4106–4116 *Nucleic Acids Research, 2005, Vol. 33, No. 13*
doi:10.1093/nar/gki717

A host–guest approach for determining drug–DNA interactions: an example using netropsin

Kristie D. Goodwin, Eric C. Long[1] and Millie M. Georgiadis*

Department of Biochemistry and Molecular Biology, Indiana University School of Medicine and [1]Department of Chemistry and Chemical Biology, Purdue School of Science, Indiana University-Purdue University Indianapolis (IUPUI), IN 46202, USA

Received June 1, 2005; Revised and Accepted July 1, 2005

ABSTRACT

Netropsin is a well-characterized DNA minor groove binding compound that serves as a model for the study of drug–DNA interactions. Our laboratory has developed a novel host–guest approach to study drug–DNA interactions in which the host, the N-terminal fragment of Moloney murine leukemia virus reverse transcriptase (MMLV RT) is co-crystallized with a DNA oligonucleotide guest in the presence and absence of drug. We have co-crystallized netropsin with the RT fragment bound to the symmetric 16mer d(CTTAATTCGAATTAAG)$_2$ and determined the structure of the complex at 1.85 Å. In contrast to previously reported netropsin–DNA structures, our oligonucleotide contains two AATT sites that bind netropsin with flanking 5′ and 3′ sequences that are not symmetric. The asymmetric unit of the RT fragment–DNA–netropsin crystals contains one protein molecule and one-half of the 16mer with one netropsin molecule bound. The guanidinium moiety of netropsin binds in a narrow part of the minor groove, while the amidinium is bound in the widest region within the site. We compare this structure to other Class I netropsin–DNA structures and find that the asymmetry of minor groove widths in the AATT site contributes to the orientation of netropsin within the groove while hydrogen bonding patterns vary in the different structures.

INTRODUCTION

Our laboratory has developed a host–guest approach to crystallize and analyze the structures of DNA sequences of interest (1–5). In developing this system, one of our goals has been to study DNA–drug interactions. The host–guest system differs from other methods of DNA crystallization in that a DNA oligonucleotide (guest) is co-crystallized with a protein (host), the N-terminal fragment of the Moloney murine leukemia virus reverse transcriptase (MMLV RT). Here, we present the use of this system to study drug–DNA interactions. This method has several advantages over DNA-only crystals in which the DNA–DNA interactions govern lattice formation. One of the primary advantages is that the crystal lattice is comprised of protein–protein and protein–DNA interactions, but does not contain DNA–DNA interactions (1–5). Thus, there is an ~10 Å-thick cylindrical shell parallel to the helical axis that surrounds each DNA molecule and is available for interactions with DNA-specific binding compounds. A second advantage is that in contrast to DNA-only crystals, any DNA sequence can be analyzed using our system as interactions with the protein are limited to the 3′-OH end of one strand and minor groove base atoms and sugar atoms of the terminal 3 bp (1–5). The third major advantage is that once crystals have been obtained, the structure can be phased by molecular replacement using the RT fragment as the search model providing unbiased electron density for the DNA (1–5).

Herein, we chose to focus our initial efforts on the well-studied drug netropsin in order to validate the host–guest approach to studying DNA–drug interactions. The netropsin study provides an opportunity to compare essential features of the DNA–drug interface found in our crystal structure with those previously reported but is also of interest in its own right. Netropsin is a site-selective minor groove DNA binder with a preference for binding AT sites (6). AT-rich sites have a deeper minor groove, as compared with GC sites that contain an N2 amine, and display a narrower minor groove due to a higher propeller twist and a deeper electrostatic potential that accommodates the charged amidinium and guanidinium groups on the termini of the netropsin molecule (7). Despite having been studied extensively, fundamental questions still remain regarding details of the netropsin–DNA interaction (e.g. factors responsible for drug–DNA orientation) that in a broad

*To whom correspondence should be addressed. Tel: +1 317 278 8486; Fax: +1 317 274 4686; Email: mgeorgia@iupui.edu
Correspondence may also be addressed to Eric C. Long. Tel: +1 317 274 6888; Fax: +1 317 274 4701; Email: long@chem.iupui.edu

sense, continue to impact our general understanding of drug–DNA interactions. Two classes of netropsin binding have been described (8). Netropsins that exhibit Class I binding have a single orientation on the DNA and are positioned such that each amide group is situated between consecutive base pairs (8–13). These amide NH groups make bifurcated hydrogen bonds to A or T bases on the opposite strands, and the pyrrole rings rest against the C2 of A in one AT pair. Class II netropsins display end-for-end disorder (14,15). This type of binding eliminates the specificity of netropsin for AT sites because the drug is shifted down the minor groove where the pyrrole rings (instead of amide nitrogens as in Class I) are between base pairs, and there are no bifurcated hydrogen bonds.

The Class I structures of the netropsin–DNA complex have included several sequences of DNA with various AT-containing sites (8–13). In the majority of these structures, netropsin is bound to an oligonucleotide containing the AATT sequence. Although these oligonucleotides and their netropsin binding sites are sequence-symmetric, netropsin was bound to the DNA in only one orientation. In the absence of structural differences within the AATT that might result in asymmetry, netropsin could theoretically bind in two different orientations leading to end-for-end disorder as is seen in the Class II structures of netropsin bound to DNA.

Using our host–guest approach, we have analyzed the interaction of netropsin with a 16mer DNA oligonucleotide containing two AATT sites in the context of the N-terminal fragment of MMLV RT. Our system has allowed us to examine netropsin binding for the first time in a non-sequence symmetric environment and has revealed that the orientation of netropsin binding is correlated with placement of the guanidinium end in the narrowest part of the minor groove, corresponding to the 3′ end of the AATT sequence within the structure. Although the structure of the 16mer DNA oligonucleotide in the absence of netropsin differs in detail from that of the netropsin complex, the minor groove widths are virtually identical suggesting that the groove width is a structural property of the AATT site. A comparative analysis of the DNA structures and hydrogen bonding patterns with previously reported netropsin–DNA structures reveals that only the minor groove width is correlated with the orientation of the netropsin in all of these structures.

MATERIALS AND METHODS

Crystallization and data collection

The DNA oligonucleotide d(CTTAATTCGAATTAAG)$_2$ containing two netropsin binding sites was synthesized by TriLink BioTechnologies (San Diego, CA) and purified using standard reverse-phase HPLC methods. Correct placement of the netropsin binding site within the oligonucleotide was important because the first 3 bp of the oligonucleotide are involved in interactions with the protein. Thus, the AATT site within the oligonucleotide must be placed at the fourth base from the terminus to allow space for the drug to interact with the DNA. The purified oligonucleotide was resuspended in 10 mM $MgCl_2$ and 10 mM HEPES pH 7.0 and annealed prior to crystallization studies by heating to 80°C followed by slow cooling to room temperature. The RT fragment (residues 24–278) was purified as previously described by Ni-NTA affinity chromatography followed by S-Sepharose ion exchange chromatography (5). The 6×His-tag was removed by thrombin digestion and again subjected to S-Sepharose ion-exchange chromatography. Typical yields are 5–10 mg/l culture. Finally, the protein was concentrated to ~2 mM in 0.3 M NaCl/100 mM MES pH 6.0 for use in crystallization experiments.

