Structure of Actinomycin D bound with $(GAAGCTTC)_2$ and $(GATGCTTC)_2$ and Its Binding to the $(CAG)_n$:$(CTG)_n$ Triplet Sequence As Determined by NMR Analysis†

Chenyang Lian,§ Howard Robinson,‡ and Andrew H.-J. Wang*,§,‡

Contribution from the Department of Cell and Structural Biology and Department of Chemistry, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801

Received May 15, 1996⊗

Abstract: The anticancer drug actinomycin D (ActD) binds to DNA by intercalating its phenoxazone ring at a GpC step with the drug's two cyclic pentapeptides located in the DNA minor groove. The binding affinity to the GpC site is influenced by the flanking sequences. We have analyzed the structure of the complexes of ActD−$d(GAAGCTTC)_2$ and ActD−$d(GATGCTTC)_2$ by NOE-restrained refinement. Binding of ActD to the −$(AGCT)_2$− sequence causes the *N*-methyl group of MeVal to wedge between the bases at the ApG step, resulting in kinks on both sides of the intercalator site. Surprisingly ActD forms a very stable complex with $d(GATGCTTC)_2$ in which the same methyl group now fits snugly in a cavity at the TpG step created by the T:T mismatched base pair. In contrast, ActD does not stabilize the unstable A:A−mismatched $d(GAAGCATC)_2$ duplex to a significant extent. Such high-resolution structural information helps reveal the sequence preference of ActD toward −XGCY− tetranucleotides. The triplet repeat $(CAG)_n$ and $(CTG)_n$ motifs, which are associated with several genetic diseases such as Huntington's disease/spinobulbar muscular atrophy and myotonic dystrophy, contain −AGCA− and −TGCT− sequences. It was found by NMR spectroscopic studies that ActD significantly stabilizes the mismatched $(CAG)_n$ and $(CTG)_n$ duplexes and prevents them from annealing with each other to form the Watson−Crick duplex. This suggests that ActD may trap the cruciform structure of the $(CAG)_n$/$(CTG)_n$ sequence and may exert certain biological actions (e.g., stopping the expansion during replication), since interference of the equilibrium between the duplex and cruciform structures by proteins or drugs may have biological consequences.

Introduction

Actinomycin D (Figure 1) is a potent anticancer drug. It has been shown that ActD binds strongly to DNA duplexes, thereby interfering with replication and transcription. The sequence specificity of ActD has been analyzed extensively by a variety of methods, including chemical footprinting,[1] NMR,[2] X-ray crystallography,[3] and photoaffinity cross-linking.[4,5] These results suggest that the GpC site is the major preferred binding site, although other sites such as GpG[4,6] have been noted to have an unusual affinity toward ActD. Moreover, the flanking sequences at the GpC binding site play an important role in the binding affinity of ActD.[7]

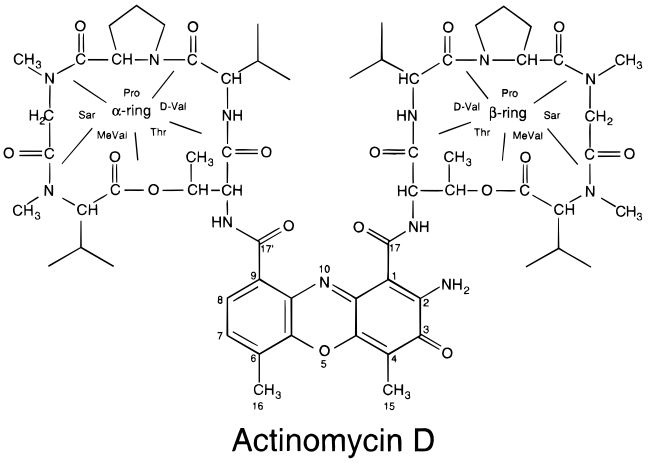

Figure 1. Molecular structure of the actinomycin D (ActD). The atomic numbering on the phenoxazone ring is shown. The amino acids on the quinonoid ring are numbered from Thr_{18} to Val_{22}, and those from the benzenoid ring from Thr_{23} to Val_{27}.

Recently a number of human genetic diseases have been correlated to expansions of triplet DNA sequence repeats. The $(CGG)_n$ repeat in the X-chromosome is responsible for fragile-X syndrome,[8] the $(CAG)_n$ repeat is associated with Huntington's disease and spinobulbar muscular atrophy,[9] and finally the

† Abbreviations used: ActD, actinomycin D; bp, base pair(s); COSY, correlated spectroscopy; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; oligo(s), oligonucleotide(s); ppm, parts per million; RMSD, root mean square deviation; SPEDREF, spectral-driven refinement; 2D, two-dimensional; Pxz, phenoxazone.

* Address for correspondence: Dr. A. H.-J. Wang, Department of Cell and Structural Biology, 506 Morrill Hall, University of Illinois, Urbana, IL, 61801. Tel.: (217)-244-6637. Fax: (217)-244-3181. E-mail: ahjwang@uiuc.edu.

‡ Department of Cell and Structural Biology.

§ Department of Chemistry.

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

(1) (a) Aivasashvilli, V. A.; Beabealashvilli, R. S. *FEBS Lett.* **1983**, *160*, 124−128. (b) Lane, M. J.; Dabrowiak, J. C.; Vournakis, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 3260−3264. (c) Fox, K. R.; Waring, M. J. *Nucleic Acids Res.* **1984**, *12*, 9271−9285. (d) Phillips, D. R.; Crothers, D. M. *Biochemistry* **1986**, *25*, 7355−7362. (e) Goodisman, J.; Rehfuss, R.; Ward, B.; Dabrowiak, J. C. *Biochemistry* **1992**, *31*, 1046−1058.

(2) (a) Patel, D. J. *Biochemistry* **1974**, *13*, 2396−2402. (b) Liu, X.; Chen, H.; Patel, D. J. *J. Biomol. NMR* **1991**, *1*, 323−347. (c) Brown, D. R.; Kurz, M.; Kearns, D. R.; Hsu, V. L. *Biochemistry* **1994**, *33*, 651−664.

(3) (a) Sobell, H. M.; Jain, S. C.; Sakore, T. D.; Nordman, C. E. *Nature (New Biol.)* **1971**, *231*, 200−205. (b) Sobell, H. M.; Jain, S. C. *J. Mol. Biol.* **1972**, *68*, 21−34. (c) Kamitori, S.; Takusagawa, F. *J. Mol. Biol.* **1992**, *225*, 445−456. (d) Kamitori, S.; Takusagawa, F. *J. Am. Chem. Soc.* **1994**, *116*, 4154−4165.

(4) Rill, R. L.; Marsch, G. A.; Graves, D. E. *J. Biomol. Struct. Dynam.* **1989**, *7*, 591−605.

(5) Bailey, S. A.; Graves, D. E.; Rill, R.; Marsch, G. *Biochemistry* **1993**, *32*, 5881−5887.

(6) Bailey, S. A.; Graves, D. E.; Rill, R. *Biochemistry* **1994**, *33*, 11493−11500.

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

S0002-7863(96)01631-9 CCC: $12.00

$(CTG)_n$ repeat is associated with myotonic dystrophy.[10] How these unusual repetitive sequences correlate with the etiology of these diseases has been under intense study.[11] These repeats are found both inside and outside the coding region of known genes. For example, the $(CAG)_n$ repeat is located in the protein coding region of the spinobulbar muscular atrophy (androgen receptor) gene, presumably resulting in a long insertion of $(gln)_n$ peptides sequence into its gene product.[9b] On the other hand, the $(CGG)_n$ repeat is located outside the coding region of the *FMR-1* gene on the 5′ side. Methylation of the CpG sequence in the $(CGG)_n$ repeat, which may affect its structure, plays an important role in the biological consequence of the repeat.[8b]

The mechanism by which those repeats are extended during replication is under intense scrutiny. Some have proposed that a "slippage" process occurs due to the ease of the formation of hairpin structures for these repeating sequences.[12] Indeed several recent studies have shown that certain triplet repeats, e.g., $(CAG)_n$ and $(CTG)_n$, but not $(CGA)_n$, have a strong propensity to form hairpin structures.[13] Therefore DNA duplex encoding the $(CAG)_n$:$(CTG)_n$ repeats may easily exchange between duplex and cruciform, especially under the negative supercoiling strain. If there are proteins or other ligands (e.g., drugs) that can stabilize the stem of the cruciform, this process would be inhibited.

Parenthetically, DNA oligos with sequences related to 5′−CGA adopt a parallel-stranded (PS) double helices with all base pairs of the nonWatson−Crick self-pairing type, i.e., A with A, T with T, G with G, and, finally, C with C^+.[14] Whether this property is related to its inability to form hairpin structures remains to be resolved.

Inspection of the $(CAG)_n$/$(CTG)_n$ triplet repeat reveals the existence of many GpC ActD binding sites. What is the DNA structure associated with the $(CAG)_n$/$(CTG)_n$ triplet sequences in the presence of ActD? In this paper we have studied the structures of the $(CAG)_n$ and $(CTG)_n$ triplets and the binding of ActD to them.

