Molecular Dynamics Simulations Shed Light on the Enthalpic and Entropic Driving Forces That Govern the Sequence Specific Recognition between Netropsin and DNA

Jožica Dolenc,[†,‡] Sarah Gerster,[†,§] and Wilfred F. van Gunsteren*[,†]

Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, CH-8093 Zürich, Switzerland, and Faculty of Chemistry and Chemical Technology, University of Ljubljana, SI-1000 Ljubljana, Slovenia

Received: January 18, 2010; Revised Manuscript Received: May 8, 2010

With the aim to gain a better understanding of the various driving forces that govern sequence specific DNA minor groove binding, we performed a thermodynamic analysis of netropsin binding to an AT-containing and to a set of six mixed AT/GC-containing binding sequences in the DNA minor groove. The relative binding free energies obtained using molecular dynamics simulations and free energy calculations show significant variations with the binding sequence. While the introduction of a GC base pair in the middle or close to the middle of the binding site is unfavorable for netropsin binding, a GC base pair at the end of the binding site appears to have no negative influence on the binding. The results of the structural and energetic analyses of the netropsin−DNA complexes reveal that the differences in the calculated binding affinities cannot be explained solely in terms of netropsin−DNA hydrogen-bonding or interaction energies. In addition, solvation effects and entropic contributions to the relative binding free energy provide a more complete picture of the various factors determining binding. Analysis of the relative binding entropy indicates that its magnitude is highly sequence-dependent, with the ratio $|T\Delta\Delta S|/|\Delta\Delta H|$ ranging from 0.07 for the AAAGA to 1.7 for the AAGAG binding sequence, respectively.

Introduction

Thermodynamic data on the sequence specific binding affinities of different small molecules to the DNA minor groove are available from a number of experimental[1−15] and theoretical[16−25] investigations, and they importantly complement the structural studies in the field.[26−42] Structural studies on drug−DNA complexes show the fine structure of ligand−DNA complexes defining hydrogen bonds and van der Waals interactions, which are responsible for the sequence specificity of the minor groove binding. However, because they cannot take into account sequence-dependent solvation and conformational changes in DNA and the ligands that occur upon binding, structural studies alone do not provide an explanation of the sequence specific thermodynamics of drug−DNA association.

Sensitive microcalorimeters nowadays allow measurement of binding constants and enthalpies of complex formation of drug−DNA interactions;[43−45] yet, the driving forces behind the observed binding affinities are still ambiguous.[46,47] This presents a bottleneck for the rational design of new compounds with altered or enhanced DNA sequence selectivity, particularly because the drug−DNA binding events characterized by similar binding free energies can be due to different thermodynamic driving forces.[48−52] An essential difficulty in the interpretation of the thermodynamic effects detected by microcalorimetry is that many different noncovalent interactions, solvation effects, and configurational changes contribute to the measurable heat effects. Because one cannot distinguish the enthalpic and entropic contributions that drive the process from the generally much larger solvent−solvent interactions, the molecular interpretation of the underlying thermodynamics is rather difficult to assess.[53−56]

From a theoretical point of view, the problems of developing a rigorous view of the thermodynamics of DNA recognition arise first from the challenge of modeling a polyanionic DNA at an atomic level in an explicit aqueous solution including the mobile counterions[57−74] and second from the difficulties in computing binding free energies and the corresponding enthalpic and entropic components for large biomolecular systems, the convergence of which can require very long computation times.[53,75−86]

With the aim to gain a better understanding of the thermodynamics that governs the sequence specific DNA minor groove recognition, binding of netropsin to an AT-containing and to a set of six mixed AT/GC-containing sequences in the DNA minor groove was used as a model for the investigations presented in this work. Netropsin is a well-characterized small cationic polyamide with a positive charge at both ends of the molecule (see Figure 1). It is known to strongly bind to AT sequences, while GC-containing sequences are known to be unfavorable for its binding.[10,27,87,88] Employing molecular dynamics (MD) simulations and free energy calculations followed by a detailed analysis of the resulting data, we address the following key questions: (i) How does the binding affinity of netropsin for the DNA minor groove depend on the position of the GC base pair in the binding site? (ii) Is there a correlation between the sequence dependence of the relative binding free energies and the interaction energies between netropsin and DNA? (iii) How do the detailed thermodynamic values for the enthalpic and entropic contributions to the relative free energy of binding, $\Delta\Delta H$ and $T\Delta\Delta S$, compare for the different netropsin binding sequences, and (iv) what is the origin of the differences between the enthalpic contributions for different netropsin binding sequences? We address these questions by a systematic characterization of the thermodynamic features of netropsin binding

* To whom correspondence should be addressed. Tel: +41 44 632 55 01. Fax: +41 44 632 10 39. E-mail: wfvgn@igc.phys.chem.ethz.ch.
† Swiss Federal Institute of Technology.
‡ University of Ljubljana.
§ Current address: Seminar für Statistik, ETH CH-8092 Zürich, Switzerland.

10.1021/jp100483f
Published on Web 08/06/2010

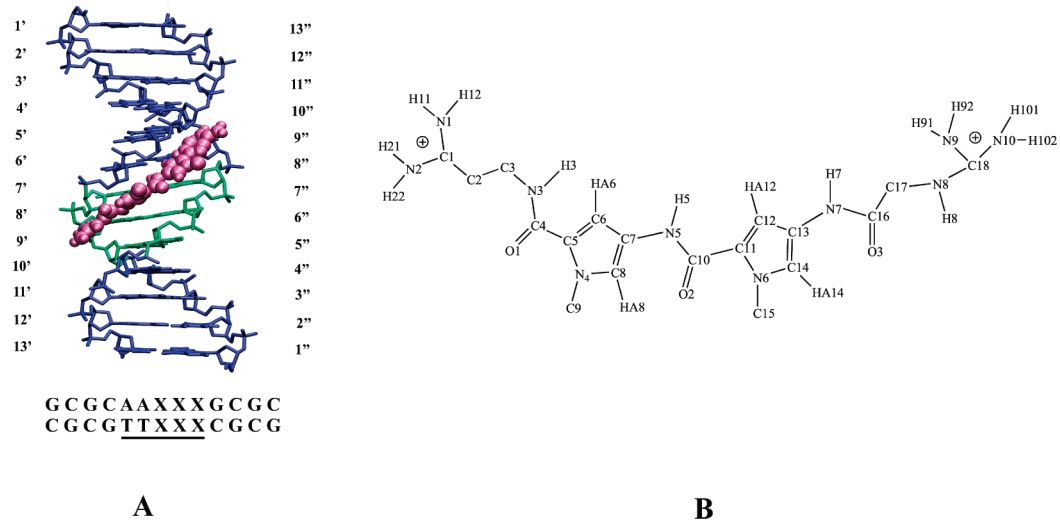


Figure 1. (A) Netropsin−DNA complex. Base pairs that were perturbed in the thermodynamic integration calculations are shown in green, and netropsin is shown in purple. The base pair sequence of the oligonucleotide is indicated with the binding site underlined and the perturbed bases depicted as X. (B) The chemical structure of netropsin with the atomic numbering scheme used in this work.

