Sequence-specific binding of [*N*-MeCys³,*N*-MeCys⁷]TANDEM to TpA

Manuel LAVESA,* Richard K. OLSEN† and Keith R. FOX*

*Department of Physiology and Pharmacology, University of Southampton, Bassett Crescent East, Southampton, U.K., and †Department of Chemistry and Biochemistry, Utah State University, Logan, UT, 84322, U.S.A.

The sequence selective binding of [*N*-MeCys³,*N*-MeCys⁷]TANDEM to DNA has been studied by footprinting experiments on DNA fragments containing the self-complementary sequences CGCGATATCGCG, CGCGTATACGCG, CGCGTTAACGCG and CGCGAATTCGCG. DNAase I and micrococcal nuclease reveal drug-induced footprints with the central sequences

ATAT, TATA and TTAA, but not AATT, suggesting that the ligand binds to the dinucleotide TpA. The ligand renders certain adenines hyper-reactive to diethyl pyrocarbonate. These are observed with ATAT, TATA and TTAA, but not AATT, and are located both within, and distal to, the TpA-binding sites.

INTRODUCTION

The quinoxaline group of antibiotics have long been known to bind to DNA by bifunctional intercalation [1], recognizing CG-rich sequences [2]. Footprinting [3,4], n.m.r. [5,6] and X-ray-crystallographic [7,8] studies have shown that the naturally occurring members of this series, echinomycin and triostin A, are selective for the dinucleotide CpG. In contrast, the synthetic derivatives TANDEM and [*N*-MeCys³,*N*-MeCys⁷]TANDEM (Figure 1), which lack four and two *N*-methyl groups respectively, are known to bind to AT-rich DNAs [9]. Simple footprinting experiments confirmed the AT-selectivity, but could not identify whether binding was to TpA or ApT, since there appeared to be a requirement for flanking AT base-pairs [10,11]. DNA 'melting' and restriction-enzyme-inhibition studies suggested TpA as the binding site [12]. By contrast, model-building studies, extrapolating from the crystal structure of TANDEM [13], suggested ApT as the binding site and proposed an elegant model accounting for this specificity. A recent n.m.r. study has shown that [*N*-MeCys³,*N*-MeCys⁷]TANDEM binds to GGTTAAC and GGATATCC, but not GGAATTCC, consistent with binding to TpA, and suggested that selectivity arises from hydrogen bonds between the alanine NH groups and adenine N-3 [14]. A series of experiments on the effects of this ligand on long stretches of (AT)_n were also best explained by specific binding to TpA not ApT [15].

In the present study we have examined the selectivity of [*N*-MeCys³,*N*-MeCys⁷]TANDEM by footprinting experiments on the sequences CGCGATATCGCG, CGCGTATACGCG, CGCGTTAACGCG and CGCGAATTCGCG, each of which has been cloned into longer DNA fragments. The results demonstrate unambiguously that the ligand binds selectively to TpA not ApT.

MATERIALS AND METHODS

Drugs and enzymes

[*N*-MeCys³,*N*-MeCys⁷]TANDEM was prepared as previously described [11]. Since this drug is only sparingly soluble in aqueous systems, it was stored as a stock solution of 2 mM in dimethyl sulphoxide and diluted to working concentrations in 10 mM Tris/HCl, pH 8.0, containing 10 mM NaCl, immediately before use. The final dimethyl sulphoxide concentration was always less than 5% (v/v); this had no effect on either DNAase I digestion or modification by diethyl pyrocarbonate. DNAase I and

micrococcal nuclease were purchased from Sigma and stored as previously described [3,16].