Protein–DNA crystals were grown by hanging drop vapor diffusion crystallization at 20°C in 7% PEG 4000, 5 mM magnesium acetate and 50 mM ADA pH 6.5 from a 1:2 ratio of protein:DNA. The crystals were $\sim 60 \times 160 \times 200\ \mu^3$. The protein:DNA crystals were then used to microseed drops including a 1:2:8 ratio of protein:DNA:netropsin. Stock concentrations of the components were 2 mM RT fragment in 50 mM MES pH 6.0/0.3 M NaCl, 2.5 mM oligonucleotide in 10 mM HEPES pH 7.0/10 mM $MgCl_2$ and 25 mM netropsin in 100 mM HEPES pH 8.0. Final concentrations in the drop were 0.38 mM RT, 0.75 mM DNA and 2.9 mM netropsin. The protein–DNA crystals were stabilized in 20% ethylene glycol, 9% PEG 4000, 5 mM magnesium acetate and 100 mM HEPES pH 8.0. Additional cryosoaks containing 0.5, 1 and 2.5 mM netropsin were performed for the drug complex crystals. A 1.8 Å data set for the protein–DNA crystals was collected at beamline 19-BM of the Advanced Photon Source (Argonne, IL), and a 1.85 Å data set for the netropsin-containing crystals was collected on an R-axis IV^{++} detector mounted on a home source (Table 1). The presence of the drug in the crystals seems to improve the diffraction data; similarly sized crystals of the DNA:protein complex in the absence of drug diffracted to ~2.3 Å using home source X-rays. Data were integrated and processed with the HKL2000 package (16). Both crystals are orthorhombic, space group $P2_12_12$, with unit cell dimensions of $a = 54.57$ Å,

Table 1. Summary of crystallographic and refinement data

Cell parameters	RT:DNA	RT:DNA:Net
Cell constants (Å)	$a = 54.93$	$a = 54.57$
	$b = 145.75$	$b = 145.61$
	$c = 46.86$	$c = 46.81$
Space group	$P2_12_12$	$P2_12_12$
Statistics		
Maximum resolution	1.8 Å	1.85 Å
Reflections (unique)	35 445	32 157
(total)	169 451	142 521
Completeness (%)	98.9 (97.6)	98.1 (99.7)
R^a_{sym} (%)	5.2 (32.3)	5.3 (44.5)
I/s	19.7 (4.4)	27.6 (2.8)
Refinement		
Resolution range (Å)	50–1.8	50–1.85
Number of waters	237	288
Average B-factor (Å^2)		
Protein	29.2	26.2
DNA	56.5	39.6
Netropsin	N/A	58.2
Water	31.9	32.1
R^b_{value} (%)	22.7	22.0
R_{free} (%)	26.4	24.6

Data in parentheses are for highest resolution shell.
[a]$R_{sym} = \Sigma\Sigma_i|I_i - \langle I\rangle|/\Sigma\langle I\rangle$ where I is the integrated intensity of a reflection.
[b]$R_{value} = \Sigma_{hkl}||F_{obs} - kF_{calc}|/\Sigma_{hkl}|F_{obs}|$. 5% of all reflections were omitted from refinement and R_{free} is the same statistic calculated for these reflections.

$b = 145.61$ Å, $c = 46.81$ Å in the presence of netropsin and $a = 54.93$ Å, $b = 145.75$ Å, $c = 46.86$ Å without the drug.

Structure determination and refinement

We performed molecular replacement with AMoRe (17) using the refined structural model of the protein fragment from a previously analyzed crystal as the search model [PDB accession ID, 1N4L (4)]. This model was subjected to rigid-body, positional and *B*-factor refinement using the data collected from protein–DNA crystals to obtain unbiased electron density of the DNA. We generated an idealized *B*-form model of the correct DNA sequence using Nucleic Acid Builder (18) with a uniform rise of 3.38 Å and twist of 36°. This model was then adjusted to fit the F_o–F_c and $2F_o$–F_c electron density maps using O (19) and refined. Next, water molecules were added to the structural model followed by further positional and individual *B*-factor refinement. The DNA model was assessed using F_o–F_c maps and simulated annealing omit maps (20) to correct errors in positioning of the backbone and base atoms. We note that there is no model bias from the DNA prior to the interpretation of the unbiased difference electron density maps and adjustment of the initial DNA model, and that significant adjustments to the idealized DNA model were required in order to fit the difference electron density maps. Ultimately, 237 water molecules were positioned around the protein molecule in this structure. As in other structures that we have determined in this crystal form (1–4), there are no ordered water molecules associated with the DNA. Alternate cycles of model building using O (19) and refinement using calculations in CNS (20) were completed until the observed and calculated structure factors were close to convergence, and there were no large peaks remaining in the F_o–F_c electron density maps.

A strategy similar to the one detailed for the RT fragment–DNA structure was used to determine the structure of the RT fragment–DNA–netropsin complex. Following molecular replacement phasing, adjustment of the DNA model in the difference electron density maps in O (19), and completion of the water structure including 288 water molecules, F_o–F_c and $2F_o$–F_c maps were generated in which nearly continuous density was observed for the netropsin molecule in the minor groove of the expected binding site. Netropsin was modeled into the difference electron density maps using a previously refined model of netropsin bound to DNA (10). The orientation of the netropsin molecule was determined based on consideration of the placement of the pyrrole methyl groups and the carboxamide carbonyl oxygens for which resolved features were observed in the difference electron density maps. As in the RT fragment–DNA structure, alternate cycles of model building using O (19) and refinement using calculations in CNS (20) were completed until the observed and calculated structure factors were close to convergence, and there were no large peaks remaining in the F_o–F_c electron density maps. There are no ordered water molecules associated with the DNA in the netropsin complex. A summary of the refinement statistics is included in Table 1.

In order to confirm the orientation selected for the netropsin molecule, a reverse orientation for netropsin (with the guanidinium and amidinium ends reversed) was fit into the initial difference maps generated prior to inclusion of netropsin in the structural refinement. The reversed netropsin model was then refined with the same DNA, protein and water model as used for the original netropsin model. Two large negative peaks ($>-3\sigma$) were observed in the F_o–F_c electron density map corresponding to the position of the N7 pyrrole methyl group and N6 carboxamide carbonyl for the reversed orientation of netropsin suggesting that this orientation is not correct. In contrast, following refinement of the model with the original orientation of netropsin, no significant peaks were observed in the F_o–F_c maps. Because the complex in this crystal includes a 30 kDa RT fragment, an 8 bp duplex and 288 water molecules, *R* values are not particularly sensitive to the inclusion of netropsin. Small differences in the *R*-free values were observed for the correct netropsin model, 24.5% versus 24.7% for the incorrect model. The *R*-free value in the absence of netropsin is 25.2%. As in the other reported structures of netropsin–DNA complexes, *B*-factors for the netropsin atoms are higher (~1.5 times higher) than the *B*-factors for the DNA atoms with which they interact. The *B*-factors associated with netropsin atoms do not appear to result from less than full occupancy of netropsin in the structure as refinement of netropsin at half occupancy results in large positive peaks in the F_o–F_c maps.

Coordinates for the two structures reported here have been deposited. The PDB identifiers are 1ZTT and 1ZTW corresponding to the RT fragment–DNA complexes with and without netropsin, respectively.

Figures including molecular renderings were generated using MOLSCRIPT (21) and RASTER 3D (22,23). Representations of electron density were generated using CONSCRIPT (21).