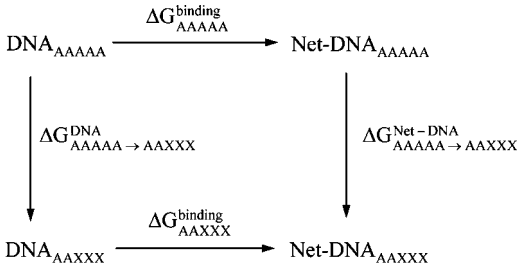
TABLE 1: Summary of the MD Simulations Discussed in This Paper

TI perturbation	code	no. of solute atoms	no. of water molecules	no. of counterions Na^+, Cl^-
$DNA_{GCGCAAAAAGCGC \rightarrow GCGCAAAAGGCGC}$	$DNA_{AAAAA \rightarrow AAAAG}$	634	11010	44, 20
$DNA_{GCGCAAAAAGCGC \rightarrow GCGCAAAGAGCGC}$	$DNA_{AAAAA \rightarrow AAAGA}$	634	11030	44, 20
$DNA_{GCGCAAAAAGCGC \rightarrow GCGCAAGAAGCGC}$	$DNA_{AAAAA \rightarrow AAGAA}$	634	11005	44, 20
$DNA_{GCGCAAAAAGCGC \rightarrow GCGCAAGGAGCGC}$	$DNA_{AAAAA \rightarrow AAGGA}$	643	11011	44, 20
$DNA_{GCGCAAAAAGCGC \rightarrow GCGCAAAGGGCGC}$	$DNA_{AAAAA \rightarrow AAAGG}$	643	11019	44, 20
$DNA_{GCGCAAAAAGCGC \rightarrow GCGCAAGAGGCGC}$	$DNA_{AAAAA \rightarrow AAGAG}$	643	11003	44, 20
$Net-DNA_{GCGCAAAAAGCGC \rightarrow GCGCAAAAGGCGC}$	$Net-DNA_{AAAAA \rightarrow AAAAG}$	681	10978	42, 20
$Net-DNA_{GCGCAAAAAGCGC \rightarrow GCGCAAAGAGCGC}$	$Net-DNA_{AAAAA \rightarrow AAAGA}$	681	11001	42, 20
$Net-DNA_{GCGCAAAAAGCGC \rightarrow GCGCAAGAAGCGC}$	$Net-DNA_{AAAAA \rightarrow AAGAA}$	681	10986	42, 20
$Net-DNA_{GCGCAAAAAGCGC \rightarrow GCGCAAGGAGCGC}$	$Net-DNA_{AAAAA \rightarrow AAGGA}$	690	11000	42, 20
$Net-DNA_{GCGCAAAAAGCGC \rightarrow GCGCAAAGGGCGC}$	$Net-DNA_{AAAAA \rightarrow AAAGG}$	690	11000	42, 20
$Net-DNA_{GCGCAAAAAGCGC \rightarrow GCGCAAGAGGCGC}$	$Net-DNA_{AAAAA \rightarrow AAGAG}$	690	11004	42, 20

to seven different DNA binding sequences, thereby reporting all thermodynamic data relative to the well-characterized poly(dA)poly(dT) binding sequence. The choice of the base pair sequences in the binding site was motivated by an earlier computational study of DNA−netropsin binding,[21,23] which involved one particular base pair sequence. The computational results matched the experimental data. Here, we consider variations on this sequence. Unfortunately, no experimental structural and thermodynamic netropsin binding data on six of the seven base pair sequences and DNA fragments considered here are available. For this reason, the emphasis will be on an analysis of the particular types of driving forces for binding, which are experimentally largely inaccessible, rather than on a comparison to experimental data.

Methods

Simulation Setup. The initial structure for the simulations was a canonical B-DNA structure d(GCGCAAAAAGCGC)·d(GCGCTTTTTGCGC) modeled with the INSIGHTII software package (Accerlys Inc., San Diego, CA). The coordinates for its complex with netropsin were obtained by superimposing the netropsin−DNA complex from the X-ray structure [sequence $d(CGCGAATTCGCG)_2$] deposited in the Protein Data Bank (PDB)[89] as entry 101d[36] on the modeled DNA oligomer, thus fitting the netropsin into the minor groove of the latter. The initial coordinates for the free (unbound) DNA were thus the same as for the DNA−netropsin complexes, which is a reasonable approximation since DNA does not undergo significant conformational changes on complex formation with monomer minor groove binding molecules.[87] The initial coordinates for the six remaining DNA duplexes and for the corresponding netropsin−DNA complexes (see Table 1) were generated from the d(GCGCAAAAAGCGC)·d(GCGCTTTTTGCGC) duplex and from its complex with netropsin by adding/removing the appropriate atoms in the DNA bases. The DNA and netropsin−DNA complexes were each solvated with approximately 11000 simple point charge (SPC) water molecules[90] in a truncated octahedron box with a minimal solute-to-wall distance of 1.4 nm, resulting in solvated structures containing approximately 33800 atoms (see Table 1). To neutralize the negative charge on the phosphate groups of the DNA backbone and to mimic the experimental conditions in drug−DNA binding experiments, the DNA oligomers were neutralized with 44 Na^+ and 20 Cl^- ions, and the netropsin−DNA complexes were neutralized with 42 Na^+ and 20 Cl^- ions, resulting in a salt concentration of 110 mM NaCl. The initial coordinates of the counterions were determined by replacing water molecules with the lowest or highest electrostatic potential by Na^+ or Cl^- ions while preserving a minimal interionic distance of 0.35 nm. All of the systems were relaxed by performing a steepest descent energy minimization with 2.5×10^4 kJ mol^{-1} nm^{-2} positional restraints on all solute atoms followed by a 100 ps long equilibration in which the positional restraints were gradually released from 2.5×10^4 to 0.0 kJ mol^{-1} nm^{-2}, and the temperature was raised from 60 to 298 K. The configurations at the end of the equilibration period were taken



$$\Delta\Delta G^{binding}_{AAAAA \rightarrow AAXXX} = \Delta G^{binding}_{AAXXX} - \Delta G^{binding}_{AAAAA} = \Delta G^{Net-DNA}_{AAAAA \rightarrow AAXXX} - \Delta G^{DNA}_{AAAAA \rightarrow AAXXX}$$

Figure 2. Thermodynamic cycle used for the calculation of relative free energies of binding of netropsin to different DNA binding sequences. The bases that were mutated in this work are labeled as X. The binding free energy of netropsin to a given DNA sequence (horizontal arrow) can be assessed experimentally but is difficult to calculate. However, the free energy change due to the base pair mutation in free DNA and in DNA−netropsin complex (vertical arrow) can be calculated using free energy simulations. Because the free energy is a state function, the relative free energy difference of netropsin binding to two different DNA binding sequences, $\Delta G^{binding}_{AAXXX} - \Delta G^{binding}_{AAAAA}$, can thus be estimated from the difference in the free energy of base pair mutation in DNA−netropsin complex and in free DNA, $\Delta G^{Net-DNA}_{AAAAA \rightarrow AAXXX} - \Delta G^{DNA}_{AAAAA \rightarrow AAXXX}$.

as initial configurations for the thermodynamic integration (TI) simulations.[91−93] All MD simulations were performed using periodic boundary conditions. The equations of motion were integrated using the leapfrog algorithm with a time step of 2 fs. The center of mass motion was stopped every 2 ps. Bond lengths were constrained by applying the SHAKE algorithm[94] with a relative geometric tolerance of 10^{-4}. The temperature and pressure were maintained at 298 K and 1 atm using the Berendsen thermostat with a coupling time $\tau_T = 0.1$ ps and barostat with a coupling time $\tau_P = 0.5$ ps and isothermal compressibility of 4.575×10^{-4} (kJ mol^{-1} nm^{-3})$^{-1}$.[95] A reaction field approach was used to treat the electrostatics employing a triple-range cutoff scheme, with cutoffs of 0.8 and 1.4 nm and a dielectric permittivity of 61.[96] The pairlist was updated every five steps. All MD simulations reported here were performed using the GROMOS biomolecular simulation package[97,98] and the GROMOS 45A4 force field.[99] The netropsin topology was taken from our previous work.[21] The chemical structure of the ligand together with the atomic numbering scheme is shown in Figure 1.

