, W. B. T., Egert, E., Sheldrick, G. M., Jones, P. G., Waring, M. J., Wakelin, L. P. G. and Olsen, R. K. (1981) *Nature (London)* **289**, 817–819
- Wang, A. H. J., Ughetto, G., Quigley, G. J., Hakoshima, T., Van der Marel, G. A., Van Broom, J. H. and Rich, A. (1984) *Science* **225**, 1115–1121
- Ughetto, G., Wang, A. H. J., Quigley, G. J., van der Marel, G. A., van Broom, J. H. and Rich, A. (1985) *Nucleic Acids Res.* **13**, 2305–2323
- Quigley, G. J., Ughetto, G., Van der Marel, G. A., Van Broom, J. H., Wang, A. H. J. and Rich, A. (1986) *Science* **232**, 1255–1258
- Addess, K. J., Gilbert, D. E., Olsen, R. K. and Feigon, J. (1992) *Biochemistry* **31**, 339–350
- Addess, K. J., Sinsheimer, J. S. and Feigon, J. (1993) *Biochemistry* **32**, 2498–2508
- Addess, K. J. and Feigon, J. (1994) *Biochemistry* **33**, 12386–12396
- Addess, K. J. and Feigon, J. (1994) *Biochemistry* **33**, 12397–12404
- Olsen, R. K., Ramasamy, K., Bhat, K. L., Low, C. M. L. and Waring, M. J. (1986) *J. Am. Chem. Soc.* **108**, 6032–6036
- Gallego, J., Ortiz, A. R. and Gago, F. (1993) *J. Med. Chem.* **36**, 1548–1561
- Ciardelli, T. L. and Olsen, R. K. (1977) *J. Am. Chem. Soc.* **99**, 2806–2807
- Low, C. M. L., Drew, H. R. and Waring, M. J. (1984) *Nucleic Acids Res.* **12**, 4865–4877
- Drew, H. R., Weeks, J. R. and Travers, A. A. (1985) *EMBO J.* **4**, 1025–1032
- Sayers, E. W. and Waring, M. J. (1993) *Biochemistry* **32**, 9094–9107
- Bloch, W. (1991) *Biochemistry* **30**, 2735–2747
- Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Johnston, R. F., Pickett, S. C. and Barker, D. L. (1990) *Electrophoresis* **11**, 355–360
- Sheldrick, G. M., Heine, A., Schmidt-Bäse, K., Pohl, E., Jones, P. G., Paulus, E. and Waring, M. J. (1995) *Acta Crystallogr.* **B51**, 987–999
- Haq, I., Ladbury, J. E., Chowdhry, B. Z., Jenkins, T. C. and Chaires, J. B. (1997) *J. Mol. Biol.* **271**, 244–257