Experimental Section

The oligonucleotides were synthesized on an automated DNA synthesizer at the Genetic Facility of UIUC. Actinomycin D was purchased from Sigma (St. Louis, MO) and dissolved in methanol as stock solutions. The concentration of ActD solutions was determined from its optical density ($\epsilon_{224nm} = 35\,280$). The solutions of various ActD−DNA complexes for NMR studies were prepared by dissolving the ammonium salt of the oligos plus the appropriate amounts of ActD stock solution in 0.55 mL of phosphate buffer solution (20 mM sodium phosphate, pH 7.0) to produce a final duplex concentration of 1−2 mM for octamers and 0.1−0.3 mM for 32mers. For 1-D H_2O spectra, the ActD−DNA complexes solutions were vacuum-dried in a SpeedVac at room temperature. The dry powder then was dissolved in 0.55 mL of 90%H_2O/10%D_2O solution. 1D-NMR spectra were collected using the 1−1 pulse sequence.[15]

For 2D NOESY spectra, each sample of the octamer oligos and their ActD complexes was dried on the SpeedVac first. The dried powder was then dissolved in 0.5 mL of 99.8% D_2O and dried again on the SpeedVac. This step was repeated three times and the sample was dried in an NMR tube with a stream of dry nitrogen gas. Finally, 0.5 mL of 99.96% D_2O (Aldrich, Milwaukee, WI) was added to produce the sample. Both 1D- and 2D-NMR spectra were recorded on a Varian VXR500 500 MHz spectrometer. The chemical shifts (in ppm) are referenced to the HDO peak which is calibrated to 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) at different temperatures. Phase sensitive NOESY spectra were recorded as 512 t_1 increments of 1024 (or 2048) complex points each (in the t_2 dimension) and averaged for 24 scans per FID. The recycle delay was 4.37 s and the mixing time was 100 ms for the NOESY. The 2D-NOESY spectra in H_2O were collected using the 1−1 pulse sequence as the read pulse.[15]

The 2D data sets were processed with the program FELIX version 1.1 (Hare Research, Woodinville, WA) on Silicon Graphics workstations. Linear prediction was used to correct the first data point in t_1. In both the t_1 and t_2 time domains, the NOE data were apodized to reduce truncation artifacts by having the last quarter of the FID smoothly attenuated to zero with a sine bell squared curve. The resulting FID was exponentially multiplied with a constant of 4 Hz. The mixing time 100 ms was selected as it gives an optimum number of NOE observables without the problem of severe spin diffusion or yielding too few measurable NOEs. The inversion recovery experiment determined the T_1 relaxation time for every spin, with an average T_1 of 1.7 s for all protons. The recycle time of 4.37 s is about 2.57 times of the average T_1, amounting to about 92.5% complete recovery of magnetization on average.

All measureable NOE crosspeak integrals have been determined by the program MYLOR and subsequently included in the refinement. The number of unique NOEs above a certain overlap cutoff level can be defined. The definition of the overlap cutoff means, using a 50% cutoff as an example, that for each integral at least 50% of the volume is due to that crosspeak.

Starting models of the octamer DNA GAAGCTTC and GATGCTTC and their 1:1 complexes were built using MIDAS (University of California). The initial duplex conformation of the dinucleotide surrounding the ActD site was that of the crystal structure of the ActD−GAAGCTTC complex.[3d] The ActD, using the atomic coordinates of its high-resolution crystal structure,[3b] was docked into the intercalation cavity. A B-DNA tetranucleotide was overlaid on both ends of the complex. In the case of the ActD−GATGCTTC complex, models incoporating two possible T:T base pair configurations were built. All initial models were energy minimized of using conjugate gradient.

The structure refinement of the octamer DNAs and their ActD complexes has been carried out by the procedure SPEDREF.[16] The force field parameters of ActD were adjusted to conform to its high

(8) (a) Fu, Y.-H.; Kuhl, D. P. A.; Pizzuti, A.; Pierreti, M.; Sutcliffe, S. S.; Richards, S.; Verkerk, A. J. M. H.; Holden, J. J. A.; Fenwick, T. G., Jr.; Warren, S. T.; Oostra, B. A.; Nelson, D. L.; Caskey, C. T. *Cell* **1991**, *67*, 1047−1058. (b) Verkerk, A. J. M. H.; Pieretti, M.; Sutcliffe, J. S.; Fu, Y.-H.; Kuhl, D. P. A.; Pizzuti, A.; Reiner, O.; Richards, S.; Victoria, M. F.; Zhang, F.; Eussen, B. E.; van Ommen, G-J. B.; Blonden, L. A. J.; Riggins, G. J.; Chastain, J. L.; Kunst, C. B.; Galjaard, H.; Caskey, C. T.; Nelson, D. L.; Oostra, B. A.; Warren, S. T. *Cell* **1991**, *65*, 905−914.

(9) (a) The Huntington's Disease Collaborative Research Group. *Cell* **1993**, *72*, 971−983. (b) LaSpada, A. R.; Wilson, E. M.; Lubahn, D. B.; Harding, A. E.; Fischbeck, K. H. *Nature* **1991**, *352*, 77−99.

(10) (a) Brook, J. D.; McCurrash, A. E.; Harley, H. G.; Buckler, A. J.; Church, D.; Aburatani, H.; Hunter, K.; Stanton, V. P.; Thirion, J.-P.; Hudson, T.; Sohn, R.; Zemelman, B.; Snell, R. G.; Rundle, S. A.; Crow, S.; Davies, J.; Shelbourne, P.; Buxton, J.; Jones, C.; Juvonen, V.; Johnson, K.; Harper, P. S.; Shaw, D. J.; Housman, D. E. *Cell* **1992**, *68*, 799−808. (b) Mahadevan, M.; Tsilfidis, C.; Sabourin, L.; Shutler, G.; Amemiya, C.; Jansen, G.; Neville, C.; Narang, M.; Barcelo, J.; O'Hoy, K.; Leblond, S.; Earle-Macdonald, J.; de Jong, P. J.; Wieringa, B.; Korneluk, R. G. *Science* **1992**, *255*, 1253−1255.

(11) (a) Caskey, C. T.; Pizzuti, A.; Fu, Y.-H.; Fenwick, T. G., Jr.; Nelson, D. L. *Science* **1992**, *256*, 784−789. (b) Nelson, D. L.; Warren, S. T. *Nature Genet.* **1993**, *4*, 107−108. (c) Martin, J. B. *Sceince* **1993**, *262*, 674−676. (d) Miwa, S. *Nature Genet.* **1994**, *6*, 3−4. (e) Richards, T. I.; Sutherland, G. R. *Cell* **1992**, *70*, 709−712.

(12) Leach, D. R. F. *BioEssays* **1994**, *16*, 893−900.

(13) (a) Chen, X.; Mariappan, S. S. V.; Catasti, P.; Ratliff, R.; Moyzis, R. K.; Laayoun, A.; Smith, S. S.; Bradbury, E. M.; Gupta, G. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 5199−5203. (b) Mitas, M.; Yu, A.; Dill, J.; Kamp, T. J.; Chambers, E. J.; Haworth, I. S. *Nucleic Acids Res.* **1995**, *23*, 1050−1059. (c) Zhu, L. M.; Chou, S. H.; Xu, J. D; Reid, B. R. *Nature Struct. Biol.* **1995**, *2*, 1012−1017. (d) Smith, G. K.; Jie, J.; Fox, G. E.; Gao, X. L. *Nucleic Acids Res.* **1995**, *23*, 4303−4311. (e) Mariappan, S. V. S.; Catasti, P.; Chen, X.; Ratliff, R.; Moyzis, R. K.; Bradbury, E. M.; Gupta, G. *Nucleic Acids Res.* **1996**, *24*, 784−792. (f) Gacy, A. M.; Goellner, G.; Juranic, N.; Macura, S.; McMurray, C. T. *Cell* **1995**, *81*, 533−540. (g) Mariappan, S. V. S.; Garcia, A. E.; Gupta, G. *Nucleic Acids Res.* **1996**, *24*, 775−783.

(14) Robinson, H.; Wang, A. H.-J. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5224−5228.

(15) Hore, P. J. *J. Magn. Res.* **1983**, *54*, 539−542.

(16) Robinson, H.; Wang, A. H.-J. *Biochemistry* **1992**, *31*, 3524−3533.

Table 1. Refinement Statistics of Two ActD−DNA Complexes

	number of NOE restraints			
	d(GAAGCTTC)$_2$ + ActD		d(GATGCTTC)$_2$ + ActD	
	reliable[a]	total	reliable[a]	total
DNA intraresidue	150	561	210	593
DNA interresidue	279	473	298	489
DNA to ActD	253	650	254	639
ActD to ActD	245	511	263	553
total	927	2195	1025	2274
R-factor (%)	23.5	22.5	24.6	23.3
structure statistics (rmsd)[b]				
NOE violations (Å)	0.32		0.35	
bond deviations of DNA (Å)	0.012		0.012	
bond deviations of ActD (Å)	0.012		0.010	
angle deviations of DNA (deg)	3.4		3.4	
angle deviations of ActD (deg)	2.6		2.7	

[a] Total NOE restraints are all NOEs from the model that are greater than 0.1% of the diagonal for each spin at zero mixing time. Reliable NOEs are defined as those whose overlap is less than 50% for this nearly 2-fold symmetric molecule. [b] Biharmonid potential wells were used for the NOE restraints. Thus, the rmsd of the NOE violations represents the differences between the target distance and the actual distance from the model. For the rmsd's of the bonds and angles the deviations are relative to the parameters as defined in X-PLOR 3.1's parah1e.dna and topallhdg.pro parameter files.

resolution crystal structure.[3b] The program X-PLOR[17] was used for the molecular dynamics simulation and energy minimization. For the first 40 refinement cycles, the molecular complex with NOE restraints were slowly annealed from 300 to 40 K, and then subject to conjugate gradient minimization with NOE-restrains. For the next 20 refinement cycles, the low temperature annealing was deleted from the procedure and the refinement was performed with only NOE-restrained conjugated gradient minimization. The NMR *R*-factor is defined as $R = \sum|N_o - N_c|/\sum N_o$, where N_o and N_c are the experimental and calculated NOE integrals, respectively. The refinement statistics of two structures are listed in Table 1. The atomic coordinates of the refined structures of the 1:1 complexes ActD−GAAGCTTC and ActD−GATGCTTC have been deposited in Brookhaven Protein Databank (Entry ID codes IDSC and IDSD, respectively).