RESULTS

Host–guest approach applied to drug–DNA interactions

We have previously reported the use of the N-terminal fragment from MMLV RT, which we refer to here as the RT fragment, to crystallize and analyze nucleic acid sequences of interest (2). We now describe the application of this host–guest system to study the interactions of netropsin with DNA. The RT-fragment was co-crystallized with a 16mer DNA sequence containing two netropsin binding sites d(CTTAATTCGAATTAAG)$_2$ in the presence and absence of netropsin (see Figure 1). Microseeding techniques were employed to obtain crystals of the desired RT fragment–DNA complex and RT fragment–DNA–netropsin complex resulting in diffraction-quality crystals within 1–2 days as detailed in the Materials and Methods. Unbiased difference electron density maps obtained from the molecular replacement phasing with the protein alone revealed well-resolved density for the DNA and significant density for the netropsin. Although it was also possible to soak netropsin into the crystals of the RT fragment complexed with DNA, the resulting density from the soaking experiment was of poor quality compared to that obtained by co-crystallization. Our analysis of the RT fragment–DNA–netropsin structure suggests that determination of the structures of other drug–DNA complexes will be possible using this system.

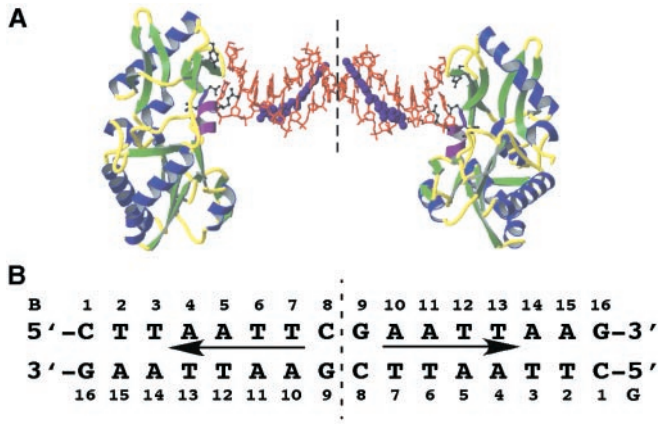

Figure 1. (**A**) The crystal structure of an RT fragment–DNA–netropsin complex. The asymmetric unit consists of one protein molecule, an 8 bp oligonucleotide duplex and one netropsin molecule representing half of the symmetric complex. The dashed vertical line divides the symmetrically equivalent halves of the 16 bp oligonucleotide. The DNA oligonucleotide is shown in a red sticks model, and netropsin is shown in a purple CPK model. RT is shown as a ribbon rendering with β-strands in green, coils in yellow and α-helices in blue except for the αD helix in magenta. Residues Tyr-64, Asp-114, Leu-115, Arg-116 and Gly-191 that make contacts with the DNA are shown in black ball-and-sticks models. (**B**) Schematic of the oligonucleotide duplex with the two complementary strands (B and G) and the numbering scheme referred to in the text. Arrows denote netropsin molecules oriented from guanidinium (tail) to amidinium end (head).

Table 2. Hydrogen bonds between RT and DNA in the presence and absence of netropsin

Residue	Atom	Atom	Nucleotide	Distance (Å)	Netropsin
Tyr-64	OH	O2	C1	3.2	+
		O4′	T2	3.4	–
Asp-114	$O^{\delta}2$	N2	G16	3.0/3.1	–/+
Leu-115	N	O3′	G16	3.0/2.9	–/+
Arg-116	$N^{\eta}2$	O4′	T3	2.8/2.0	–/+
		O2	T3	3.3	+
	$N^{\eta}2$	N2	G16	3.2/3.3	–/+
		O2	T2	2.7/2.8	–/+
		O4′	T3	3.3/3.3	–/+
Gly-191	O	O3′	G16	2.9/3.0	–/+

Protein–DNA interactions

Details of the protein structure have been described previously (24). The asymmetric unit of the Form IV structure contains one protein molecule and one half of the 16mer as shown in Figure 1A (2). The oligonucleotide is bound to conserved residues in the fingers domain of the catalytic fragment of MMLV-RT as seen in previous structures of the complex (1–4). There are a total of 8 hydrogen bonds formed between one protein molecule and the last 3 bp of one end of the 16mer duplex (Table 2). Of these, two hydrogen bonds are formed by main chain atoms and the remainder by side chain atoms. The 3′-OH of G16 (refer to Figure 1B for numbering scheme) is hydrogen bonded to the main chain of Leu-115 and Gly-191. Interactions between side chain atoms of Tyr-64, Asp-114 and Arg-116 are with minor groove base atoms and sugar atoms of G16, C1, T2 and T3 as shown in Table 2. The protein–DNA interactions in the presence and absence of netropsin are slightly different resulting in a loss of one hydrogen bond formed in the absence of netropsin and a gain of two additional hydrogen bonds. For example, Tyr-64 interacts with the second base of the 5′ end (T2 O4′) when netropsin is bound instead of the first base (C1 O2) in the absence of netropsin. Arg-116 also makes an additional contact with the DNA in the presence of netropsin. In both cases, the oligonucleotide maintains a *B*-form helix that is amenable to netropsin binding. Details of the DNA structure in the presence and absence of netropsin are described below.

Netropsin–DNA structure

Two netropsin molecules are bound to the DNA complexed with MMLV RT in the Form IV structure (Figure 1A), placing one netropsin molecule within the asymmetric unit. As detailed in the Materials and Methods, once the DNA and water structures had been completed, F_o–F_c and $2F_o$–F_c maps were generated in order to model the netropsin molecule bound to the DNA. As previously suggested, we found that evaluation of initial difference maps prior to inclusion of netropsin in the refinement were most useful in determining the orientation of netropsin (8). Well-resolved electron density for the pyrrole methyl groups and carboxamide carbonyl oxygen atoms as shown in Figure 2A allowed us to distinguish the correct orientation for the netropsin molecule, and the density for the netropsin molecule improved upon refinement as shown in Figure 2B and C. No significant peaks were observed in the final F_o–F_c maps in the vicinity of the netropsin molecule. As detailed in the Materials and Methods, fitting of netropsin in the reverse orientation resulted in two large negative peaks in the F_o–F_c maps corresponding to the positions of a pyrrole methyl group and a carboxamide carbonyl group indicating that this orientation was not correct. Thus, the orientation of the netropsin is unambiguous in our structure.

Hydrogen bonding between netropsin and the DNA

The single orientation of netropsin identified in our crystal lattice and pattern of hydrogen bonds formed between netropsin and the DNA indicate that it is similar to other Class I netropsin–DNA structures (8). Class I netropsin–DNA structures exhibit a hydrogen bonding pattern that includes bifurcated hydrogen bonds between the amide NH groups of netropsin and the A (N3) or T (O2) bases on opposite strands of the DNA. These hydrogen bonds promote DNA recognition and binding of the netropsin molecule (12). In our structure, there are a total of 14 potential hydrogen bonds formed between netropsin and the DNA of which eight are between 2.7 and 3.2 Å and the remaining six are between 3.3 and 3.4 Å (Table 3, Figure 3).