Free Energy Calculations. To evaluate the relative free energies of binding of netropsin to different DNA oligomers, free energy calculations were carried out on unbound DNA free in solution and on the netropsin−DNA complex in solution using the TI method.[91−93] In this method, the system is mutated from state A ($\lambda = 0$) to state B ($\lambda = 1$) by changing the interaction parameters that define the Hamiltonian as a function of a coupling parameter λ. The free energy difference ΔG between the two states of the system is calculated as

$$\Delta G_{A \rightarrow B} = G_B - G_A = \int_0^1 \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \qquad (1)$$

where $\langle\ \rangle_\lambda$ denotes an ensemble average at a given λ value. In the current work, the integral was evaluated using the trapezoidal rule based on simulations at 21 λ values spaced equidistantly between 0 and 1. A 0.5 ns production run was performed at

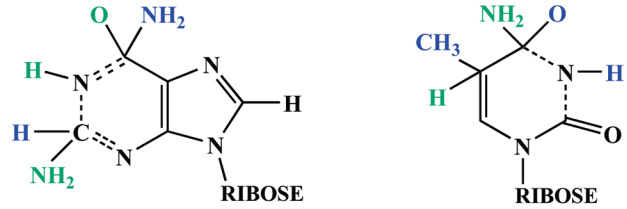


Figure 3. Structure of the two base pairs (DDAG and DDTC) used in TI calculations. Atoms that are present in the initial state ($\lambda = 0$) are shown in blue, and atoms that are present in the final state ($\lambda = 1$) are shown in green. The bonds that change during the perturbation are depicted as dashed.

each intermediate λ value, and a 1 ns production run was performed at the end points ($\lambda = 0$ or 1) of all TI simulations. The first 100 ps of each run was considered as the equilibration period and was not used in the calculation of the ensemble averages $\langle \partial H(\lambda)/\partial\lambda \rangle_\lambda$. To prevent instabilities in the simulations, the soft core approach was employed[100] with the softness parameter $\alpha^{LJ}_{ij} = 0.5$ for the Lennard−Jones interactions and $\alpha^{C}_{ij} = 0.5$ nm^2 for the electrostatic interactions. According to the thermodynamic cycle shown in Figure 2, the difference in free energy of netropsin binding to two different DNA sequences, $\Delta\Delta G^{binding}_{AAAAA \rightarrow AAXXX}$, was calculated by performing one TI simulation on unbound DNA in water, resulting in $\Delta G^{DNA}_{AAAAA \rightarrow AAXXX}$, and one on the netropsin−DNA complex in water, resulting in $\Delta G^{Net-DNA}_{AAAAA \rightarrow AAXXX}$. The difference in the binding free energy of netropsin to two different DNA sequences is then given as

$$\Delta\Delta G^{binding}_{AAAAA \rightarrow AAXXX} = \Delta G^{Net-DNA}_{AAAAA \rightarrow AAXXX} - \Delta G^{DNA}_{AAAAA \rightarrow AAXXX} \qquad (2)$$

For the work presented here, six TI simulations on DNA duplexes and six thermodynamic simulations on the netropsin−DNA complexes were performed resulting in 12 TI trajectories all together, each 11.5 ns long. A summary of the TI calculations together with the codes used in this article to refer to specific simulations is given in Table 1. As already mentioned in the Introduction to this paper, all relative free energies of binding of netropsin to different DNA oligomers were evaluated with respect to the d(GCGCAAAAAGCGC)·d(GCGCTTTTTGCGC) sequence. For the perturbation of the base pairs in the TI simulations, special molecular topology building blocks, which are presented in Figure 3, were used. Atoms treated as soft are depicted with blue and green color, and the bonds that changed during the perturbation are depicted as dashed. A detailed specification of the λ dependence of the Hamiltonian is given in refs 90 and 91.

Enthalpy and Entropy Calculations. Because of the large number of solvent−solvent interactions that are present in the simulated system, it is very difficult to calculate the total enthalpy change $\Delta H_{A \rightarrow B} = \langle H_B \rangle_B - \langle H_A \rangle_A$ that occurs in a TI simulation because of the perturbation. However, several authors have shown that in the process of solvation or molecular association, only the λ-dependent term of the Hamiltonian contributes to the relative free energy difference, while the λ-independent enthalpy and entropy contributions exactly cancel each other.[53,54,81] Following this argument, one can calculate the enthalpic and entropic contributions that drive the process of molecular association of two solutes in a solvent from the

enthalpy and entropy contribution of the perturbed part of the system only

$$\Delta G = \Delta H_{\text{pert}} - T\Delta S_{\text{pert}} \quad (3)$$

where ΔH_{pert} stands for enthalpy change and ΔS_{pert} stands for the entropy change of the perturbed (λ-dependent) part of the system between the final and the initial state of the TI simulation. In the present work, we have therefore calculated the enthalpy change that occurs either in the unbound DNA, $\Delta H^{\text{DNA}}_{\text{AAAAA}\rightarrow\text{AAXXX}}$, or in the DNA−netropsin complex, $\Delta H^{\text{Net}\rightarrow\text{DNA}}_{\text{AAAAA}\rightarrow\text{AAXXX}}$, due to the perturbation of base pairs from initial to the end state, from the difference between the interaction energy of the perturbed (λ-dependent) part with itself and with the unperturbed (λ-independent) part in the final state and in the initial state of the corresponding TI simulation. The difference between these two enthalpic changes is denoted as the relative binding enthalpy, $\Delta\Delta H^{\text{binding}}_{\text{AAAAA}\rightarrow\text{AAXXX}}$ (see Figure 2). The corresponding estimate of the entropic contribution to the relative binding free energy, $T\Delta\Delta S^{\text{binding}}_{\text{AAAAA}\rightarrow\text{AAXXX}}$, was calculated from the Gibbs equation as a difference between the relative binding free energy and the relative binding enthalpy

$$T\Delta\Delta S^{\text{binding}}_{\text{AAAAA}\rightarrow\text{AAXXX}} = \Delta\Delta H^{\text{binding}}_{\text{AAAAA}\rightarrow\text{AAXXX}} - \Delta\Delta G^{\text{binding}}_{\text{AAAAA}\rightarrow\text{AAXXX}} \quad (4)$$

We note, however, that ΔH_{pert} and ΔS_{pert} can be computed but not experimentally measured,[54] whereas ΔH and ΔS as calculated for all interactions can be measured, but only calculated for rather small systems due to the dominance of the λ-independent contributions $\Delta H_{\text{nonpert}}$ and $\Delta S_{\text{nonpert}}$, which exactly compensate each other, that is, $\Delta H_{\text{nonpert}} = T\Delta S_{\text{nonpert}}$.[53,54,81] Thus, these λ-independent contributions to ΔG may mask the real driving forces contained in ΔH_{pert} and ΔS_{pert}.[54]

Trajectory Analysis. Atom coordinates and energies of the trajectories were stored every 0.5 and 0.1 ps, respectively. The trajectory configurations for all of the initial and end states of the TI simulations were analyzed in terms of atom-positional root-mean-square deviation (rmsd) from the energy-minimized initial structures and in terms of intrasolute Watson−Crick (WC) hydrogen bonds. The rmsd values for DNA were calculated along each trajectory at 0.5 ps intervals, based on either all of the heavy atoms or only the heavy atoms of the DNA backbone, bases, or netropsin, using the backbone atoms to perform the superposition of centers of mass and rotational least-squares fit superposition[101] of the successive structures onto the reference one. WC hydrogen bonds were assumed to exist between the base pairs if the distance between the hydrogen and the acceptor atom was smaller than 0.25 nm and the angle between the donor-hydrogen and acceptor-hydrogen vectors was larger than 135°. The error estimates for the ensemble averages in the simulations were calculated using the block averaging method.[102] Visualization of the DNA duplexes and DNA−netropsin complexes was carried out using VMD.[103]