Results and Discussion

Structure of ActD−DNA Complexes. We have chosen the DNA octamer d(GAAGCTTC) as the canonical sequence for NMR structural study, since the crystal structure of its 1:1 complex with ActD has been determined at 3 Å resolution[3c,d] and it may serve as a reference structure for comparison with other structures. We have studied three related DNA sequences, GAAGCTTC, GATGCTTC, and GAAGCATC. Their 1D NMR spectra suggested that the first octamer formed a stable duplex as expected, but the latter two did not (Figure 2 and Figure 1 of the supporting information). The structure of GAAGCTTC has been refined using the NOE-restrained refinement procedure[16] (data not shown) and it is similar to the canonical B-DNA, in contrast to its crystal structure being an A−DNA.[3d]

Addition of ActD to solutions of the DNA octamers caused extensive changes in their NMR spectra. Figure 2 shows the titration of ActD to DNA solutions as monitored by the imino proton resonances. This region of the spectra was chosen for two reasons. First, since the resonances are located far downfield from most other resonances (i.e., overlapping less likely), they can be conveniently used to follow the titration of ActD. Second, the appearance of the imino proton resonances (e.g., peak height and sharpness) provides information on the stability and dynamics of the DNA duplexes. It can be seen that binding of ActD to GpC−containing DNA sequences stabilized their duplex structure to varying degrees, depending on the flanking sequences.

(17) Brunger, A. T. X-PLOR, 1993, version 3.1, The Howard Hughes Medical Institute and Yale University, New Haven, CT.

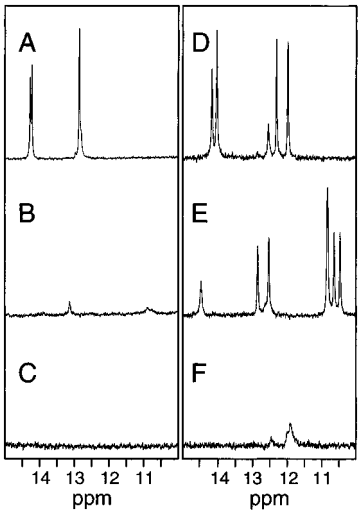

Figure 2. The imino proton region of the proton 1D NMR spectra (2 °C) from three DNA octamers and their complexes with ActD. The left panels are from DNA alone and the right panels are from the DNA−ActD complexes. (A) GAAGCTTC, (B) 1:1 ActD−GAAGCTTC complex, (C) GATGCTTC, (D) 1:1 ActD−GATGCTTC complex, (E) GAAGCATC, and (F) 1:1 ActD−GAAGCATC complex.

The 1:1 complexes of ActD−GAAGCTTC and ActD−GATGCTTC have well-resolved 2D-NOESY NMR spectra (Figure 3) for which resonance assignment and refinement have been carried out using the procedure developed in this laboratory.[16]

ActD−GAAGCTTC Complex. Figure 3A shows the aromatic-H1′ region of the nonexchangeable proton 2D-NOESY spectrum from which the sequential assignment was deduced. Since ActD does not possess an exact 2-fold symmetry, its binding to a palindromic DNA causes both the ActD and DNA protons to become nonequivalent (thus having different chemical shifts). For example, the chemical shifts of C_5H1′ (6.16 ppm) and C_{13}H1′ (6.21 ppm) are split apart by 0.05 ppm. In fact the identity of these two protons was made by the different NOE crosspeak intensities to the drug protons. For example, PxzH8 has significant NOEs to C_{13}H1′ and G_{12}H1′, but not to the respective C_5H1′ and G_4H1′. Thus the orientation of ActD in the DNA intercalation cavity is unambiguously defined. Extension of the assignment into other regions could be made systematically, as exemplified by the aromatic to H2′/H2″ region (Figure 2 of supporting information) in which a number of

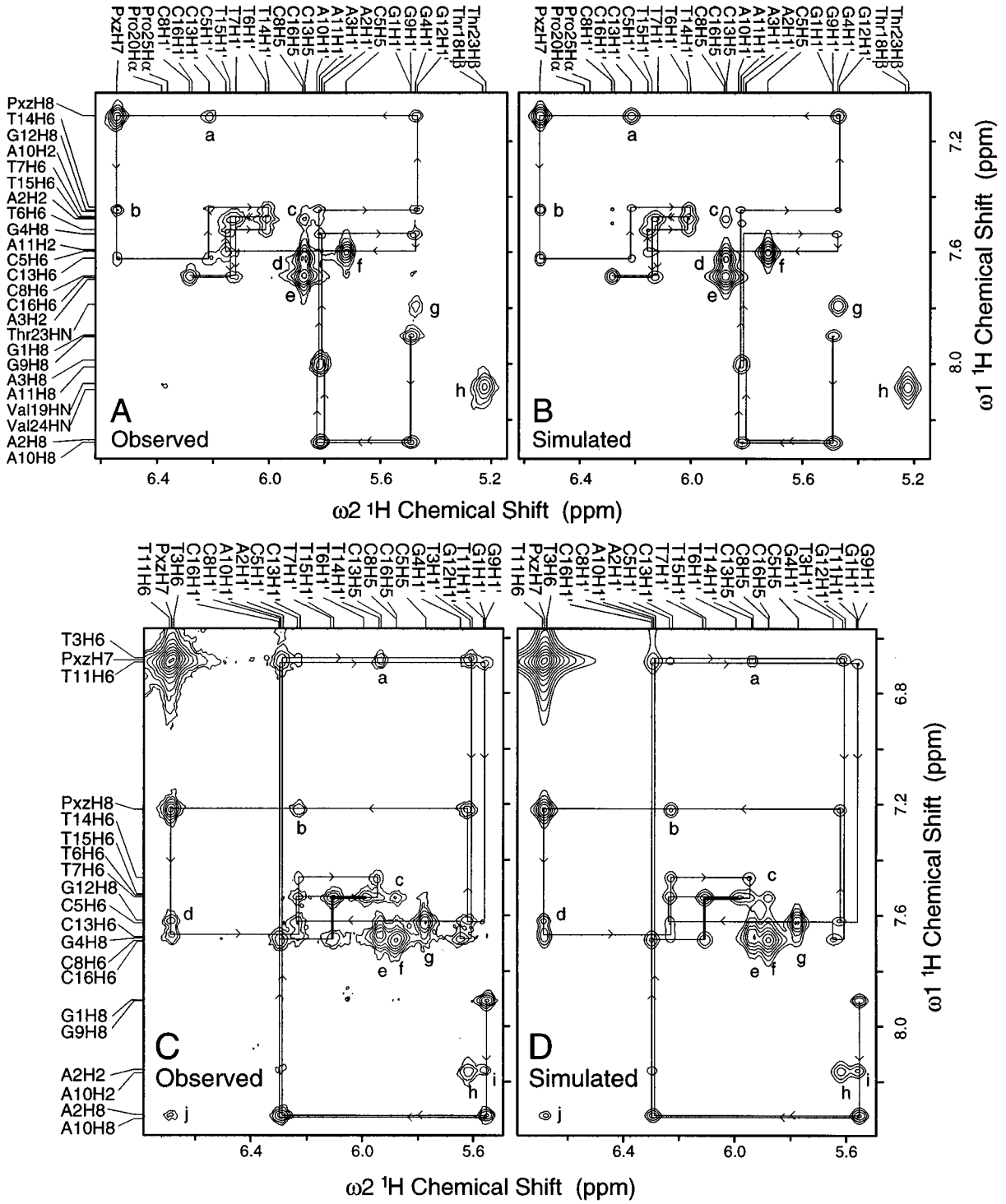

Figure 3. Experimental and simulated 2D-NOESY spectra of (A and B) the 1:1 complex of ActD−GAAGCTTC duplex and (C and D) the 1:1 complex of ActD−GATGCTTC. The aromatic-H1′/H5 region is shown, with the sequential assignment pathway indicated. The simulated spectra were calculated using a full spin relaxation matrix based on the refined structures as described in the text. Note that the spectra from both 1:1 complexes have two assignment pathways. This is due to the asymmetry of ActD phenoxazone (Pxz) chromophore. The nucloetides in the first strand were labeled numerically from 1 to 8. Likewise, the ones in the second strand were labeled from 9 to 16. The phenoxazone ring was oriented in the way that H7, H8 are near C_{13}, T_{14} in the second strand. Those peaks that are not in the pathways are labeled alphabetically: In parts A and B, **a**, PxzH8−C_{13}H1′; **b**, PxzH7−G_{12}H8; **c**, T_7H6−C_8H5; **d**, C_{13}H5−C_{13}H6; **e**, C_8H5−C_8H6 and C_{16}H5−C_{16}H6; **f**, C_5H5−C_5H6; **g**, Thr_{23}HN−G_4-H1′; **h**, Val_{19}HN−Thr_{23}Hβ and Val_{24}HN−Thr_{11}Hβ. In parts C and D, **a**, PxzH7−C_{13}H5; **b**, PxzH8−C_{13}H1′; **c**, T_7H6−C_8H5; **d**, PxzH7−G_{12}H8; **e**, C_{13}H5−C_{13}H6; **f**, C_8H5−C_8H6; **g**, C_5H5−C_5H6; **h**, A_2H2−T_3H1′; **i**, A_{10}H2−T_{11}H1′; **j**, A_2H8−T_3H6 and A_{10}H8−T_{11}H6.

informative NOE crosspeaks are found. Notably, A_3H2 to $MeVal_{22}$NMe and A_{11}H2 to $MeVal_{27}$NMe crosspeaks are very intense, suggesting that the *N*-methyl group of MeVal nudges tightly against the floor of the AT−base pair in the minor groove.