A comparison of the hydrogen bonds formed between netropsin and DNA in our structure with other class I structures reveals interesting similarities and differences in the hydrogen bonding patterns. For the purposes of this analysis, we have considered potential hydrogen bonds as those for which donor–acceptor distances are 2.4–3.4 Å. Three other structures including the AATT site have been used for this comparison including complexes of netropsin with the following sequences d(CGCGAATTCGCG) (8,13), which we will refer to in this discussion by PDB identifier, 101D, d(CGCAAATTTGCG) (11), 121D and d(CGCAATTGCG) (12), 261D. The number of hydrogen bonds between netropsin and the DNA varies

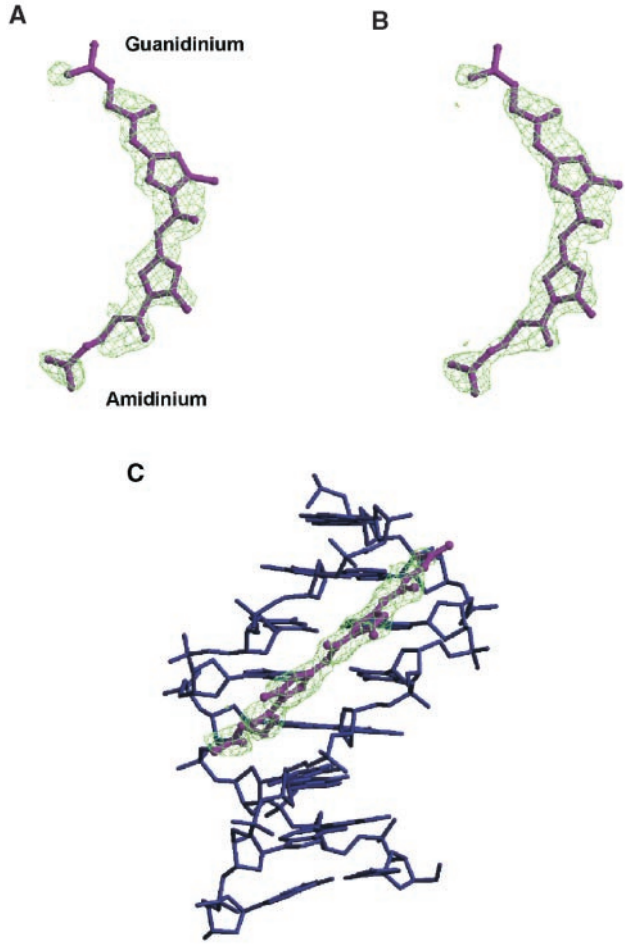

Figure 2. (**A**) The initial $2F_o$–F_c map of the netropsin density contoured at 1σ is shown in a green cage rendering with the final netropsin model superimposed in magenta. (**B**) The final $2F_o$–F_c map of netropsin density contoured at 1σ is shown also in a green cage rendering with the final refined model of netropsin superimposed in magenta. (**C**) The final $2F_o$–F_c map is shown superimposed on the final DNA model in blue sticks and the final netropsin model in a magenta ball-and-stick rendering.

from 7 in 101D to 14 in our structure. All seven hydrogen bonds found in 101D, which include the bifurcated hydrogen bonds involving N4, N6 and N8 atoms of netropsin with O2 and N3 atoms from T and A bases (that are not base-paired) from complementary strands and an interaction between N10 and N3 of A, are present in our structure (see Figure 3B).

The hydrogen bonding pattern observed in the 121D structure includes eight hydrogen bonds with only one bifurcated hydrogen bond for the N8 atom of netropsin. Of the possible hydrogen bond donors present in netropsin, N1, N3, N4, N6, N8 and N9 are involved in hydrogen bonds with the DNA. As in our structure, N3 forms a hydrogen bond to an O4′ sugar atom between the first two A in the AATT site. The lack of bifurcated hydrogen bonds for N4 and N6 atoms of netropsin makes the 121D structure the least similar to the others within the group of Class I structures.

The 261D structure includes 12 hydrogen bonds between netropsin and the octamer duplex to which it is bound and also has bifurcated hydrogen bonding interactions with the N4, N6 and N8 atoms of netropsin. Of the hydrogen bonds observed in the 261D structure, eight are also found in our structure. A detailed comparison of the hydrogen bonding interactions observed in 261D and our structure is shown in Figure 3B and C. While our structure includes interactions between the O4′ sugar atoms of one strand of the DNA duplex and the N1, N3, N4, N6 and N8 atoms of netropsin and an N10–O4′ hydrogen bond to the opposite strand, 261D includes interactions between O4′ sugar atoms and N1 and N10 on only one strand of the DNA. Our structure also includes hydrogen bonding interactions with six of the possible NH donors present in netropsin, while 261D and 121D include interactions with only five of the NH donors and 101D, with only four. It is perhaps of some interest to note that the interactions within the central region including N4 to N8 of the netropsin molecule are more highly conserved in the different structures than those with the N1 and N10 atoms at the ends of the molecule.

Minor groove widths within the netropsin binding site

The netropsin molecule is an asymmetric molecule whose orientation with respect to the DNA can be defined by the position of its guanidinium and amidinium ends. Although all sites in previously reported structures are symmetrical with respect to sequence in both strands of the DNA (i.e. GAATTC/gaattc, small letters denote the complementary strand), the netropsin binding site in our sequence is asymmetrical because the flanking bases immediately adjacent to the site are different (TAATTC/gaatta). Comparison of our structure with other class I structures suggests that there is no correlation between the orientation of netropsin bound to the DNA and the nature of the flanking sequence (i.e. pyrimidine versus purine base). In addition, the structures of the AATT sites within the Class I netropsin–DNA structures including those compared above, 101D, 121D and the inverted TTAA site within d(CGCGTTAACGCG) (11) (PDB identifier 195D) are very similar with an average r.m.s.d. of 0.5 Å for superimpositioning of C1′ atoms as shown in Figure 4A. The AATT sites of 261D and our structure are also very similar as shown in Figure 4B in which the AATT sites are shown along with the netropsin molecules following superimpositioning of the C1′ DNA atoms within the AATT sites. As was true for the other structures compared above, the AATT sites within our structure and the 261D structure are very similar (r.m.s.d. of 0.3 Å). The netropsin molecules are also very similar. The only significant differences occur in the amidinium ends and result from the hydrogen bonding to opposite strands of the DNA in the two structures as shown in Figure 3.

In contrast to previous reports, a detailed analysis of our DNA structure revealed that while the AATT sequence is symmetric, the widths of the minor groove are asymmetric within this site. Unless otherwise stated, 3DNA (25) has been used to analyze the DNA structures discussed here. Minor groove widths are calculated in 3DNA based on the cross-strand distances between phosphate groups (25,26). In our structure, netropsin is bound such that the guanidinium end is in the narrow part of the groove at the AATT site while the amidinium end is in the widest region within the site. Indeed, analysis of the same netropsin–DNA structures (8,9,11) that were superimposed on our structure (see Figure 4) revealed that the increase in minor groove width within the AATT site