Results and Discussion

Structural Properties. The time series of the atom-positional rmsd of the DNA and netropsin−DNA heavy atoms with respect to the energy-minimized modeled structures are displayed in Figure S1 of the Supporting Information. The DNA double helices diverge from the modeled structure on average by 0.5 nm in the case of unbound DNA duplexes and by 0.4 nm in the case of netropsin−DNA complexes, indicating that binding of netropsin to DNA stabilizes the structure of the DNA duplexes. Furthermore, rmsd values for the DNA backbone are larger than for the DNA bases, reflecting the larger flexibility of the backbone when compared to that of the bases. Generally, the structures of all DNA duplexes and netropsin−DNA complexes remained stable throughout the simulations. The high occurrence of WC hydrogen bonds between the bases in DNA duplexes and in DNA−netropsin complexes presented in Figure S2 of the Supporting Information is additional evidence that no big structural changes occurred. A disruption of the terminal WC base pairing, which is rather common in MD simulations,[59,99] occurs for the terminal base pairs of the DNA_{AAAAG}, DNA_{AAAGA}, DNA_{AAGAA}, and DNA_{AAGAG} duplexes and the Net−DNA_{AAAGA} complex. All of the other WC hydrogen bonds except for 9′GUA(N2)−5″CYT(O2) in DNA_{AAAGG} and 5′ADE(N6)−9″THY(O4) in Net−DNA_{AAAGG}, the occurrence of which is only about 40%, are well preserved. The hydrogen-bonding patterns between netropsin and DNA that are of particular importance for this study are discussed in detail in the section on the structural and energetic analysis of the netropsin−DNA complexes.

Relative Binding Free Energies of Netropsin to Different DNA Binding Sequences. The calculated free energy differences for all of the inspected DNA and netropsin−DNA sequences and the resulting relative free energies of binding are collected in Table 2, and in Figure S3 of the Supporting Information, the corresponding free energy profiles are shown. As described in the Methods section of this paper, the 5′-AAAAA-3′ binding sequence, to which netropsin binds in a 1:1 high affinity mode,[5,10] was taken as a reference in all free energy calculations.

The results of the free energy calculations listed in Table 2 show a high sequence specificity of netropsin binding to the DNA minor groove, which has also been demonstrated in various experimental studies.[1,4,5,10] Introduction of a GC base pair in the middle or close to the middle of the netropsin binding site is unfavorable to binding. As expected, the most unfavorable relative free energy of binding of 43.6 ± 12.1 kJ/mol was calculated for the 5′-AAGGA-3′ binding sequence. The relative free energies of binding to 5′-AAGAA-3′ and 5′-AAAGA-3′ sequences (14.3 ± 7.2 and 14.3 ± 7.6 kJ/mol) are much smaller, however, still rather unfavorable, most likely indicating the negative effect of the exocyclic $-NH_2$ group of the guanine base on the netropsin−DNA interaction. Interestingly, the results of the free energy calculations also show that the substitution of an AT with a GC base pair at the end of the binding site has no negative influence on netropsin binding. The relative binding free energy for the 5′-AAGAG-3′ sequence is the same as for the 5′-AAGAA-3′ sequence, and the relative binding free energy of netropsin for the 5′-AAAAG-3′ sequence is nearly zero (see Table 2). We guess that the structural differences between the rigid body and the flexible tails of the netropsin molecule are a major reason for this observation.[23,104] It should be emphasized that the observation that netropsin is not sensitive to GC base pairs located at the end of its binding site is of importance in the development of the strategies for the rational design of new polyamide-based minor groove binders that can recognize mixed AT/GC-containing binding sequences.

The free energy changes that occur upon base pair perturbations in the DNA duplexes and in the netropsin−DNA complexes, $\Delta G^{\text{DNA}}_{\text{AAAAA}\rightarrow\text{AAXXX}}$ and $\Delta G^{\text{Net-DNA}}_{\text{AAAAA}\rightarrow\text{AAXXX}}$, from which the relative binding free energies discussed above were obtained (see Figure 2), can also be loosely compared to experimental observations. Assuming that the free energy changes of base

TABLE 2: Relative Free Energies, Enthalpies, and Entropies of Netropsin Binding to Six Different DNA Duplexes with Respect to the 5′-AAAAA-3′ Duplex along with the Corresponding Changes in Free Energy and Enthalpy of the Investigated DNA Duplexes and Netropsin−DNA Complexes[a]

	AAAAG	AAAGA	AAGAA	AAGGA	AAAGG	AAGAG
$\Delta G^{DNA}_{AAAAA\rightarrow}$	−166.9 ± 3.7	−171.8 ± 3.5	−170.5 ± 3.8	−341.9 ± 6.1	−330.3 ± 5.8	−336.5 ± 5.5
$\Delta G^{Net-DNA}_{AAAAA\rightarrow}$	−169.8 ± 3.1	−157.5 ± 4.1	−156.2 ± 3.4	−298.3 ± 6.0	−324.0 ± 5.9	−322.2 ± 5.8
$\Delta\Delta G^{binding}_{AAAAA\rightarrow}$	−2.9 ± 6.8	14.3 ± 7.6	14.3 ± 7.2	43.6 ± 12.1	6.3 ± 11.7	14.3 ± 11.3
$\Delta H^{DNA}_{AAAAA\rightarrow}$	−83.0 ± 2.8	−92.9 ± 2.9	−95.3 ± 4.9	−217.1 ± 6.7	−194.5 ± 4.2	−182.7 ± 4.1
$\Delta H^{Net-DNA}_{AAAAA\rightarrow}$	−92.1 ± 6.2	−79.5 ± 4.1	−79.0 ± 2.6	−177.6 ± 5.4	−181.0 ± 5.2	−177.5 ± 4.7
$\Delta\Delta H^{binding}_{AAAAA\rightarrow}$	−9.1 ± 9.0	13.4 ± 7.0	16.3 ± 7.5	39.5 ± 12.1	13.5 ± 9.4	5.2 ± 8.8
$T\Delta\Delta S^{binding}_{AAAAA\rightarrow}$	−6.2 ± 15.8	−0.9 ± 14.6	2.0 ± 14.7	−4.1 ± 24.2	7.2 ± 21.1	−9.1 ± 20.1

[a] All values are in kJ/mol. Simulation codes refer to Table 1 and Figure 2.

pair mutations in the gas phase of these duplexes and complexes are negligible, the corresponding changes in solution yield indications of the relative stability of these duplexes and complexes. From Table 2, one can see that a replacement of one AT with a GC base pair in a DNA duplex is always favorable, indicating the strong contribution of WC interactions to the DNA duplex stability. In the case of the unbound DNA, the calculated free energy changes due to the substitution of an AT with a GC base pair are very similar, ranging from −166.9 ± 3.7 kJ/mol for the 5′-AAAAG-3′ binding sequence to −171.8 ± 3.5 kJ/mol for the 5′-AAAGA-3′ binding sequence. In the case of the netropsin−DNA complex, larger variations with the binding sequence occur, and the free energy differences range from −156.2 ± 3.4 kJ/mol for the 5′-AAGAA-3′ sequence to −169.8 ± 3.1 kJ/mol for the 5′-AAAAG-3′ binding sequence, supporting the previously discussed differences in the relative free energies of netropsin binding to DNA. Similar observations can also be made for the free energy changes that occur upon substitution of two AT base pairs with GC base pairs. In this case, the calculated free energy differences for free and complexed DNA oligomers are approximately twice as big as in the case of a single AT to GC mutation. The sequence variations of the free energy differences in free DNA duplexes are again small, ranging from −330.3 ± 5.8 kJ/mol for the 5′-AAAGG-3′ sequence to −341.9 ± 6.1 kJ/mol for the 5′-AAGGA-3′ sequence. The free energy differences in the case of the netropsin−DNA complexes show, as expected, bigger variations, ranging from −324.0 ± 5.9 kJ/mol for the 5′-AAAGG-3′ sequence to −298.3 ± 6.0 for the 5′-AAGGA-3′ sequence.