The assignment of the exchangeable protons was further carried out. Figure 4 shows the amide/amino protons to aromatic/H1′/H5 regions of the 2D-NOESY spectra. Note that the G_4N2/G_{12}N2 amino protons are detected as well-resolved peaks, suggesting that the rotation about the *C*2−N2 bond in these two guanines is slow. This is consistent with the fact that the G-N2 amino group is involved in the hydrogen bonding interaction with the peptide carbonyl of the Thr residue from ActD. Some amide NH protons in the complex exchange with

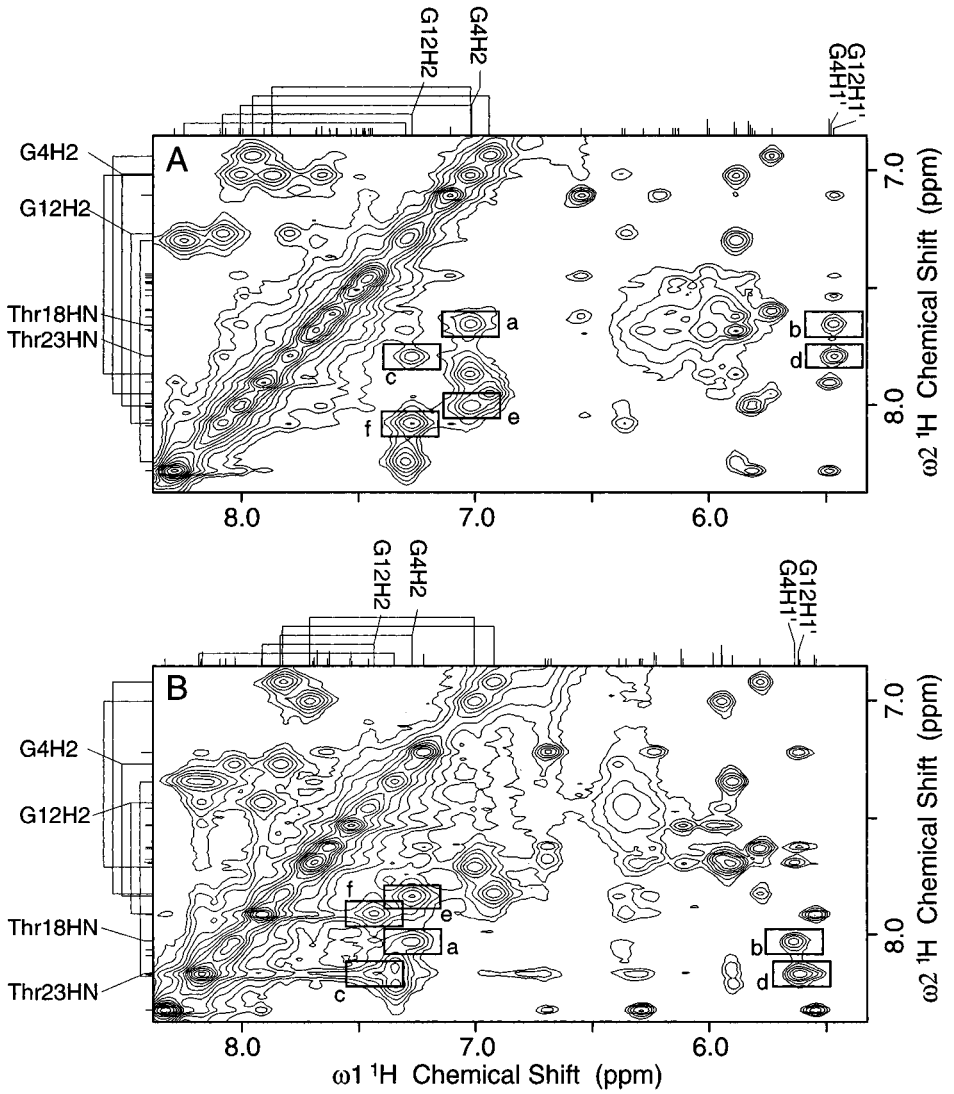

Figure 4. The imino/amide to aromatic/H1′/H5 crosspeak region of the proton 2D-NOESY spectra (2 °C) from (A) the 1:1 ActD−GAAGCTTC complex and (B) the 1:1 ActD−GATGCTTC complex. The marked key crosspeaks in both spectra are a, $Thr_{18}HN-G_4H2b$; b, $Thr_{18}HN-G_4H1'$; c, $Thr_{23}HN-G_{12}H2b$; d, $Thr_{23}HN-G_{12}H1'$; e, G_4H2a-G_4H2b; f, $G_{12}H2a-G_{12}H2b$. They are consistent with the hydrogen-bonding interactions between the Thr amide group of ActD and the N2,N3 of guanine of DNA in the minor groove.

H_2O extremely slowly. For example, both Thr NH protons remained unexchanged to a great extent in D_2O for days in the NMR tube. This is evident in Figure 3A, where peak g is the crosspeak between $Thr_{23}HN$ and G_4H1'.

Table 2 lists the chemical shifts of all nonexchangeable resonances and most of the exchangeable resonances of ActD and DNA. A total of 927 reliable restraints were derived for the NOE refinement which resulted in an NMR *R*-factor of 23.5%. Some of the important intermolecular NOE crosspeaks between ActD and DNA are listed in Table 3. The refined structure has an RMSD of 2.77 Å with the starting model.

The refined structure (Figure 5A) shows that the ActD binds in and significantly widens the minor groove. ActD in the complex maintains its pseudo-2-fold symmetry to a large extent, with the RMSD between the two penptadepsipeptide rings being 1.56 Å. There are two interpeptide hydrogen bonds between NH of D-Val on one peptide ring and the carbonyl C═O of D-Val on the other peptide ring which provide conformational rigidity to ActD. Such strong hydrogen bonds explain the slow exchange rate of the Val_{19}/Val_{24} amide protons, as seen in Figure 3A (peak h). Several strong intermolecular ActD−DNA hydrogen bonds are found, including those between the $G-NH_2$ and the carbonyl C═O of the Thr residue which defines the sequence specificity for the GpC step. The amino group from the Pxz aglycon ring causes the backbone of the $G_{13}-C_{14}$ strand to move by ∼2.0 Å. This amino group is in a good position to form a hydrogen bond to the O4′ of C_5 (N···O4′ distance 2.7 Å). The offset position of the ActD in the GpC intercalation cavity makes the complex no longer 2-fold symmetrical, consistent with the observed NMR data. In both complexes the G−C base pairs next to the aglycon have large buckles, similar to those seen in other complexes of intercalators (e.g., daunorubicin and nogalamycin) and DNA.[18]

The local structure of ActD−$(GpC)_2$ at the binding site in the 1:1 ActD−GAAGCTTC complex is in general similar to that from two recent crystal structures, one from the *C*2 space group[3c] and the other from the *F*222 space group,[3d] although several important deviations were found. In the *C*2 crystal, the entire 1:1 ActD−GAAGCTTC complex is in the asymmetric unit of the crystal unit cell. Interestingly the ActD in the complex of the *C*2 form is significantly deviated from the pseudo-2-fold symmetry with the two pentapeptide rings adopting very different conformations. This is likely due to crystal packing, since the NOE data did not reveal such a large difference. In the *F*222 crystal, however, only half of the 1:1 ActD−GAAGCTTC complex is in the asymmetric unit of the

(18) Wang, A. H.-J. *Current Opin. Struct. Biol.* **1992**, *2*, 361−368.