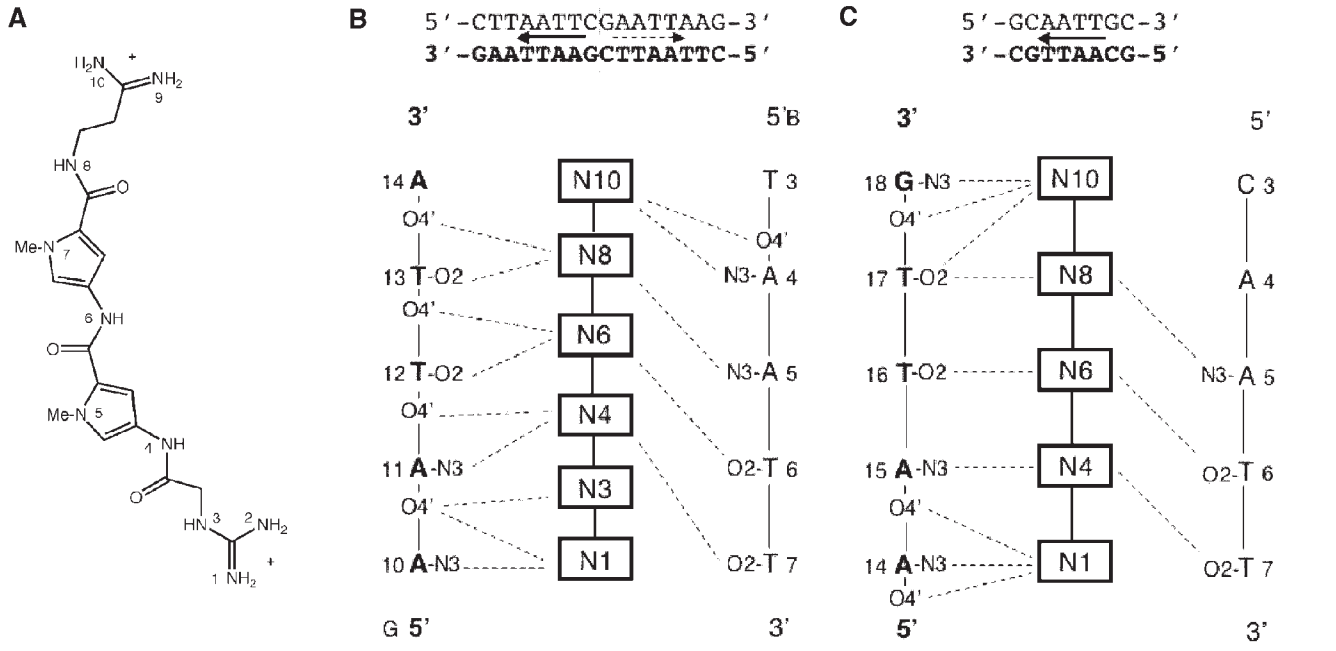

Figure 3. (**A**) Diagram of netropsin including the numbering scheme for the atoms referred to in the subsequent panels. (**B**) The hydrogen bonds between netropsin and the DNA to which it is bound are shown schematically for this work and (**C**) the structure reported by Nunn *et al*. (12), PDB identifier 261D. Arrows denote netropsin oriented from guanidinium (N1, tail) to amidinium end (N10, head). The numbering of the bases in (B) refers to the bold arrowhead representing netropsin. The symmetry-related netropsin molecule is represented as a dashed arrow and the vertical dashed line divides the two symmetry related halves of the 16 bp oligonucleotide. Dashed lines between the N atoms of netropsin and the DNA atoms represent hydrogen bonds. Of the 14 hydrogen bonds present in our structure, 8 are also found in the 261D structure.

Table 3. Hydrogen bonds between netropsin and DNA

Netropsin atom	DNA atom	Chain	Distance (Å)
N1	N3 of A10	G	2.7
N1	O4′ of A11	G	3.2
N3	O4′ of A11	G	3.3
N4	O2 of T7	B	2.9
N4	N3 of A11	G	3.0
N4	O4′ of T12	G	3.3
N6	O2 of T6	B	3.3
N6	O2 of T12	G	3.2
N6	O4′ of T13	G	3.3
N8	N3 of A5	B	3.4
N8	O2 of T13	G	2.8
N8	O4′ of A14	G	3.4
N10	N3 of A4	B	2.9
N10	O4′ of A4	B	3.2

is consistent in the other netropsin–DNA structures as shown in Figure 5A. Further, the orientation of netropsin with respect to the minor groove width is also consistent, with the guanidinium moiety bound in the narrower end of the groove within the AATT site. For the inverted site TTAA, the narrowest part of the groove is the 3′ end of the site as is the case for the AATT sites. The minor groove width increases by 1.5–3 Å within the netropsin binding site from the guanidinium end to the amidinium end of the netropsin molecule. We have not included 261D in this comparison as minor groove widths can be calculated for only 3 bp within this octamer duplex using 3DNA. This correlation of minor groove width with the positioning of the more rigid guanidinium moiety within the narrower part of the minor groove and amidinium in the wider part of the minor groove within the AATT site has not been previously noted.

It has been previously noted that the propeller twist values are higher within AT sites in DNA structures and that this feature of the DNA structure contributes to the interaction with netropsin (8,13). However, the propeller twist values corresponding to each base pair within the AATT binding site, defined by the orientation of the netropsin binding, are not similar in magnitude in the different netropsin–DNA structures compared in Figure 5B. Therefore, we conclude that there is no simple correlation between the propeller twist for each base pair within the AATT site and the orientation of the netropsin binding within the minor groove. Our structure differs from the reported structures in that the base pair that interacts with the amidinium end has almost no propeller twist (−.4°); however, the same base pair in the absence of netropsin has a propeller twist at that site of −12.4°. Thus, while propeller twist may provide the narrowed minor groove environment that draws netropsin to the DNA, it appears to be the minor groove width that contributes significantly to the directionality of binding.

Comparison of structures with and without netropsin bound to the DNA

The observed variation in groove width within the AATT site does not result from the binding of netropsin to the DNA; our DNA models in the presence and absence of drug possess nearly identical minor groove widths at the AATT site and beyond (Figure 6A). However, a comparison of our structures with and without netropsin bound reveal significant differences consistent with the r.m.s.d. of 0.9 Å obtained for

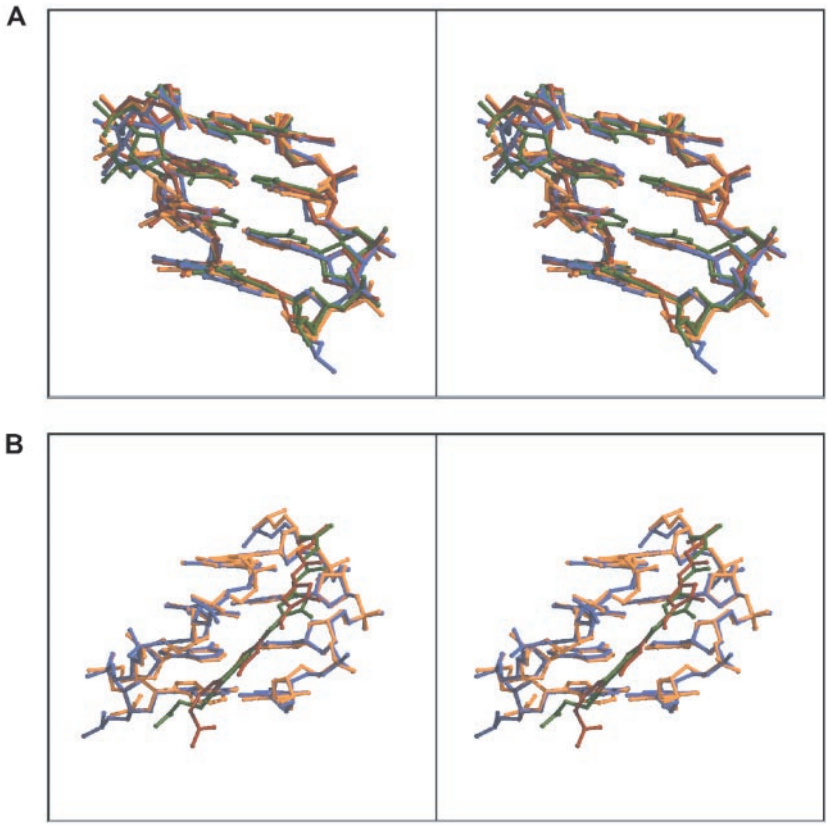

Figure 4. (**A**) Stereodiagram rendering (22,23) of the superimposed AATT netropsin binding sites of this work and several known netropsin–DNA structures. C1′ atoms from all 4 bp comprising the AATT sites were superimposed in O (19) based on directionality of netropsin binding. Reported structures are labeled by their PDB identifier. This work (AATT) in blue; 101D (AATT) in red, r.m.s.d = 0.4 Å; 121D (AATT) in orange, r.m.s.d. = 0.4 Å; 195D (TTAA') in green, r.m.s.d. = 0.6 Å. (**B**) The structures of the AATT site in our structure (blue) and that in the 261D structure (yellow) were superimposed as above using C1′ atoms with an r.m.s.d. = 0.3 Å. The netropsin molecules are shown for our structure (green) as well as the 261D structure (red).