Structural and Energetic Analysis of the Netropsin−DNA Complexes. To put the calculated relative free energies of binding in the context of the structural differences between the netropsin−DNA complexes, we first performed a geometry-based hydrogen bond analysis on each of the seven investigated complexes. The favorable interaction between the netropsin and the DNA minor groove depends on the number and arrangement of hydrogen-bonding donors and acceptors in the DNA minor groove, which is sequence-dependent.[37,41,42,87,105−107] An open question is the size of the contribution of the hydrogen-bonding interactions to the free energy of binding or, in other words, how importantly do the differences in netropsin−DNA hydrogen bonding contribute to the differences in the binding affinities of netropsin for various DNA binding sequences. The results of the hydrogen-bonding analysis presented in Figure 4 give an answer to this question.

As one can see from Figure 4, the total occurrence of netropsin−DNA hydrogen bonds in the case of the binding sequences 5′-AAGAA-3′ and 5′-AAGGA-3′, which are, according to the calculated relative binding free energies, the most unfavorable for netropsin, is much lower than in the case of the favorable 5′-AAAAA-3′ binding sequence. The total occurrence of netropsin−DNA hydrogen bonding is also low for the 5′-AAGAG-3′ binding sequence and for the 5′-AAAGA-3″ binding sequence. Interesting enough, the binding sequences 5′-AAAAG-3′ and 5′-AAAGG-3′ should, according to the total occurrence of netropsin−DNA hydrogen bonding, be more favorable than the 5′-AAAAA-3′ sequence. However, the comparison of the total occurrence of hydrogen bonding between the netropsin and the studied DNA oligomers with the corresponding differences in the relative free energies of binding indicates that a direct correlation between the two does not exist. The binding site preferences for netropsin according to the calculated relative free energies of binding are AAAAA ≈ AAAAG > AAAGG > AAGAA ≈ AAAGA ≈ AAGAG > AAGGA, while the total occurrence of netropsin−DNA hydrogen bonds for different binding sequences descends in the order AAAGG > AAAAG > AAAAA > AAAGA > AAGAA > AAGAG > AAGGA. This observation clearly shows that although hydrogen bonds between netropsin and DNA certainly contribute to the interaction energy between netropsin and DNA and may also be very important in positioning netropsin in the minor groove, they cannot explain the calculated differences in the binding affinities of netropsin for different DNA sequences.

Could inclusion of all DNA−netropsin nonbonding interactions in the analysis explain the order of the calculated relative free energies of binding? To address this question, we have calculated the total energy of interaction of netropsin with the DNA for all of the binding sequences investigated in this work. Moreover, to analyze where the differences in the total nonbonding energy between netropsin and various DNA oligomers come from, we performed an energy partitioning of the total netropsin−DNA interaction energy into contributions that come from each strand and, furthermore, from the DNA backbone and the DNA bases. The results of this analysis are listed in Table 3, and several characteristic features of netropsin−DNA binding emerge from a comparison of the energy components.

Considering the differences in the chemical structure of AT and GC base pairs, one would expect that the low binding affinity of netropsin for the sequences containing GC base pairs comes from the unfavorable interactions between netropsin and DNA. However, inspection of the results collected in Table 3 reveals that despite the presence of the exocyclic −C2 amino group of the guanine base, the variations of the total nonbonding energy between netropsin and DNA are not large for all investigated binding sequences, except for the most unfavorable sequence 5′-AAGGA-3′ with a nonbonding energy of −715.7 ± 26.2 kJ/mol. For the other DNA sequences, the total nonbonding energy between netropsin and DNA ranges from −922.6 ± 15.8 kJ/mol for the 5′-AAAGG-3′ sequence to

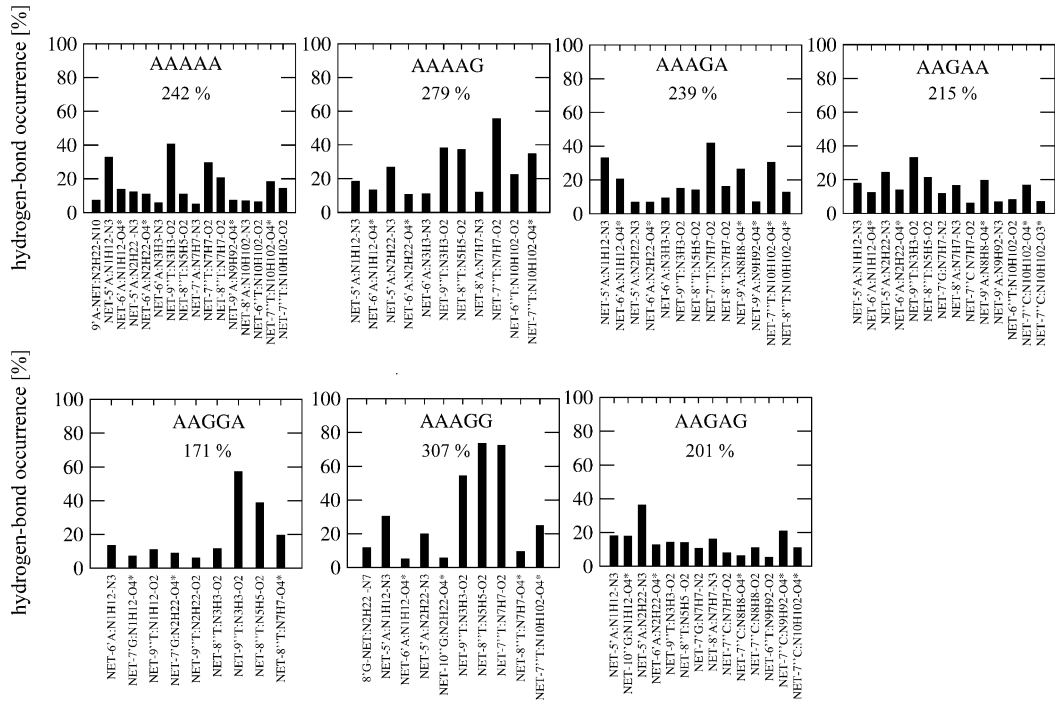


Figure 4. Occurrence of netropsin−DNA hydrogen bonds for the initial and six end states of TI simulations of netropsin−DNA complexes. A hydrogen bond is labeled as a donor molecule or base−acceptor molecule or base: donor atom−acceptor atom. The atom labeling is indicated in Figure 1. Only hydrogen bonds with occurrence greater than 5% are shown. For comparison of the netropsin−DNA hydrogen bonding between the investigated complexes, the sum of the occurrences of the depicted hydrogen bonds is also given.