Table 2. Chemical Shifts (ppm) for d(GAAGCTTC)$_2$ + ActD and d(GATGCTTC)$_2$ + ActD complexes at 2 °C

d(GAAGCTTC)$_2$ + Act−D

	H5/Me/H2	H8/6	H1′	H2′	H2″	H3′	H4′	H5′	H5″	H1/3	H2/4/6(a)[a]	H2/4/6(b)[a]
G1		7.9	5.49	2.53	2.70	2.53	4.84	4.15	3.66	12.54		
A2	7.50	8.28	5.81	2.79	2.89	5.06	4.38	4.15	4.04		na	na
A3	7.70	7.99	5.82	2.35	2.35	5.01	4.29	4.16	4.16		na	na
G4		7.53	5.48	2.41	2.65	5.00	3.79	4.08	3.89	11.98	8.01	7.02
C5	5.72	7.60	6.16	2.35	2.35	4.63	4.23	4.15	3.70		7.95	6.94
T6	1.73	7.52	6.01	2.13	2.56	4.86	3.85	4.23	4.09	14.02		
T7	1.81	7.48	6.12	2.20	2.52	4.86	4.14	4.08	4.08	14.17		
C8	5.88	7.69	6.28	2.30	2.30	4.61	4.02	4.15	4.03		8.24	7.30
G9		7.90	5.49	2.53	2.70	4.15	4.15	3.66	3.66	12.54		
A10	7.44	8.28	5.82	2.80	2.91	5.06	4.38	4.15	4.04		na	na
A11	7.60	8.01	5.82	2.44	2.44	5.01	4.32	4.15	4.15		na	na
G12		7.45	5.47	2.47	2.64	5.01	3.76	4.09	3.97	12.31	8.08	7.27
C13	5.87	7.62	6.21	1.95	2.56	4.50	3.96	4.25	3.95		7.87	7.02
T14	1.75	7.44	6.01	2.09	2.62	4.85	3.86	4.24	4.08	14.02		
T15	1.80	7.48	6.13	2.20	2.51	4.87	4.14	4.08	4.08	14.17		
C16	5.88	7.69	6.28	2.30	2.30	4.61	4.02	4.15	4.03		8.24	7.30

	HN/MeN	Ha	Ha	Hb	Hb	Hg/Me	Hg/Me	Hd	Hd	H7	H8	H4Me	H6Me
Pxz17										6.54	7.11	1.50	1.85
Thr18	7.66	4.54		5.22		1.37							
Val19	8.07	3.65		2.15		1.05	0.83						
Pro20		6.38		3.35	1.99	2.27	2.05	3.94	3.66				
Sar21	2.96	4.36	4.44										
Mva22	2.84	2.81		2.52		0.92	0.92						
Thr23	7.79	4.75		5.22		1.38							
Val24	8.09	3.64		2.16		1.05	0.83						
Pro25		6.36		3.25	2.13	2.27	2.04	3.87	3.66				
Sar26	3.00	4.40	4.55										
MeVal27	2.93	2.82		2.53		0.92	0.92						

d(GATGCTTC)$_2$ + ActD

	H5/Me/H2	H8/6	H1′	H2′	H2″	H3′	H4′	H5′	H5″	H1/3	H2/4/6(a)[a]	H2/4/6(b)[a]
G1		7.91	5.55	2.56	2.70	4.83	4.12	3.66	3.66	na	na	na
A2	8.16	8.32	6.29	2.78	2.86	5.00	4.45	4.128	4.08		7.46	6.36
T3	1.68	6.67	5.61	0.61	1.25	4.71	3.85	4.28	4.02	10.83		
G4		7.68	5.65	2.60	2.66	4.84	3.90	4.12	3.92	12.52	7.84	7.27
C5	5.78	7.63	6.23	2.37	2.46	4.66	4.24	4.18	3.78		7.82	6.92
T6	1.66	7.53	5.99	1.98	2.53	4.89	3.89	4.22	4.08	10.63		
T7	1.84	7.53	6.10	2.24	2.49	4.90	4.11	4.08	3.97	14.47		
C8	5.88	7.69	6.30	2.30	2.30	4.60	4.03	4.18	4.04		8.21	7.34
G9		7.91	5.55	2.56	2.70	4.83	4.12	3.66	3.66	na	na	na
A10	8.16	8.32	6.29	2.78	2.86	5.00	4.48	4.18	4.08		7.46	6.36
T11	1.64	6.70	5.56	0.59	1.18	4.60	3.92	4.28	3.97	10.81		
G12		7.62	5.62	2.61	2.73	4.80	3.80	3.97	4.18	12.84	7.91	7.43
C13	5.94	7.67	6.23	2.01	2.60	4.54	4.00	4.28	4.01		7.71	7.01
T14	1.64	7.46	5.95	1.89	2.56	4.85	3.91	4.24	4.07	10.46		
T15	12.84	7.54	6.11	2.24	2.49	4.90	4.11	4.08	3.97	14.47		
C16	5.88	7.69	6.30	2.30	2.30	4.60	4.03	4.18	4.04		8.15	7.35

	HN/MeN	Hα	Hα	Hβ	Hβ	Hγ/Me	Hγ/Me	Hδ	Hδ	H7	H8	H4Me	H6Me
Pxz17										6.68	7.22	1.56	1.92
Thr18	7.97	4.61		5.21		1.40							
Val19	8.07	3.66		2.16		0.83	1.05						
Pro20		6.40		3.36	1.96	2.26	2.06	4.00	3.70				
Sar21	2.91	4.30	4.36										
Mva22	2.79	2.93		2.47		0.81	0.92						
Thr23	8.16	4.80		5.24		1.40							
Val24	8.09	3.65		2.15		0.83							
Pro25		6.36		3.22	2.07	2.26	2.06	3.90	3.70				
Sar26	2.94	4.33	4.46										
MeVal27	2.82	2.97		0.82	0.93								

[a] H2/4/6(a) are base-pair hydrogen bonded amino protons, H2/4/6(b) are not.

crystal unit cell, forcing the complex to adopt an exact 2-fold symmetry due to the crystallographic disorder. This is again likely due to crystal packing, since the NOE data show that the two strands of DNA are not equivalent. The RMSD between the NMR structure and the *F*222 crystal structure is 2.47 Å.

ActD−GATGCTTC Complex. It is interesting to note that ActD significantly stabilizes the GATGCTTC duplex in which two T:T mismatched base pairs are incorporated (Figure 2C,D). The structure of the 1:1 ActD−GATGCTTC complex has been similarly determined by NOE refinement using 1025 reliable restraints with an NMR *R*-factor of 24.6% (Table 1). Two initial models incorporating different T:T base pair configuraions were refined and they converged to essentially the same structure

Table 3. Largest observed NOE[a] for each ActD spin to $d(GAAGCTTC)_2$ and $d(GATGCTTC)_2$

			$d(GAAGCTTC)_2$				
Pxz17	H8	C13	H1′	2.22	G12	H2″	6.05
	H7	C13	H5	1.80	G12	H2′	2.28
	Me6	C13	H5	0.44	G12	H8	0.36
	Me4	G4	H8	0.29	C5	H5	0.19
		To	18−22		To	23−27	
Thr	Hα	G4	H1′	0.66	G12	H1′	0.82‘
18/23	Hβ	C5	H4′	0.58	G12	H1′	0.64
	Meγ	G4	H1′	2.32	G12	H1′	1.78
Val	Hα	T14	H1′	0.58	T6	H1′	0.27
19/24	Meγ1				T6	H1′	0.28
	Meγ2	T15	H5′	0.15	T7	H5′	0.12
Pro	Hα	C13	H1′	1.33	C5	H1′	1.32
20/25	Hβ1	C13	H1′	5.08	C5	H1′	4.69
	Hβ2	T14	H4′	4.91	C5	H1′	4.63
	Hγ1	C13	H4′	5.13	C5	H4′	9.21
	Hγ2	C13	H4′	6.20	T6	H5′	7.38
	Hδ1	C13	H1′	1.60	C5	H4′	2.32
	Hδ2	C13	H4′	1.65	C5	H4′	1.90
Sar	MeN	T15	H5′	1.77	T7	H5′	1.76
21/26	Hα1	T14	H1′	2.37	A11	H2	1.32
	Hα2	T14	H1′	1.86	T6	H1′	1.34
MeVal	MeN	A3	H2	2.38	A11	H2	2.59
22/27	Hα	G4	H4′	3.90	G12	H4′	2.89
	Hβ	G4	H4′	1.04	G12	H4′	0.43
	Meγ1	T15	H4′	0.25	G12	H4′	1.09
	Meγ2	G4	H5′	1.43	T7	H1′	0.37
			$d(GATGCTTC)_2$				
Pxz17	H8	C13	H1′	2.39	G12	H2″	5.52
	H7	C13	H6	2.41	G12	H2′	2.35
	Me6	C13	H5	0.49	G12	H8	0.41
	Me4	G4	H8	0.27	C5	H5	0.15
		To	18−22		To	23−27	
Thr	Hα	G4	H1′	0.93	G12	H1′	0.82
18/23	Hβ	G4	H1′	0.99	G12	H1′	0.69
	Meγ	G4	H1′	3.19	G12	H1′	1.79
Val	Hα	T14	H1′	0.60	T6	H1′	0.62
19/24	Meγ1	T14	H1′	0.13	T6	H1′	0.22
	Meγ2	G4	H1′	0.13	T6	H1′	0.18
Pro	Hα	C13	H1′	1.09	C5	H1′	1.67
20/25	Hβ1	C13	H1′	5.14	C5	H1′	4.54
	Hβ2	T14	H4′	4.63	C5	H1′	5.14
	Hγ1	C13	H4′	3.60	T6	H5′	5.99
	Hγ2	C13	H4′	4.28	T6	H5′	9.75
	Hδ1	C13	H4′	0.86	C5	H4′	1.39
	Hδ2	C13	H4′	.09	T6	H5′	2.47
Sar	MeN	T14	H1′	1.24	T6	H1′	1.51
21/26	Hα1	T14	H1′	2.01	T6	H1′	1.37
	Hα2	T14	H1′	1.26	T6	H1′	1.40
MeVal	MeN	T3	H1′	1.27	T11	H1′	1.10
22/27	Hα	G4	H5″	5.09	G12	H5′	5.82
	Hβ	G4	H5″	3.56	G12	H4′	0.47
	Meγ1	G4	H5″	0.75	G12	H5′	1.13
	Meγ2	G4	H5″	2.65	G12	H4′	0.94

[a] NOE intensities are percents of the scaled total for that spin at zero mixing time at complete relaxation. The NOE is the average of the two observables. The average cytosine H5−H6 NOE intensity is 8.94%. Methyl intensities are reported per spin, thus the intensity is divided by 3. NOE intensities between 0.1 and 0.2 are weak, between 0.21 and 1.0 are medium, and above 1.0 are strong.

with an RMSD of 0.85 Å bewteen the final two refined structures. Representative intermolecular drug−DNA NOE crosspeaks are listed in Table 3. The agreement between the experimental and simulated 2D-NOESY spectra (Figure 3B) is excellent, supporting a well-refined structure (Figure 5B) which shows again that the ActD binds in and significantly widens the minor groove. Similar strong intermolecular ActD−DNA hydrogen bonds as those of the ActD−GAAGCTTC complex are found.