DNA fragments

The self-complementary oligonucleotides dCGCGATATCGCG, dCGCGTATACGCG, dCGCGTTAACGCG, dCGCGAATTCGCG were prepared on an Applied Biosystems DNA synthesizer and used without further purification. The dodecamers were treated with polynucleotide kinase and ATP and ligated into *Sma*I (CCC/GGG)-cut pUC19 and transformed into *Escherichia coli* TG2. White colonies were picked from 5-bromo-4-chloroindol-3-yl β -D-galactoside/isopropyl thiogalactoside plates containing 100 μ g/ml ampicillin. Their sequences were checked by dideoxy sequencing using a T7 sequencing kit (Pharmacia). Since the insertion of 12 bp fragment preserves the reading frame, and so generates blue colonies, some of the white colonies obtained were found to contain one or two base deletions around the *Sma*I site. In particular the clone of CGCGATATCGCG lacks one GC base-pair from each end of the insert (i.e., as if it was cloned into the sequence CC/GG), the clone of CGCGTATACGCG lacks a single CG base-pair on the 5' (*Eco*R1) side. Full-length clones were obtained for CGCGTTAACGCG (since this contains an in-frame stop codon, TTA) and CGCGAATTCGCG. Other full-length inserts, not used in the present work, were later obtained from the blue colonies.

DNA fragments containing the dodecamer inserts were obtained by cutting with *Hind*III, labelling at the 3'-end with α -[³²P]dATP using reverse transcriptase and cutting again with *Sac*I or *Kpn*I. The radiolabelled fragments were separated from the rest of the plasmid on 8%-polyacrylamide gels.

DNA cleavage

The fragments were digested with DNAase I [3], micrococcal nuclease [16] and diethyl pyrocarbonate [17,18] as previously described. The products of digestion were resolved on 12%-polyacrylamide gels containing 8 M urea and 20% formamide. The presence of formamide was essential to prevent unacceptable band compressions in the GC regions. Gels were run at 1500 V for about 2 h, fixed in 10% acetic acid, transferred to Whatmann 3MM paper, dried under vacuum at 80 °C and subjected to autoradiography at -70 °C with an intensifying screen. Autoradiographs were scanned with a Joyce-Loebl Chromoscan III microdensitometer.

Figure 1 Structures of Triostin A, TANDEM and [*N*-MeCys³,*N*-MeCys⁷]TANDEM

Triostin A, $R_1 = R_2 = \text{CH}_3$; TANDEM, $R_1 = R_2 = \text{H}$; [*N*-MeCys³,*N*-MeCys⁷]TANDEM, $R_1 = \text{CH}_3$, $R_2 = \text{H}$.

Figure 2 DNAase I digestion of fragments containing the inserts CGCGATATCGCG, CGCGTATACGCG, CGCGTTAACGCG and CGCGAATTCGCG in the presence (TAN) and absence (CON) of 100 μM [*N*-MeCys³,*N*-MeCys⁷]TANDEM

The square brackets show the positions of the inserts, the filled boxes indicate the central four AT base-pairs. Each pair of lanes corresponds to digestion by the enzyme for 1 and 5 min, as indicated at the bottom of the gel lanes. Tracks labelled 'G' are dimethyl sulphate/piperidine markers specific for guanine. The hatched boxes show the position of the TpA site (CTAG) outside the insert within the polylinker.

RESULTS

DNAase I cleavage

Figure 2 presents the results of DNAase I digestion of the four DNA fragments in the presence and absence of 100 μM [*N*-MeCys³,*N*-MeCys⁷]TANDEM. Clear footprints can be seen in the central regions of ATAT, TATA and TTAA; no changes are apparent with AATT. These results suggest that [*N*-MeCys³,*N*-MeCys⁷]TANDEM recognizes the dinucleotide TpA, since this is present in ATAT, TATA and TTAA, but not in AATT. The footprints with ATAT and TATA are clearer than with TTAA, possibly suggesting that the latter presents a poorer binding site. Experiments performed with various concentrations of [*N*-MeCys³,*N*-MeCys⁷]TANDEM revealed that, although ATAT and TTAA required 100 μM ligand to generate a footprint, TATA produced a footprint with only 25 μM ligand.