TABLE 3: Nonbonding Energies (in kJ/mol) of Netropsin−DNA Interactions, E(Net−DNA), Partitioned into Contributions from Interactions of Netropsin with DNA Bases, E(Net−DNA_{bases}), and Backbone, E(Net−$DNA_{backbone}$), for Each of the DNA Strands

	E(Net−DNA_{bases})		E(Net−$DNA_{backbone}$)		E(Net−DNA)
binding site	strand 1	strand 2	strand 1	strand 2	
AAAAA	−156.0 ± 3.2	−153.8 ± 3.5	−313.6 ± 7.0	−310.9 ± 7.4	−934.3 ± 12.3
AAAAG	−142.0 ± 1.4	−167.5 ± 2.2	−347.8 ± 5.2	−335.2 ± 4.6	−992.5 ± 8.6
AAAGA	−104.6 ± 2.0	−165.3 ± 5.5	−322.4 ± 13.2	−353.6 ± 9.4	−945.9 ± 12.4
AAGAA	−134.9 ± 3.4	−180.9 ± 3.5	−308.5 ± 4.8	−312.7 ± 3.6	−937.0 ± 8.1
AAGGA	−47.1 ± 5.4	−154.0 ± 3.1	−160.4 ± 17.0	−354.0 ± 10.6	−715.7 ± 26.2
AAAGG	−84.6 ± 3.4	−167.5 ± 4.5	−310.6 ± 18.6	−359.8 ± 10.6	−922.6 ± 15.8
AAGAG	−108.4 ± 2.8	−195.1 ± 2.8	−295.8 ± 7.0	−329.2 ± 4.4	−928.5 ± 8.13

−992.5 ± 8.6 kJ/mol for the 5′-AAAAG-3′ binding sequence. So, the differences in the free energies of binding cannot only be due to the differences in the interactions between netropsin and DNA but must also originate from differences in DNA solvation and entropic contributions to the free energy of binding. A similar observation has also been made in the case of the interaction of Hoechst 33258 with the d(CG-CAAATTTGCG)$_2$ sequence[3] in which the interactions between the ligand and the DNA appear to contribute little to the overall binding free energy.

An additional insight gained from the partitioning of the total nonbonding energy between netropsin and DNA is that for the sequences where a GC base pair is located in the middle or close to the middle of the binding site, as expected, netropsin is not able to maintain a strong interaction with the bases in the DNA strand containing the guanine base. However, for the sequences 5′-AAGAA-3′ and 5′-AAGAG-3′, this effect is compensated by an increase of favorable interactions of netropsin with the bases in the opposite DNA strand. As can be seen in Table 3, the interaction energy between the netropsin and the bases of the DNA strand 1 increases from −156.0 ± 3.2 kJ/mol for the 5′-AAAAA-3′ sequence up to −47.1 ± 5.4 kJ/mol for the 5′-AAGGA-3′ sequence, while the interaction energy between the netropsin and the nucleotide bases of the DNA strand 2 drops from −153.8 ± 3.5 kJ/mol for the 5′-AAAAA-3′ sequence to −180.9 ± 3.5 and −195.1 ± 2.8 kJ/mol for the 5′-AAGAA-3′ and 5′-AAGAG-3′ sequences, respectively. A similar although not so pronounced compensation also occurs for the 5′-AAAGA-3′ and 5′-AAAGG-3′ binding sequences (see Table 3). The evidence that netropsin can avoid the DNA strand that contains a guanine base and can compensate the loss of these interactions by approaching the opposite DNA strand is further supported by the calculated differences in the interactions between the netropsin and the backbone of DNA strands 1 and 2. In the case of the most favorable binding sequence 5′-AAAAA-3′, the interaction energy between the netropsin and the DNA backbone of strands 1 and 2 is similar, that is, −313.6 ± 7.0 and −310.9 ± 7.4 kJ/mol for strands 1 and 2, respectively. The difference between the interaction energy of netropsin with the backbone of DNA strands 1 and 2 remains small in the case of the sequence 5′-AAAAG-3′ and surprisingly also in the case of the sequence 5′-AAGAA-3′. However, for the 5′-AAAGA-3′, 5′-AAAGG-3′, and 5′-AAGAG-3 binding sequences, the difference in the interaction energies of netropsin with the backbone of the two DNA strands increases to 31.2, 49.2, and 33.4 kJ/mol, respec-

TABLE 4: Contributions to $\Delta H^{DNA}_{AAAAA\rightarrow AAXXX}$ and $\Delta H^{Net-DNA}_{AAAAA\rightarrow AAXXX}$ Values from DNA−DNA, DNA−Solvent, and DNA−Netropsin Interactions for the DNA Duplexes and Netropsin−DNA Complexes Investigated in This Work[a]

	AAAAG	AAAGA	AAGAA	AAGGA	AAAGG	AAGAG
$\Delta H^{DNA}_{AAAAA\rightarrow}$	−83.0 ± 2.8	−92.9 ± 2.9	−95.3 ± 4.9	−217.1 ± 6.7	−194.5 ± 4.2	−182.7 ± 4.1
$\Delta H^{DNA}_{AAAAA\rightarrow}$(DNA−DNA)	−38.4 ± 4.2	−45.4 ± 3.1	−55.1 ± 3.2	−87.1 ± 8.1	−84.7 ± 4.7	−89.5 ± 5.2
$\Delta H^{DNA}_{AAAAA\rightarrow}$(DNA−solv)	−44.6 ± 3.8	−47.5 ± 3.9	−40.2 ± 3.9	−130.0 ± 10.2	−109.8 ± 6.5	−93.2 ± 4.9
$\Delta H^{Net-DNA}_{AAAAA\rightarrow}$	−92.1 ± 6.2	−79.5 ± 4.1	−79.0 ± 2.6	−177.6 ± 5.4	−181.0 ± 5.3	−177.5 ± 4.7
$\Delta H^{Net-DNA}_{AAAAA\rightarrow}$(DNA−DNA)	−36.6 ± 2.7	−55.0 ± 3.3	−30.5 ± 3.5	−81.1 ± 5.9	−83.0 ± 6.6	−72.0 ± 6.7
$\Delta H^{Net-DNA}_{AAAAA\rightarrow}$(DNA−Net)	−3.7 ± 5.6	36.8 ± 4.8	6.0 ± 3.7	43.2 ± 4.1	50.0 ± 8.6	7.6 ± 4.6
$\Delta H^{Net-DNA}_{AAAAA\rightarrow}$(DNA−solv)	−51.8 ± 8.7	−61.3 ± 5.5	−54.5 ± 5.1	−139.7 ± 7.5	−148.0 ± 8.7	−113.1 ± 5.9

[a] All values are in kJ/mol.

tively. The biggest difference in the interactions occurs in the case of the most unfavorable binding site 5′-AAGGA-3′ in which case the difference in the interaction energies of netropsin with the backbone of the DNA strands 1 and 2 equals 193.6 kJ/mol. Note that according to the results listed in Table 3, for all investigated DNA oligonucleotides, the average interaction energy between the netropsin and the DNA backbone is about twice as large as the interaction energy between the netropsin and the DNA bases, which is due to the favorable electrostatic interactions between the positively charged ends of the netropsin molecule and the negatively charged phosphate groups of the DNA backbone.

Enthalpic and Entropic Contributions to the Relative Free Energy of Netropsin−DNA Binding. The results obtained from comparing the binding affinities and interaction energies of netropsin with different DNA sequences motivated us to further address the question of the origin of the differences between the relative binding free energies of netropsin to DNA in terms of the enthalpic and entropic driving forces, which we have estimated using the approach described in the Methods section of this article.