All four T−N3 imino protons in the complex are clearly detected (Figure 2E), suggesting that the T:T base pairs are uniquely defined as far as the pairing configuration relative to the ActD position is concerned. T_3O2 and T_3N3 are hydrogen bonded to T_{14}N3 and T_{14}O4, respectively, whereas T_6O4 and T_6N3 are hydrogen bonded to T_{11}N3 and T_{11}O2, respectively. The T:T base pairing conformation places the H2′/H2″ of T_3 and T_{11} directly over the guanine base ring of G_4 and G_{12}, respectively. The strong ring current effect of the guanine bases causes the chemical shift of H2′ of T_3/T_{11} to move extremely upfield to 0.61/0.59 ppm, fully consistent with the observed data (Table 2). Similarly, the chemical shift of H2 of A_2/A_{10} moves far downfield to 8.16 ppm, due to the deshielding effect resulting from the destacking of the adenine bases from the T_3/T_{11} bases.

Our structures may offer a plausible explanation for the binding preference of ActD to −XGCY− sequences embedded in ATA−(XGCY)−TAT: TGCA > AGCT ∼ CGCG >> GGCC.[7] Figure 6 shows the close-up views of the complexes at the ActD intercalation site. Note that in the complex the outer edges of the peptide rings (i.e., the Sar-MeVal dipeptide part) from ActD reach the minor groove side of the flanking base pairs. In the ActD−GAAGCTTC complex, the *N*-methyl of MeVal is in very close contact with the base A_3, forcing the A_3pG_4 step to be wedged open by the methyl group, as evident from the very strong NOE crosspeaks of A_3H2-MeVal$_{22}$MeN (and similarly A_{11}H2-MeVal$_{27}$MeN) (Table 3). Such a wedging interaction by a methyl group is reminiscent of the observation in complexes of several HMG−box proteins (e.g., SRY, LEF-1) and DNA.[19]

When the −AGCT− sequence is switched to −TGCA−, the van der Waals clashes between adenine base and MeVal-MeN found in −AGCT− become diminished in −TGCA−, making it energetically more favorable. Such a property may explain why −TGCA− is a slightly better binding sequence for ActD than for −AGCT−. However, through our computer modeling studies, in both the −CGCG− and −GGCC− sequences, the N2 amino groups from the flanking G:C base pairs are shown to have serious van der Waals clashes with the *N*-methyl of MeVal since they protrude into the minor groove, resulting in greater instability. However, ActD binds nearly equally well to −CGCG− and −AGCT−,[7] seemingly contradicting the modeling studies. One possible reason may be related to the sequence context in which the tetranucleotides are embedded. In the decamer ATA−(CGCG)−TAT, ActD may actually bind to the −CGTA−, not the −CGCG−, site. Here the loss and the gain of the binding energies due to the flanking sequences compensate each other so that ActD has a binding affinity for −CGTA− comparable to that for −AGCT−. In the decamer ATA−(GGCC)−TAT, no alternative ActD binding site, except the −GGCC− site, exists. Consequently, ActD has the least binding affinity with −GGCC− due to the unfavorable G/C flanking sequences.

In the complex of ActD−GATGCTTC, the unique T:T base pair conformation generates a cavity in the minor groove (Figure 6B) where the *N*-methyl group of MeVal is tucked snugly against the floor of the T:T base pair, without the TpG step being wedged open. Such a tight fit apparently provides an increased stability to the binding of ActD to the −TGCT− sequence. In contrast, the A:A mismatched base pair is known to be very destabilizing to B-DNA. Addition of ActD to GAAGCATC did slightly stabilize the duplex as judged by the weak imino proton signals (Figure 1F). However, the poor

(19) Werner, M. H.; Gronenborn, A. M.; Clore, G. M. *Science* **1996**, *271*, 778−784.

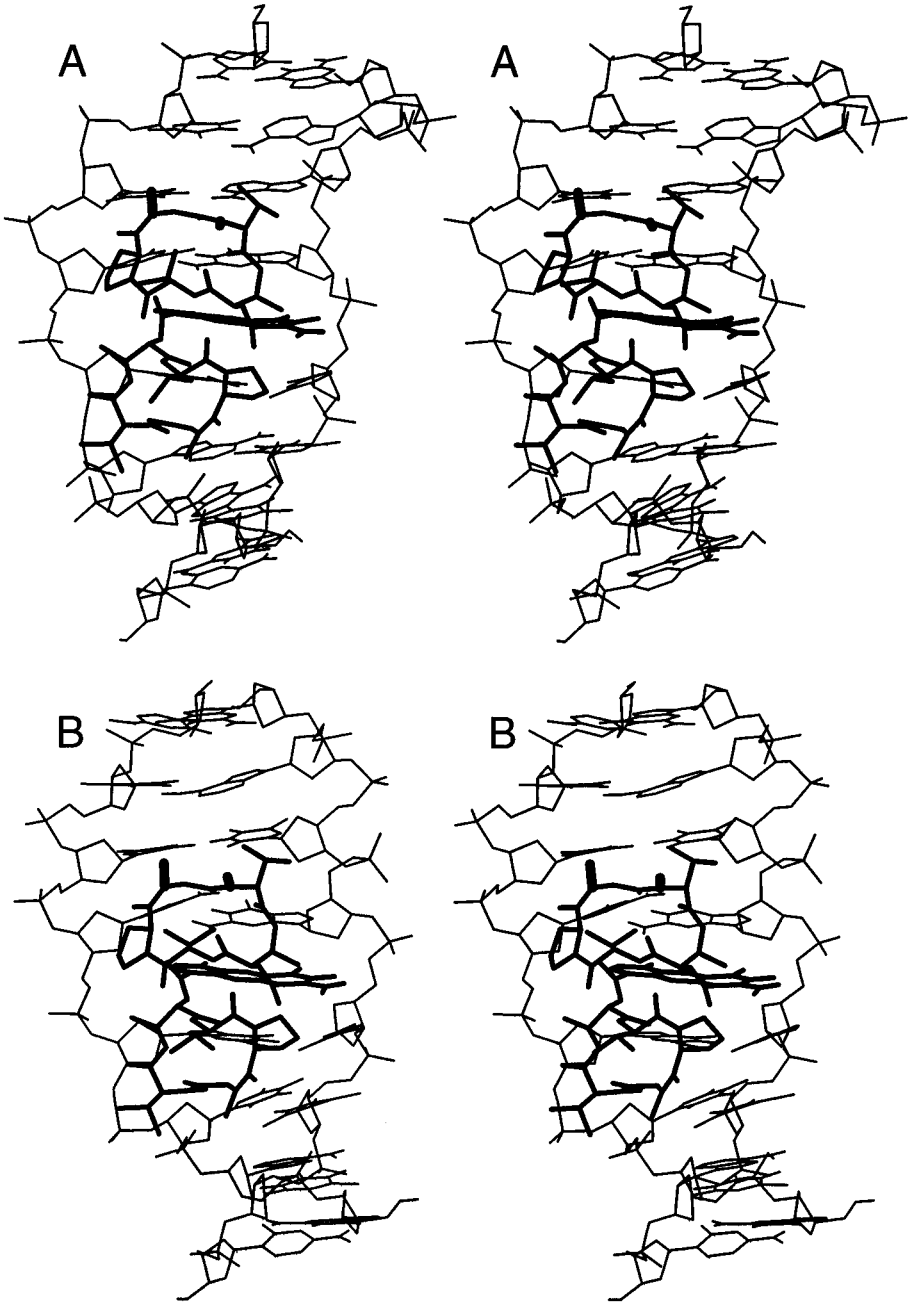

Figure 5. The refined structure of the 1:1 complex of (A) ActD−GAAGCTTC and (B) ActD−GATGCTTC.

quality of its 2D-NOESY spectrum did not afford a reliable structural analysis.

ActD Binds to $(CAG)_n$ and $(CTG)_n$ Sequences

Inspection of the $(CAG)_n$ and $(CTG)_n$ triplet repeat sequences suggests that they have numerous ActD binding sites of −TGCT− and −AGCA−, thus ActD should have significant effects on the structure and stability of these triplet sequences. As a first step, the structures associated with two DNA octamers containing two repeats of each triplet, GCTGCTGC and GCAGCAGC, have been investigated by ^{1}H NMR. Their 1D- and 2D-NMR spectra suggest that both octamers form duplex structures (Figure 7A,B, bottom traces). However they have different stability. The GCTGCTGC octamer has a well-formed duplex, as evident from the sharp resonances and the clear imino proton signals (Figure 7A), but the GCAGCAGC does not. Two resonances at 13.15 and 13.27 ppm of the GCTGCTGC spectrum are from G_4 and G_7 NH1 protons, respectively, whereas the other more upfield resonances at 10.98 and 10.96 ppm are from T_3- and T_6NH3 protons, respectively. The observation of the latter T imino resonances suggests that the two T's are base-paired with each other.