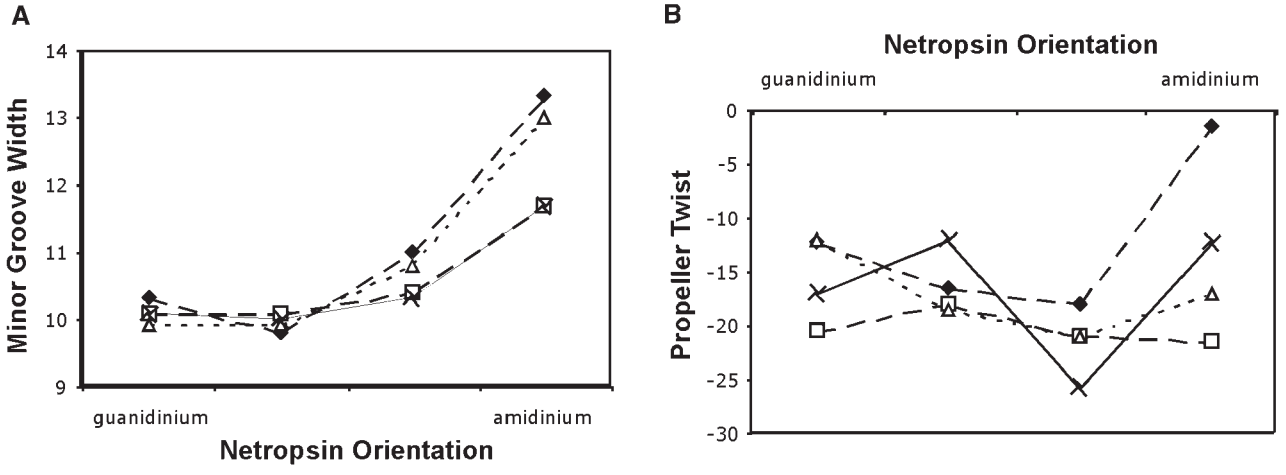

Figure 5. (**A**) Minor groove widths (in Å), calculated in 3DNA based on cross-strand distances of phosphate groups (25,26) and (**B**) propeller twist (in degrees) of base-pair steps in netropsin binding sites in the presence of netropsin are graphed according to binding directionality of netropsin to the DNA. Reported structures are labeled as their Protein Data Bank ID. This work (black diamond), 101D (open square) and 121D (X) are AATT sites; 195D (Δ) is a TTAA site.

superimpositioning of the C1′ atoms within each DNA structure (Figure 6B). These differences are most significant for the DNA residues that interact with the guanidinium and amidinium moieties of the netropsin upon comparison with the structure lacking netropsin. Major changes occur near the amidinium end where the backbone and base atoms of T3 are pulled in toward the center of the DNA, while those of A4 are pushed away from the center upon binding of netropsin (see Figure 1 for numbering scheme). Meanwhile, the structures of the residues base-paired to T3 and A4, A14 and T13,

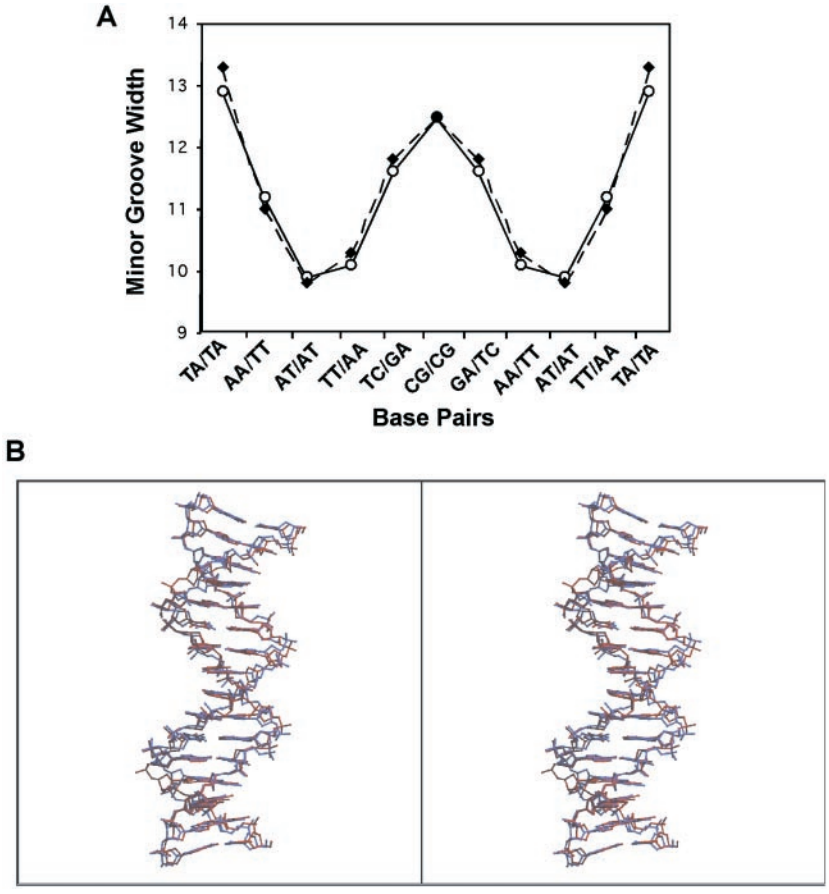

Figure 6. (**A**) Minor groove widths of DNA base-pair steps (from 3DNA calculations) in the absence (open circle) and presence (black diamond) of netropsin. (**B**) Stereo diagram of structures of DNA in the absence (red) and presence (blue) of netropsin (r.m.s.d = 0.9 Å). Superimpositioning of C1′ of all 16 bp was done using O (19).

respectively, are similar to that of the DNA in the absence of netropsin. Near the guanidinium end of netropsin, the backbone and base atoms of A10 and A11 are shifted relative to their positions in the structure without netropsin. The differences observed for T3 and A4 on one strand and A10 and A11 of the complementary strand appear to be a direct result of netropsin binding as they coincide with the pattern of hydrogen bonding between DNA and the drug as shown in Figure 3B; hydrogen bonds with the backbone may affect the positions of the sugar rings so that the DNA can accommodate netropsin. In addition to the changes observed in the positions of the DNA atoms upon binding of netropsin, a number of other structural changes in the DNA occur near the ends of the netropsin binding site: increased base-pair opening at the T7–A10 step (−7.8° versus 5.2°), increased roll (4.8° versus 13.2°) and inclination (7.7° versus 24.3°) of T3–A14/A4–T13, and decreased tip (7.1° versus −1.9°) of T3–A14/A4–T13 for the structures in the absence and presence of netropsin, respectively. Although netropsin binding to DNA causes alterations in structure for DNA residues surrounding the netropsin binding site, the minor groove widths of the DNA remain the same whether the netropsin is bound or not. Furthermore, we analyzed the previously reported DNA structure d(CGCGAATTCGCG) in the presence (PDB accession ID 6BNA) and absence (3BNA) of netropsin, and found that minor groove widths were the same even though a number of structural changes were evident upon netropsin binding (data not shown) (13,27). Thus, we conclude that the differences in minor groove widths within the AATT site are structural features of the DNA that promote directionality of netropsin binding within the site.