Results of the analysis of the enthalpic and entropic contributions to the relative free energy of binding listed in Table 2 reveal that the observed low binding affinity of netropsin to DNA binding sequences that possess a GC base pair in the middle or close to the middle of the netropsin binding site is almost entirely due to the unfavorable enthalpic contributions to the relative free energy of binding. The enthalpic contribution clearly dominates over the entropic contribution for the 5′-AAGAA-3′, 5′-AAGGA-3′, and 5′-AAAGA-3′ binding sequences. The corresponding $|T\Delta\Delta S|/|\Delta\Delta H|$ ratios equal 0.12, 0.10, and 0.07, and the relative binding enthalpies of 16.3 ± 7.5, 39.5 ± 12.1, and 13.4 ± 7.0 kJ/mol correlate thus well with the relative free energies of binding (see Table 2). In the case of the other binding sequences, in which a GC base pair is located at the end of the binding site, the entropic contributions to the relative free energy of binding become important, with the $|T\Delta\Delta S|/|\Delta\Delta H|$ ratio ranging from 0.53 for the 5′-AAAGG-3′ to 1.7 for the 5′-AAGAG-3′ sequence, respectively. The observed sequence dependence of the enthalpic and entropic driving forces can be due to the structural and solvation differences between different DNA sequences as well as to the apparent differences in the chemical structure and flexibility of the body and the tails of the netropsin molecule,[23,104] since first, the rigid body of netropsin has less freedom to move away from the bulky $-NH_2$ group positioned at the guanine ring than the flexible tails in the case when the GC base pair is positioned at the end of the binding site, and second, the nonbonding interactions between the uncharged pyrole rings in the body of netropsin differ from the nonbonding interactions between the charged netropsin tails and the DNA. From this point of view, the favorable enthalpic contribution to the relative free energy of netropsin binding to the AAAAG sequence is also a rational exception among all of the investigated sequences (see Table 2). The favorable relative enthalpy in this case probably arises from the favorable electrostatic interactions between the netropsin tail and the DNA duplex, which is in agreement with the total netropsin−DNA interaction energy reported in Table 3 and with the total occurrence of netropsin−DNA hydrogen bonds shown in Figure 4. Note that enthalpy and entropy may both (dis)favor binding, as is observed in particular for the 5′-AAGAG-3′ sequence, or may counteract each other in binding, for example, for the 5′-AAAAG-3′ binding sequence. The nearly identical relative free energies of binding of netropsin to the 5′-AAGAG-3′, 5′-AAGAA-3′, and 5′-AAAGA-3′ sequences are due to different enthalpic and entropic contributions.

Origin of the Differences in the Relative Enthalpies of Binding. So far, we have demonstrated that the relative free energies of binding of netropsin to DNA do not correlate with netropsin−DNA interaction energies and that they can only be partially explained in terms of enthalpic and entropic driving forces that govern the netropsin−DNA association. These observations suggest that the calculated differences in the relative enthalpies of netropsin binding to different sequences should arise not only from the changes in the interactions between DNA and netropsin but also from the changes in the structure and solvation of the different DNA oligomers. To confirm this hypothesis, we have partitioned the total relative enthalpy of binding into contributions coming from (i) interactions of the perturbed (λ-dependent) part of the DNA with the rest of the DNA, (ii) interaction of the perturbed part of the DNA with the solvent, and, in the case of the DNA−netropsin complex, (iii) the interaction of the perturbed part of DNA with netropsin:

$$\Delta H^{DNA}_{AAAAA\rightarrow AAXXX} = \Delta H^{DNA}_{AAAAA\rightarrow AAXXX}(DNA-DNA) + \Delta H^{DNA}_{AAAAA\rightarrow AAXXX}(DNA-solv) \quad (5)$$

and

$$\Delta H^{Net-DNA}_{AAAAA\rightarrow AAXXX} = \Delta H^{Net-DNA}_{AAAAA\rightarrow AAXXX}(DNA-DNA) + \Delta H^{Net-DNA}_{AAAAA\rightarrow AAXXX}(DNA-Net) + \Delta H^{Net-DNA}_{AAAAA\rightarrow AAXXX}(DNA-solv) \quad (6)$$

The results are collected in Table 4 and show interesting differences for the different binding sequences. In particular, the difference in the relative enthalpy of the interaction of DNA with the solvent in the case of the unbound DNA and in the case of the complex, $\Delta H^{Net-DNA}_{AAAAA\rightarrow AAXXX}(DNA-solv)-$

$\Delta H^{DNA}_{AAAAA\rightarrow AAXXX}$(DNA−solv), is highly sequence-dependent and can vary from −7.2 kJ/mol for the AAAAG site to −38.2 kJ/mol for the AAAGG site. This shows that the solvent plays an important role in defining the netropsin binding affinities, which is not surprising as it is well-known that for different sequences of DNA the amount of structured water and the DNA−ion interactions differ considerably.[58,61,63−65,69−72,74] Similarly, the relative change in enthalpy of the DNA due to the base pair perturbation varies with the base pair sequences for the unbound DNA duplexes and for the netropsin−DNA complexes. The corresponding differences, $\Delta H^{Net-DNA}_{AAAAA\rightarrow AAXXX}$(DNA−DNA)−$\Delta H^{DNA}_{AAAAA\rightarrow AAXXX}$(DNA−DNA), calculated from Table 4, range from 1.7 kJ/mol in the case of the 5′-AAAGG-3′ sequence to 24.6 kJ/mol in the case of the 5′-AAGAA-3′ sequence and may arise from the differences in the WC and stacking interactions between the base pairs as well as from the slight distortion of the structure of the DNA duplex upon netropsin binding.

Conclusions

Despite the large number of thermodynamic studies in the field of DNA minor groove binding, the driving forces of binding a specific ligand with respect to different binding sequences have not yet been thoroughly investigated. The main goal of this work is to provide a detailed thermodynamic characterization of the binding of the small polyamide netropsin to different binding sequences in the DNA minor groove and to gain understanding of the driving forces that govern the sequence specificity of this binding. Assuming the same binding site for different DNA base pair sequences and employing free energy calculations, we demonstrated that binding of netropsin to the mixed AT/GC sequences is unfavorable unless the GC base pair is located at the end of the binding site. Furthermore, using MD simulations and free energy calculations, we were able to perform a parallel thermodynamic and structural study on the investigated netropsin−DNA complexes. A detailed analysis of the netropsin−DNA hydrogen-bonding and interaction energies showed that these particular interactions cannot fully explain the sequence specific netropsin−DNA binding affinities. Changes in solvation and entropy upon binding show also sequence specific contributions to the free energy of binding. Specifically, the relative free energies of netropsin binding to the mixed AT/GC sequences with the GC base pair located in the middle of the binding site are characterized by large unfavorable enthalpic contributions with entropy playing a minor role. However, for the sequences 5′-AAAAG-3′, 5′-AAAGG-3′, and 5′-AAGAG-3′, where a GC base pair is located at the end of the binding site, the entropy of binding plays an important role as well. Despite the differences in the binding modes, the relative free energies of netropsin binding to the 5′-AAGAA-3′, 5′-AAAGG-3′, and 5′-AAGAG-3′ sequences are, due to enthalpy−entropy compensation or enhancement effects, very similar. Entropic driving forces are rarely taken into account when designing new drugs. Our study suggests that at least in the case of DNA minor groove binders it would be appropriate to do so. Finally, the detailed analysis of the enthalpic contributions to the relative free energies of binding clearly shows that the differences in relative binding enthalpies do not arise only from the differences in the interactions of netropsin with DNA but also from the structural and solvation differences between the different DNA sequences, which demonstrates that interpretation of the drug−DNA binding only in terms of favorable and unfavorable drug−DNA interactions can be rather misleading.

Summarizing, the results presented here demonstrate that a proper statistical−mechanical modeling of the binding or complexation process of relatively flexible biomolecules based on Boltzmann ensembles generated by computer simulation may offer insights in regards to the specific driving forces of binding that are inaccessible by experimental means. This nicely illustrates the complementary roles of simulation and experiment in the process of delineating the driving factors for biomolecular association.

Acknowledgment. Financial support from the National Centre of Competence in Research (NCCR) in Structural Biology, the Swiss National Science Foundation (SNSF) (Grant Number 200020-121913), the European Research Council (ERC) to W.F.G. (Grant Number 228076), and the Slovenian Research Agency (ARRS) (Grant Number Z1-9576 to J.D.) is gratefully acknowledged. We thank Urban Borštnik, Clara Christ, and Chris Oostenbrink for valuable discussions and suggestions.