2D-NOESY spectra of both GCTGCTGC (Figure 3 of the supporting information, aromatic to H5/H1′ region) and GCAGCAGC (data not shown) octamers were collected. The spectrum from GCTGCTGC is of sufficient quality to permit unequivocal assignment of all resonances and the measurement of 665 crosspeak integrals. Refinement of the B-DNA model resulted in an NMR *R*-factor of 18.6%. The refined structures (Figure 4 of the supporting information) are in the B-DNA family with most of the sugar puckers in the S-type conformation. T_3O2 and T_3N3 are hydrogen bonded to T_{14}N3 and T_{14}O4, respectively, whereas T_6O4 and T_6N3 are hydrogen bonded to T_{11}N3 and T_{11}O2, respectively. The NMR spectrum of GCAGCAGC had broad resonances which made the interpretation of the 2D-

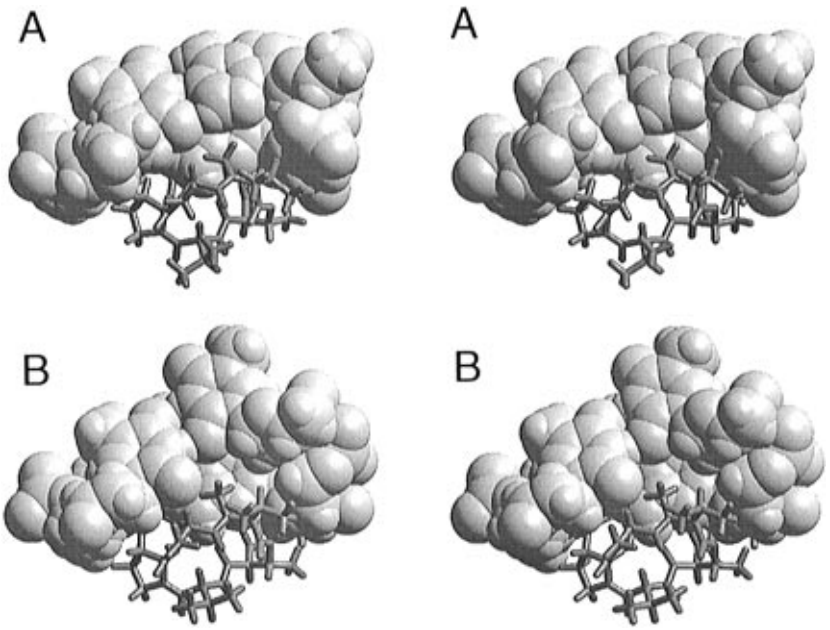

Figure 6. The close-up view of the refined structure of the 1:1 complex at the ActD intercalation site. Only one-half of ActD and the A_3pG_4:$C_{13}pT_{14}$ (or T_3pG_4:$C_{13}pT_{14}$) base pair steps are shown. (A) ActD−GAAGCTTC and (B) ActD−GATGCTTC.

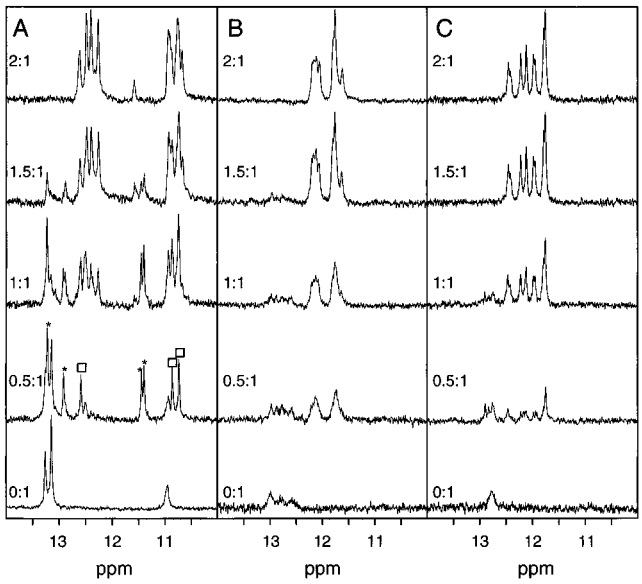

Figure 7. Titration of ActD with three different DNA octamers: (A) GCTGCTGC, (B) GCAGCAGC and (C) GCAGCAIC. Addition of more ActD to the 2:1 ActD−DNA complexes did not change their spectra, suggesting that the octamer can only accommodate two ActDs, despite the three GpC sites.

NOESY spectrum difficult. Nonetheless weak imino proton-resonances were detected (Figure 7B), supporting the formation of a less stable duplex structure.

Addition of ActD to either GCTGCTGC or GCAGCAGC solutions caused extensive changes in their NMR spectra. Figure 7 shows the titration of ActD to DNA solutions as monitored by the imino proton resonances.

ActD−GCTGCTGC Complex. The free DNA forms a duplex as discussed above. At a 0.5:1 (ActD/duplex) ratio, new resonances from the ActD−DNA complexes appear and coexist with the free DNA resonances, indicating that the complexes and free DNA are in slow exchange on the NMR time scale. Close inspection of the spectrum suggests that there are at least two sets of resonances arising from the complexes. One set (marked with *) is likely from the 1:1 complex with the ActD bound at the center GpC sequence. This interpretation is supported by the fact that this set of resonances disappears at higher ActD/DNA ratios. Another set (labeled with the squares) may arise from the ActD−DNA complexes with the ActD bound at the GpC step at the ends of the DNA duplex.

At a 2:1 ratio, the spectrum is significantly simplified with two well-resolved sets of resonances, centered around 12.45 ppm for the G:C base pairs and 10.81 ppm for the T:T base pairs. This is due to the presence of a 2:1 ActD:DNA complex in which the ActDs are bound at the two outer GpC steps. Since ActD has only a quasi 2-fold symmetry, the binding of ActD to a GpC site will destroy the local 2-fold symmetry associated with the palindromic GpC sequence, causing the two G imino protons to have slightly different chemical shifts. As there are two ActDs in the complex in which each ActD has two possible binding orientations, there are three unique combinations of the 2:1 complexes (see Figure 5 of supporting information). Consequently each DNA proton in the complex will appear as four resonances, though they will have similar chemical shifts. In Figure 7A, the four resonances centered around 12.45 ppm are from G_4 and G_7 imino protons and those around 10.81 ppm are from T_3 and T_6 imino protons. Figure 5 of supporting information presents a schematic drawing to illustrate the observations of multiple imino proton resonances.

Further addition of ActD to the solution of the 2:1 complex did not have any effect on the spectrum, suggesting that the octamer duplex can only bind a maximum of **two** ActD molecules. In other words, the two nearest neighboring GpC sites (separated by a T:T base pair) cannot bind ActD simultaneously, namely, an exclusion effect. As noted in Figures 5B and 6B, the outer edges of the peptide rings (i.e., the Sar-MeVal dipeptide part) from ActD reach the minor groove side of the T:T base pairs in the 1:1 ActD−GATGCTTC complex. This explains the observation that two ActD molecules cannot bind to the two adjacent GpC sites simultaneously in a sequence like TGCTGCT, since there would be severe van der Waals clashes between two neighboring bound ActDs. There is a small resonance at 11.58 ppm the identity of which has not been resolved, and its likely from other conformations, e.g., the hairpins.[13]

ActD−GCAGCAGC Complex. The free DNA duplex of GCAGCAGC is significantly less stable, as evident from the

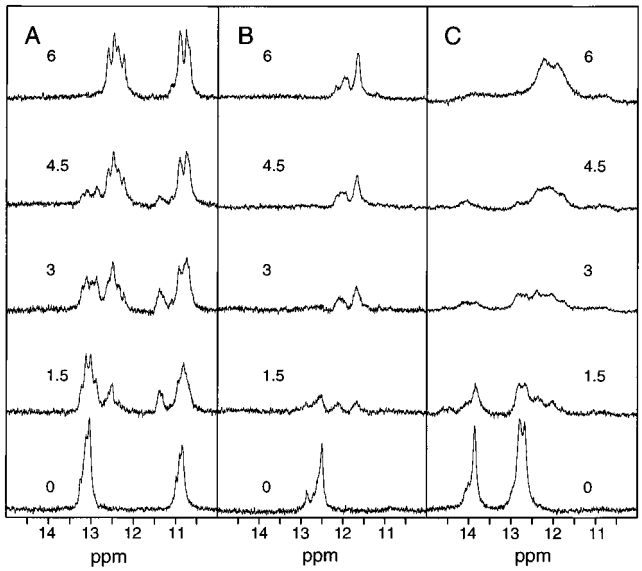

Figure 8. Titration of ActD with three different DNA 32mer duplexes: (A) $G(CTG)_{10}C$. (B) $G(CAG)_{10}C$ and (C) $G(CAG)_{10}C + G(CTG)_{10}C$. Addition of more ActD to the 6:1 ratio of the ActD−DNA complexes did not change the spectra, suggesting that the 32mers can only accommodate ~6 ActD, despite the 11 GpC sites. Note that in part C the A:T imino resonances vanish at high ActD/DNA ratio, suggesting that ActD destabilizes the duplex.

weak imino proton resonances near 12.5−13.0 ppm (Figure 7B). At a 0.5:1 (ActD/duplex) ratio, new resonances at 11.75 and 12.16 ppm from the ActD−DNA complexes appeared and coexisted with the free DNA resonances, again indicating that the complexes and free DNA are in slow exchange on the NMR time scale (Figure 7B). At higher ActD/DNA ratios, the spectrum from the complex remains more or less the same, while the resonances from free DNA diminished significantly. We interpret this as follows. The GpC sequence embedded between two A:A base pairs is not a good ActD binding site, i.e., AGCA is a less favorable site than TGCT. In fact, AGCA is even less favorable than either the AGC(null) or the (null)GCA sites. Therefore, the central AGCA sequence in GCAGCAGC is mostly not occupied at any ActD concentration.