Comparison of the netropsin structures

To compare the curvature of the netropsin molecules bound in the various netropsin–DNA structures, the netropsin models were superimposed onto our model of netropsin using those atoms in the guanidinium end through the first pyrrole ring (Figure 7) (8,9,11–13). This region of the netropsin molecule is quite similar (average r.m.s.d. = 0.3 Å), whereas the orientation of the amidinium ends of the netropsin molecule differ such that the superimposed molecules form a fan shape. The degree of curvature for the netropsin models observed appears to correlate with the interactions of the ends of the netropsin molecule with the DNA. For the most curved netropsin models including our structure, the 261D (12) and 195D (9) models, both N1 and N10 atoms of netropsin form hydrogen bonds with the DNA. In the least curved models including 101D and 121D, interactions with the DNA include either N1 or N10 netropsin atoms but not both. The different conformations of netropsin and the alterations in DNA structure when netropsin is bound suggest that there is a compromise between

Figure 7. Superimposed netropsin molecules from our structure and several known netropsin–DNA structures. Structures of the atoms from the guanidinium end through the first pyrrole ring were superimposed using LSQKAB in CCP4i (33,34). Reported structures are labeled by their Protein Data Bank ID. This work in blue; 101D in red, r.m.s.d = 0.4 Å; 121D in orange, r.m.s.d. = 0.5 Å; 195D in green, r.m.s.d. = 0.2 Å; 261D in magenta, r.m.s.d = 0.3 Å.

the netropsin and DNA in order to promote hydrogen bonding and structural changes that stabilize netropsin–DNA interactions, potentially an induced fit, akin to that reported for an analogue of Hoechst 33258 binding to A-tract DNA (28). Thus, netropsin binds with the more rigid guanidinium moiety near the narrow part of the groove allowing the more adjustable amidinium end to be accommodated in the wider part of the groove within the AATT binding site.

In conclusion, netropsin in our structure is a Class I binder with a clearly distinguished orientation. Analysis of the minor groove widths in the netropsin binding site reveals for the first time that the guanidinium end is located in the narrow part of the groove, while the amidinium end binds at the wider part of the groove. Further analyses of previously reported Class I netropsin–DNA complexes support and confirm this observation. We propose that the asymmetrical groove widths in this otherwise sequence-symmetrical site contribute to the orientation of the netropsin molecule within the minor groove.

DISCUSSION

Structures of netropsin–DNA interactions have been reported for >20 years and have led to theories about the directionality of netropsin binding. However, a comparative analysis of all the available Class I netropsin–DNA structures suggests that the factors previously suggested as contributing to the orientation of netropsin binding are not correlated in all of the structures. The structures and comparative analysis presented here provide new insights on the directionality of netropsin binding.

It has been proposed that the specificity of netropsin binding results in part from the high propeller twist of AT sites leading to a narrow minor groove, in comparison with the surrounding GC sites (27). Analysis of our netropsin-bound DNA structure and several previously reported structures has revealed that the propeller twist values within the AATT sites are not correlated with the orientation of netropsin binding (Figure 5B). Thus, although the netropsin molecule may recognize the DNA based on the overall narrowed groove environment created by the high propeller twist of AT sites, this factor does not provide a basis for understanding the direction of netropsin binding.

Another possible factor in directional binding is the contribution of the sequence flanking the AATT site. In contrast to previously reported structures, our structure is the first netropsin–DNA structure in which two netropsin molecules are bound to one DNA molecule in two distinct AATT sites (Figure 1) rather than end-to-end in IC-containing decamers and bridged by one water molecule (29). The TAATTC site within our DNA is a non-symmetric site within a symmetric oligonucleotide, whereas the sites within the reported netropsin–DNA structures are completely symmetrical [i.e. GAATTC (13)]. Thus, symmetry surrounding the site is not required for netropsin binding. Furthermore, the bases immediately adjacent to the AT sites in the known structures vary but do not appear to have an effect on orientation of the netropsin molecule within the site.

The first crystal structure of netropsin bound to DNA revealed that molecular specificity of the drug for AT sites resulted from hydrogen bonds formed between the drug and DNA (7,13). Assessment of the hydrogen bonding pattern of netropsin bound to AATT or TTAA in subsequent work suggested that the netropsin molecule is oriented based on the presence or absence of hydrogen bonds between the amide N6 and the DNA (9). The amide N6 of netropsin displays bifurcated hydrogen bonding to each strand of the AATT site in our structure, the original structure, d(CGCGAATTCGCG) (13) and the d(CGCAATTGCG) (12), but forms a single hydrogen bond to one strand of the DNA in the AATT site within the sequence d(CGCAAATTTGCG) (11) and the TTAA sequence d(CGCGTTAACGCG) (9,11). Further analysis of hydrogen bonding patterns reveals that both the number of hydrogen bonds as well as the specific hydrogen bonds formed varies in the different structures that have been analyzed. Thus, hydrogen bonding patterns may affect the affinity of netropsin for the DNA, as netropsin selectively binds AATT over TTAA sites (6), but are not correlated with the orientation of netropsin binding within the AATT site.

The factor that we have identified in all of the structures analyzed that contributes to the orientation of netropsin binding is the minor groove width within the AATT site. Specifically, the planar guanidinium end is bound in the narrow region of the AATT site whereas the more flexible amidinium end is situated in the wider part of the site. Thus, the narrow minor groove width within AT-rich sites probably contributes to the sequence selectivity of netropsin and other minor groove binding drugs (30,31), while variations of the minor groove width within the AATT site contribute to the orientation of netropsin binding.

Our finding regarding differences in minor groove width within a site leading to an orientational preference for netropsin is similar to that reported for the Hoechst 33258 (H33258) bound to an A-tract sequence in a 2:1 H33258:DNA complex. In that complex, H33258 was found to bind with a preferred orientation dictated by the fit of the benzimidazole portion of the compound in the narrowest part of the minor groove and

the *N*-methylpiperazine ring binding in the wider part of the groove (30).

In this study of netropsin–DNA interactions, we have validated our host–guest approach to the study of DNA–drug interactions and provided new information about the directional DNA binding of netropsin. Significant advantages of this host–guest system include the ability to use any DNA sequence and to obtain unbiased density for the DNA and the ligand by molecular replacement phasing using a previously refined model of the RT fragment. In addition, we envision that the lack of DNA–DNA interactions within the lattice may allow us to study DNA binding molecules that have previously proven refractory to crystallographic analysis. A limitation of the system is the length of oligonucleotide that can be crystallized. Furthermore, as is true in all crystallographic studies of drug–DNA complexes, determination of the interactions of the drug with the DNA in this system requires site-specific drugs. Netropsin is not only site-specific, but is also a very tight binder with a K_d of 4×10^{-9} M for 1:1 binding for AT-rich DNA sites (32). Our initial success with netropsin suggests that the system will be amenable to the study of new DNA binding compounds.

ACKNOWLEDGEMENTS

We thank Steve Ginell, Marianne Cuff and Andrzej Joachimiak from the Structural Biology Center Collaborative Access Team at the Advanced Photon Source and the members of the Georgiadis lab for helpful discussions. Data were collected at beamline 19-BM in the facilities of the SBC-CAT at APS. Use of the Argonne National Laboratory Structural Biology Center beamline at the Advanced Photon Source was supported by the US Department of Energy, Office of Energy Research, under Contract No. W-31-109-ENG-38. This work was funded by grants from the National Institutes of Health (GM55026 to M.M.G. and GM62831 to E.C.L.). Funding to pay the Open Access publication charges for this article was provided by NIH.

Conflict of interest statement. None declared.