Figure 3 Micrococcal-nuclease digestion of fragments containing the inserts CGCGATATCGCG, CGCGTATACGCG, CGCGTTAACGCG and CGCGAATTCGCG in the presence (TAN) and absence (CON) of 100 μM [*N*-MeCys³,*N*-MeCys⁷]TANDEM

The square brackets show the positions of the inserts, the filled boxes indicate the central four AT base-pairs. Each pair of lanes corresponds to digestion by the enzyme for 1 and 5 min, as indicated at the bottom of the gel lanes. The hatched boxes indicate cleavage of the TpA step outside the insert in the polylinker.

In addition to the clear footprints in the centre of the inserts, a further region of protection can be seen towards the bottom of the gels in the remainder of the polylinker. This is clearest with TTAA and TATA and represents binding of the ligand to the sequence CTAG previously shown with the isolated pUC13 polylinker [11].

Micrococcal-nuclease cleavage

We have confirmed the selectivity of [*N*-MeCys³,*N*-MeCys⁷]TANDEM for TpA by performing similar footprinting using micrococcal nuclease as the probe. The results are presented in Figure 3. Since this enzyme cuts almost exclusively at pA and pT bonds [16], the major cleavage products in the controls correspond to cutting at the centre of the inserts. Cleavage of ATAT, TATA and TTAA is clearly protected by the ligand, while that of AATT is unaffected. The two strong bands present towards the bottom of each gel correspond to cutting of the TpA

Figure 4 Effect of [*N*-MeCys³,*N*-MeCys⁷]TANDEM on modification by diethyl pyrocarbonate

The hatched boxes represent the relative modification by diethyl pyrocarbonate in the control; the filled boxes correspond to 100 μ M [*N*-MeCys³,*N*-MeCys⁷]TANDEM. The height of each bar represents the relative intensity of each band; data were taken from densitometer scans of the autoradiograms. Only the central six base-pairs of each sequence are presented.

step in the remainder of the polylinker. These are also protected by the ligand.

Modification by diethyl pyrocarbonate

We have further examined the interaction of [*N*-MeCys³,*N*-MeCys⁷]TANDEM with TpA by studying its effect on the modification of purines by diethyl pyrocarbonate. Echinomycin has previously been shown to render adenines distal to its CG-binding site hyper-reactive to diethyl pyrocarbonate [17–20]. TANDEM also affects the reactivity of adjacent adenines to modification by diethyl pyrocarbonate [17]. In a recent study we suggested that, within long regions of alternating AT, the adenine sandwiched between the chromophores of TANDEM might also become hyper-reactive [15]. The results of these experiments are presented in Figure 4. Once again it can be seen that the ligand has affected the modification of ATAT, TATA and TTAA, but not AATT. The strongest reaction can be seen at the adenine immediately distal to the TpA step in TTAA. For both ATAT and TTAA the adenines sandwiched between the quinoxaline chromophores (ATAT and TTAA) are also rendered hyper-reactive. Both adenines of TATA become equally reactive. The simplest explanation of this is that they are the bases sandwiched between the chromophores at the two TpA-binding sites. It seems unlikely to be due to modification of adenines adjacent to each TpA step, since the first enhancement is proximal to the second TpA site, while the second enhancement is two bases distal to the first TpA.

DISCUSSION

All three of the probes used in this work demonstrate that [*N*-MeCys³,*N*-MeCys⁷]TANDEM interacts with the tetranucleotides ATAT, TATA and TTAA, but not AATT. This provides compelling evidence that the ligand recognizes the dinucleotide TpA. This is in contrast with the prediction of ApT selectivity made on the basis of the crystal structure of the ligand [13], but is consistent with more recent n.m.r. and footprinting studies [14,15]. Previous studies have shown that, in addition to its primary selectivity, the ligand requires the presence of flanking

regions of A and T [10,11]. We might therefore expect that the strongest binding would be to ATAT or TTAA, containing a single central TpA step, flanked by AT base-pairs. However, the fragment requiring the lowest ligand concentration to produce a clear footprint was TATA. This could be because it possesses two adjacent overlapping binding sites. In addition, TTAA presented the least clear footprints, possibly because the tetranucleotide TTAA adopts a less favourable conformation than TATA or ATAT. Indeed, this is consistent with the different cleavage patterns found for the various fragments. It is tempting to suggest that [*N*-MeCys³,*N*-MeCys⁷]TANDEM binds in the order TATA > ATAT > TTAA. It is also worth noting that the concentrations required to produce footprints in the present study are higher than those needed with longer stretches of (AT)_n [15]. In addition TANDEM has been shown to bind in a highly co-operative fashion to poly(dA-dT) [9], suggesting that isolated TpA steps may not represent the best binding sites.