To assess the sequence preference of ActD, another octamer GCAGCAIC (I = inosine) was prepared and a similar ActD titration was performed (Figure 7C). The replacement of G_7 by I_7 is expected to diminish the binding affinity to the I_7-C_8: G_9-C_{10} step since I_7 no longer has an N2 amino group which is involved in ActD binding. It was surprising to note that despite the G → I substitution the titration behavior is still similar to that of GCAGCAGC. Here note that the eight distinct G/I imino proton resonances (4 from G_4 and 4 from I_7) are particularly evident. The number of resonances in the 2:1 complex again indicates that ActD prefers to bind at the two ends.

ActD Stabilizes the Stems of the $(CAG)_n/(CTG)_n$ Cruciform

Two long DNA oligos, $G(CAG)_{10}C$ and $G(CTG)_{10}C$, were synthesized to serve as better models for the extended triplet repeats in chromosomes. Both strands form duplex structures, as evident from their imino proton resonances (bottom spectra in Figure 8). $G(CTG)_{10}C$ has two main sets of resonances, each associated with two GN1 imino (13.10 and 13.15 ppm) and two TN3 imino (10.80 and 10.85 ppm) protons (Figure 8A). (The nonequivalent G's are due to the asymmetric T:T base pairs). There are some small downfield peaks for both groups of protons which may be the base pairs near the termini. $G(CAG)_{10}C$ has one main set of resonances (12.50 ppm), associated with more equivalent G−N1 imino protons, possibly due to the weak and symmetrical A:A base pair (Figure 8B). When the two strands were mixed together without ActD, a normal Watson−Crick double helix was formed *rapidly* (Figure 8C). Note that all resonances associated with the individual $G(CAG)_{10}C$ and $G(CTG)_{10}C$ duplexes disappeared entirely. Three new resonances appeared at 13.85 ppm for the A:T base pairs and 12.65 and 12.75 ppm for the two types of G:C base pairs.

Addition of ActD to the three solutions of the $G(CAG)_{10}C$, $G(CTG)_{10}C$, and $G(CAG)_{10}C + G(CTG)_{10}C$ duplexes produced new resonances that are characteristic of the ActD−DNA complexes. Their chemical shifts are nearly identical to those from the ActD−octamer complexes seen in Figure 7. At a ratio of 6:1 the resonances from the free DNA duplexes are completely gone and the new resonances from the complexes are relatively sharp. This is consistent with the interpretation that ActD occupies every other GpC sites in the 32mer duplexes. There are theoretically six such sites. Some resonances (e.g., 11.30 ppm in ActD−$(G(CTG)_{10}C)$ appeared at lower drug to DNA ratio (e.g., 3:1), but disappeared at higher ratio. They may be due to other intermediate conformations.

It is interesting to note that the resonances from the ActD−$G(CAG)_{10}C + G(CTG)_{10}C$ complex are considerably broader than those from the other two complexes. Furthermore, the imino proton resonances from the A:T base pairs (~14.5 ppm) essentially become invisible. This is attributed to the fact that A:T base pairs flanking the ActD−GpC site are destabilized by the bound ActD. It is evident from the structural analysis of the two ActD−octamer complexes, ActD−$(GA\underline{AGCTTC})_2$ (Figure 5A) and ActD−$(GA\underline{TGCTTC})_2$ (Figure 5B), which showed that the *N*-methyl group and the side chains of Sar-MeVal dipeptides have to wedge between the ApG step in the former complex, but fit very well in the hole created by the T:T mismatch in the latter complex.

To further explore the relative stability of various ActD complexes, temperature-dependent studies of two systems, (1) the 7:1 ActD−$G(CAG)_{10}C$:$G(CTG)_{10}C$ complex and (2) the mixture of the 7:1 ActD−$G(CAG)_{10}C$ complex and the 7:1 ActD−$G(CTG)_{10}C$ complex, were carried out (Figure 6 of the supporting information). For each panel of Figure 6S, the experiments were carried out as follows. The temperature was raised slowly from 2 to 70 °C at 10 °C increments and equilibrated for 20 min (Figure 6S-C,F), and then the 1D NMR spectra in H_2O were collected. The temperature was slowly cooled back to 2 °C (Figure 6S-B,E) in a similar procedure. In another experiment the sample was heated at 100 °C in water bath for 5 min and then quenched by dropping the NMR tube into ice directly (Figure 6S-A,D). The spectrum was subsequently taken at 2 °C and labeled as quench 2.

In the spectra of the 7:1 ActD−$G(CAG)_{10}C$:$G(CTG)_{10}C$ complex, there are four visible G:C imino proton resonances and one broad A:T imino proton resonance (Figure 6S-A−C). Those resonances persist up to 70 °C. When the sample was slowly cooled down to 2 °C, an essentially identical spectrum was regenerated, indicating that very little rearrangement of the duplex structure occurred. Interestingly, when the sample was quenched from 100 to 2 °C, the same spectrum was seen.

The spectra of the 7:1 ActD−$G(CAG)_{10}C$ complex plus the ActD−$G(CTG)_{10}C$ complex (Figure 6S-D−F) are essentially a composite of two corresponding individual spectra (see Figure 8A,B). This suggests that there is again very little exchange between the two complexes. Nevertheless a small peak appears

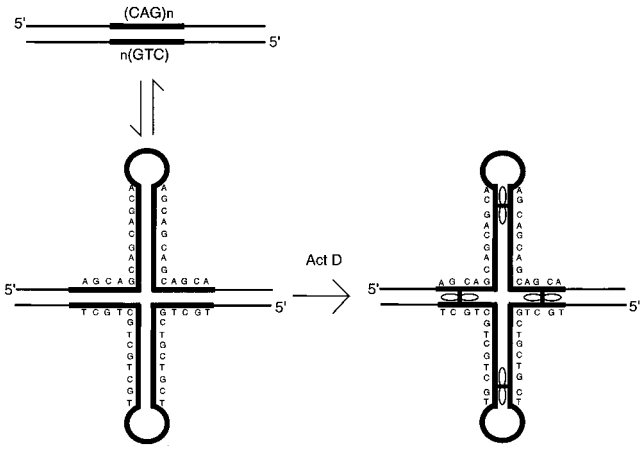

Figure 9. A schematic showing the binding of ActD to $(CAG)_n$:$(CTG)_n$ triplet sequences.

at 14.25 ppm which is due to the Watson−Crick A:T base pair, an indication of a very limited formation of the heteroduplex during mixing. The same temperature-dependent studies, i.e., slow cooling and quenching, were carried out. Both experiments resulted in very similar spectra in which a slight increase in the A:T imino resonance was detected. This is in contrast to the case in the free DNA study where an almost instantaneous formation of the Watson−Crick heteroduplex was noted. Our results strongly support the interpretation that ActD significantly stabilizes the structure of the individual $(CAG)_n$ and $(CTG)_n$ duplexes and prevents the DNA duplexes from annealing with each other at physiological temperatures.

Figure 9 shows a possible role ActD may play in the formation of a stable duplex structure of $(CAG)_n$ and $(CTG)_n$. The DNA triplet expansion in the DNA of the associated genetic diseases usually involves tens or even hundreds of the repeats. It is conceivable that the $(CAG)_n$:$(CTG)_n$ sequence may transform into a cruciform structure with extended arms composed of $(CAG)_n$ and $(CTG)_n$ duplexes, respectively,[13] possibly aided by the negative supercoiling density. If ActD is present at this time, it will bind to the stem regions and essentially trap the DNA in the cruciform configuration. This would have serious consequence on the subsequent functions associated with the normal $(CAG)_n$:$(CTG)_n$ heteroduplex. Whether other ligands (e.g., anthracycline drugs[18]) or proteins have similar properties is an intriguing topic to be pursued. Other triplet sequences such as $(CCG)_n$ and $(CGG)_n$ have also been shown to adopt unusual conformations.[20] The effect of ActD binding to those sequences will be another interesting question to be answered.

Acknowledgment. This paper is dedicated to Professor Nelson J. Leonard on the occasion of his 80th birthday. This work was supported by the American Cancer Society grant DHP-114 to A.H.-J.W. C.L. acknowledges the support from an NIH institutional NRSA in molecular biophysics (GM-08276). The Varian VXR500 NMR spectrometer was supported in part from the NIH Shared Instrumentation Grant 1S10RR06243.

Supporting Information Available: 1D NMR and 2D-NOESY spectra of ActD complexes, the structure of d(GCT-GCTGC)$_2$, a schematic of the binding of ActD to GCTGCTGC, and temperature dependent studies of ActD−DNA complexes (8 pages). See any current masthead page for ordering and Internet access instructions.

JA961631P

(20) Gao, X.; Huang, X.; Smith, G. K.; Zheng, M.; Liu, H. *J. Am. Chem. Soc.* **1995**, *117*, 8883−8884.