REFERENCES

1. Najmudin,S., Cote,M.L., Sun,D., Yohannan,S., Montano,S.P., Gu,J. and Georgiadis,M.M. (2000) Crystal structures of an N-terminal fragment from Moloney murine leukemia virus reverse transcriptase complexed with nucleic acid: functional implications for template-primer binding to the fingers domain. *J. Mol. Biol.*, **296**, 613–632.
2. Cote,M.L., Yohannan,S.J. and Georgiadis,M.M. (2000) Use of an N-terminal fragment from Moloney murine leukemia virus reverse transcriptase to facilitate crystallization and analysis of a pseudo-16mer DNA molecule containing G–A mispairs. *Acta Crystallogr. D Biol. Crystallogr.*, **56**, 1120–1131.
3. Cote,M.L. and Georgiadis,M.M. (2001) Structure of a pseudo-16mer DNA with stacked guanines and two G–A mispairs complexed with the N-terminal fragment of Moloney murine leukemia virus reverse transcriptase. *Acta Crystallogr. D Biol. Crystallogr.*, **57**, 1238–1250.
4. Cote,M.L., Pflomm,M. and Georgiadis,M.M. (2003) Staying straight with A-tracts: a DNA analog of the HIV-1 polypurine tract. *J. Mol. Biol.*, **330**, 57–74.
5. Sun,D., Jessen,S., Liu,C., Liu,X., Najmudin,S. and Georgiadis,M.M. (1998) Cloning, expression, and purification of a catalytic fragment of Moloney murine leukemia virus reverse transcriptase: crystallization of nucleic acid complexes. *Protein Sci.*, **7**, 1575–1582.
6. Abu-Daya,A., Brown,P.M. and Fox,K.R. (1995) DNA sequence preferences of several AT-selective minor groove binding ligands. *Nucleic Acids Res.*, **23**, 3385–3392.
7. Kopka,M.L., Yoon,C., Goodsell,D., Pjura,P. and Dickerson,R.E. (1985) The molecular origin of DNA–drug specificity in netropsin and distamycin. *Proc. Natl Acad. Sci. USA*, **82**, 1376–1380.
8. Goodsell,D.S., Kopka,M.L. and Dickerson,R.E. (1995) Refinement of netropsin bound to DNA: bias and feedback in electron density map interpretation. *Biochemistry*, **34**, 4983–4993.
9. Balendiran,K. (1995) X-ray structures of the B-DNA dodecamer d(CGCGTTAACGCG) with an inverted central tetranucleotide and its netropsin complex. *Acta Cryst. D*, **51**, 190–198.
10. Abrescia,N.G., Malinina,L. and Subirana,J.A. (1999) Stacking interaction of guanine with netropsin in the minor groove of $d(CGTATATACG)_2$. *J. Mol. Biol.*, **294**, 657–666.
11. Tabernero,L., Verdaguer,N., Coll,M., Fita,I., van der Marel,G.A., van Boom,J.H., Rich,A. and Aymami,J. (1993) Molecular structure of the A-tract DNA dodecamer d(CGCAAATTTGCG) complexed with the minor groove binding drug netropsin. *Biochemistry*, **32**, 8403–8410.
12. Nunn,C.M., Garman,E. and Neidle,S. (1997) Crystal structure of the DNA decamer d(CGCAATTGCG) complexed with the minor groove binding drug netropsin. *Biochemistry*, **36**, 4792–4799.
13. Kopka,M.L., Yoon,C., Goodsell,D., Pjura,P. and Dickerson,R.E. (1985) Binding of an antitumor drug to DNA, Netropsin and C-G-C-G-A-A-T-T-BrC-G-C-G. *J. Mol. Biol.*, **183**, 553–563.
14. Sriram,M., van der Marel,G.A., Roelen,H.L., van Boom,J.H. and Wang,A.H. (1992) Structural consequences of a carcinogenic alkylation lesion on DNA: effect of O6-ethylguanine on the molecular structure of the d(CGC[e6G]AATTCGCG)-netropsin complex. *Biochemistry*, **31**, 11823–11834.
15. Coll,M., Aymami,J., van der Marel,G.A., van Boom,J.H., Rich,A. and Wang,A.H. (1989) Molecular structure of the netropsin-d(CGCGATATCGCG) complex: DNA conformation in an alternating AT segment. *Biochemistry*, **28**, 310–320.
16. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. In Carter,J.C.W. and Sweet,R.M. (eds), *Macromolecular Crystallography, part A*. Academic Press, NY, Vol. 276, pp. 307–326.
17. Navaza,J. (1994) AMoRe: an automated package for molecular replacement. *Acta Cryst.*, **A50**, 157–163.
18. Macke,T. and Case,D.A. (1998) Modeling unusual nucleic acid structures. In Leontes,N.B. and SantaLucia,,J.J. (eds), *Molecular Modeling of Nucleic Acids*. American Chemical Society, Washington, DC, Vol. 1, pp. 379–393.
19. Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, **47**, 110–119.
20. Brunger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.
21. Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.*, **24**, 946–950.
22. Merritt,E.A. and Murphy,M.E.P. (1994) Raster3D Version 2.0—a program for photorealistic molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 869–873.
23. Bacon,D.J. and Anderson,W.F. (1988) A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graphics*, **6**, 219–220.
24. Georgiadis,M.M., Jessen,S.M., Ogata,C.M., Telesnitsky,A., Goff,S.P. and Hendrickson,W.A. (1995) Mechanistic implications from the structure of a catalytic fragment of Moloney murine leukemia virus reverse transcriptase. *Structure*, **3**, 879–892.
25. Lu,X.J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
26. El Hassan,M.A. and Calladine,C.R. (1998) Two distinct modes of protein-induced bending in DNA. *J. Mol. Biol.*, **282**, 331–343.
27. Fratini,A.V., Kopka,M.L., Drew,H.R. and Dickerson,R.E. (1982) Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBrCGCG. *J. Biol. Chem.*, **257**, 14686–14707.
28. Bockstock-Smith,C.E., Harris,S.A., Laughton,C.A. and Searle,M.S. (2001) Induced fit DNA recognition by a minor groove binding analogue

of Hoechst 33258: fluctuations in DNA A tract structure investigated by NMR and molecular dynamics. *Nucleic Acids Res*., **29**, 693–702.
29. Chen,X., Mitra,S.N., Rao,S.T., Sekar,K. and Sundaralingam,M. (1998) A novel end-to-end binding of two netropsins to the DNA decamers $d(CCCCCIIIII)_2$, $d(CCCBr5CCIIIII)_2$ and $d(CBr5CCCCIIIII)_2$. *Nucleic Acids Res*., **26**, 5464–5471.
30. Gavathiotis,E., Sharman,G.J. and Searle,M.S. (2000) Sequence-dependent variation in DNA minor groove width dictates orientational preference of Hoechst 33258 in A-tract recognition: solution NMR structure of the 2:1 complex with d(CTTTTTGCAAAAAG). *Nucleic Acids Res*., **28**, 728–735.
31. Clark,G.R., Squire,C.J., Gray,E.J., Leupin,W. and Neidle,S. (1996) Designer DNA-binding drugs: the crystal structure of a meta-hydroxy analogue of Hoechst 33258 bound to $d(CGCGAATTCGCG)_2$. *Nucleic Acids Res*., **24**, 4882–4889.
32. Lah,J. and Vesnaver,G. (2000) Binding of distamycin A and netropsin to the 12mer DNA duplexes containing mixed AT.GC sequences with at most five or three successive AT base pairs. *Biochemistry*, **39**, 9317–9326.
33. Kabsch,W. (1976) A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A*, **32**, 922–923.
34. Potterton,E., Briggs,P., Turkenburg,M. and Dodson,E. (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D Biol. Crystallogr*., **59**, 1131–1137.