TANDEM unwinds DNA by 46° [9], similar to the value of 48° for echinomycin [1]. It is therefore not surprising that [*N*-MeCys³,*N*-MeCys⁷]TANDEM produces enhancements in diethyl pyrocarbonate reactivity at adenines distal to its TpA-binding site (i.e. at TAA) as does echinomycin at CGA, since these enhancements are generally interpreted as arising from drug-induced unwinding of the local DNA helix [17,19]. However, it also appears that [*N*-MeCys³,*N*-MeCys⁷]TANDEM renders adenines within the sandwiched TpA site hyper-reactive to diethyl pyrocarbonate. This is not surprising, since the ligand binds in the minor groove, while diethyl pyrocarbonate reacts at purine N-7 located in the major groove. This observation does, however, suggest that some of the unwinding is taken up by the sandwiched base-pairs, which therefore may not be in a standard B-DNA configuration.

This work was supported by grants from the Cancer Research Campaign and the Science and Engineering Research Council. K.R.F. is a Lister Institute Research Fellow.

REFERENCES

- Waring, M. J. and Wakelin, L. P. G. (1974) *Nature* (London) **252**, 653–657
- Wakelin, L. P. G. and Waring, M. J. (1976) *Biochem. J.* **157**, 721–740
- Low, C. M. L., Drew, H. R. and Waring, M. J. (1984) *Nucleic Acids Res.* **12**, 4865–4879
- Van Dyke, M. W. and Dervan, P. B. (1984) *Science* **225**, 1122–1127
- Gao, X. and Patel, D. J. (1988) *Biochemistry* **27**, 1744–1751
- Gilbert, D. A. and Feigon, J. (1991) *Biochemistry* **30**, 2483–2494
- Wang, A. H.-J., Ughetto, G., Quigley, G. J., Hakoshima, T., van der Marel, G. A., van Boom, J. H. and Rich, A. (1984) *Science* **225**, 1115–1121
- Ughetto, G., Wang, A. H.-J., Quigley, G. J., van der Marel, G., van Boom, J. H. and Rich, A. (1985) *Nucleic Acids Res.* **13**, 2305–2323
- Lee, J. S. and Waring, M. J. (1978) *Biochem. J.* **173**, 128–144
- Low, C. M. L., Olsen, R. K. and Waring, M. J. (1984) *FEBS Lett.* **176**, 414–419
- Low, C. M. L., Fox, K. R., Olsen, R. K. and Waring, M. J. (1986) *Nucleic Acids Res.* **14**, 2015–2033
- Evans, D. H., Lee, J. S., Morgan, A. D. and Olsen, R. K. (1982) *Can. J. Biochem.* **60**, 131–136
- 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
- Addess, K. J., Gilbert, D. A., Olsen, R. K. and Feigon, J. (1992) *Biochemistry* **31**, 339–350
- Waterloh, K., Olsen, R. K. and Fox, K. R. (1992) *Biochemistry* **31**, 6246–6253
- Fox, K. R. and Waring, M. J. (1987) *Biochim. Biophys. Acta* **909**, 145–155
- Portugal, J., Fox, K. R., McLean, M. J., Richenberg, J. L. and Waring, M. J. (1988) *Nucleic Acids Res.* **16**, 3655–3670
- Fox, K. R. and Kentebe, E. (1990) *Biochem. J.* **269**, 217–221
- McLean, M. J. and Waring, M. J. (1988) *J. Mol. Recogn.* **1**, 138–151
- Waterloh, K. and Fox, K. R. (1991) *Nucleic Acids Res.* **19**, 6719–6724