Supporting Information Available: Figures of the time evolution of the atom-positional rmsd from the energy-minimized ideal B-DNA model structures for unbound DNA oligomers and for DNA−netropsin complexes (Figure S1), occurrence of WC hydrogen bonds in simulations of DNA duplexes and netropsin−DNA complexes (Figure S2), and free energy profiles for the perturbations in the unbound DNA duplexes and the corresponding netropsin−DNA complexes (Figure S3). This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Rentzeperis, D.; Marky, L. A.; Dwyer, T. J.; Geierstanger, B. H.; Pelton, J. G.; Wemmer, D. E. *Biochemistry* **1995**, *34*, 2937.
(2) Chaires, J. B. *Biopolymers* **1997**, *44*, 201.
(3) Haq, I.; Ladbury, J. E.; Chowdhry, B. Z.; Jenkins, T. C.; Chaires, J. B. *J. Mol. Biol.* **1997**, *271*, 244.
(4) Ren, J. S.; Chaires, J. B. *Biochemistry* **1999**, *38*, 16067.
(5) Lah, J.; Vesnaver, G. *Biochemistry* **2000**, *39*, 9317.
(6) Mazur, S.; Tanious, F. A.; Ding, D. Y.; Kumar, A.; Boykin, D. W.; Simpson, I. J.; Neidle, S.; Wilson, W. D. *J. Mol. Biol.* **2000**, *300*, 321.
(7) Haq, I. *Arch. Biochem. Biophys.* **2002**, *403*, 1.
(8) Wang, L.; Kumar, A.; Boykin, D. W.; Bailly, C.; Wilson, W. D. *J. Mol. Biol.* **2002**, *317*, 361.
(9) Lacy, E. R.; Nguyen, B.; Le, M.; Cox, K. K.; O'Hare, C.; Hartley, J. A.; Lee, M.; Wilson, W. D. *Nucleic Acids Res.* **2004**, *32*, 2000.
(10) Lah, J.; Vesnaver, G. *J. Mol. Biol.* **2004**, *342*, 73.
(11) Barceló, F.; Scotta, C.; Ortiz-Lombardía, M.; Méndez, C.; Salas, J. A.; Portugal, J. *Nucleic Acids Res.* **2007**, *35*, 2215.
(12) Freyer, M. W.; Buscaglia, R.; Cashman, D.; Hyslop, S.; Wilson, W. D.; Chaires, J. B.; Lewis, E. A. *Biophys. Chem.* **2007**, *126*, 186.
(13) Freyer, M. W.; Buscaglia, R.; Hollingsworth, A.; Ramos, J.; Blynn, M.; Pratt, R.; Wilson, W. D.; Lewis, E. A. *Biophys. J.* **2007**, *92*, 2516.
(14) Liu, Y.; Kumar, A.; Boykin, D. W.; Wilson, W. D. *Biophys. Chem.* **2007**, *131*, 1.
(15) Munde, M.; Lee, M.; Neidle, S.; Arafa, R.; Boykin, D. W.; Liu, Y.; Bailly, C.; Wilson, W. D. *J. Am. Chem. Soc.* **2007**, *129*, 5688.
(16) Singh, S. B.; Ajay; Wemmer, D. E.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7673.
(17) Singh, S. B.; Kollman, P. A. *J. Am. Chem. Soc.* **1999**, *121*, 3267.
(18) Harris, S. A.; Gavathiotis, E.; Searle, M. S.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **2001**, *123*, 12658.
(19) Špačková, N.; Cheatham, T. E.; Ryjáček, F.; Lankaš, F.; van Meervelt, L.; Hobza, P.; Šponer, J. *J. Am. Chem. Soc.* **2003**, *125*, 1759.
(20) Shaikh, S. A.; Ahmed, S. R.; Jayaram, B. *Arch. Biochem. Biophys.* **2004**, *429*, 81.
(21) Dolenc, J.; Oostenbrink, C.; Koller, J.; van Gunsteren, W. F. *Nucleic Acids Res.* **2005**, *33*, 725.
(22) Marco, E.; Negri, A.; Luque, F. J.; Gago, F. *Nucleic Acids Res.* **2005**, *33*, 6214.
(23) Dolenc, J.; Baron, R.; Oostenbrink, C.; Koller, J.; van Gunsteren, W. F. *Biophys. J.* **2006**, *91*, 1460.
(24) Vargiu, A. V.; Ruggerone, P.; Magistrato, A.; Carloni, P. *Nucleic Acids Res.* **2008**, *36*, 5910.
(25) Vargiu, A. V.; Ruggerone, P.; Magistrato, A.; Carloni, P. *Biophys. J.* **2008**, *94*, 550.

(26) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 1376.
(27) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *J. Mol. Biol.* **1985**, *183*, 553.
(28) Pjura, P. E.; Grzeskowiak, K.; Dickerson, R. E. *J. Mol. Biol.* **1987**, *197*, 257.
(29) Teng, M.; Usman, N.; Frederick, C. A.; Wang, A. H. J. *Nucleic Acids Res.* **1988**, *16*, 2671.
(30) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723.
(31) Parkinson, J. A.; Barber, J.; Douglas, K. T.; Rosamond, J.; Sharples, D. *Biochemistry* **1990**, *29*, 10181.
(32) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393.
(33) Quintana, J. R.; Lipanov, A. A.; Dickerson, R. E. *Biochemistry* **1991**, *30*, 10294.
(34) Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingam, M. *Nat. Struct. Biol.* **1994**, *1*, 169.
(35) Geierstanger, B. H.; Jacobsen, J. P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry* **1994**, *33*, 3055.
(36) Goodsell, D. S.; Kopka, M. L.; Dickerson, R. E. *Biochemistry* **1995**, *34*, 4983.
(37) Neidle, S. *Biopolymers* **1997**, *44*, 105.
(38) Trotta, E.; Paci, M. *Nucleic Acids Res.* **1998**, *26*, 4706.
(39) Gavathiotis, E.; Sharman, G. J.; Searle, M. S. *Nucleic Acids Res.* **2000**, *28*, 728.
(40) Anthony, N. G.; Johnston, B. F.; Khalaf, A. I.; MacKay, S. P.; Parkinson, J. A.; Suckling, C. J.; Waigh, R. D. *J. Am. Chem. Soc.* **2004**, *126*, 11338.
(41) Dolenc, J.; Borštnik, U.; Hodošček, M.; Koller, J.; Janežič, D. *J. Mol. Struct.: THEOCHEM* **2005**, *718*, 77.
(42) Goodwin, K. D.; Long, E. C.; Georgiadis, M. M. *Nucleic Acids Res.* **2005**, *33*, 4106.
(43) Ladbury, J. E.; Chowdhry, B. Z. *Chem. Biol.* **1996**, *3*, 791.
(44) Horn, J. R.; Russell, D.; Lewis, E. A.; Murphy, K. P. *Biochemistry* **2001**, *40*, 1774.
(45) Ladbury, J. E.; Klebe, G.; Freire, E. *Nat. Rev. Drug Discovery* **2010**, *9*, 23.
(46) Cooper, A.; Johnson, C. M.; Lakey, J. H.; Nollmann, M. *Biophys. Chem.* **2001**, *93*, 215.
(47) Lafont, V.; Armstrong, A. A.; Ohtaka, H.; Kiso, Y.; Amzel, L. M.; Freire, E. *Chem. Biol. Drug Des.* **2007**, *69*, 413.
(48) Breslauer, K. J.; Remeta, D. P.; Chou, W. Y.; Ferrante, R.; Curry, J.; Zaunczkowski, D.; Snyder, J. G.; Marky, L. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 8922.
(49) Dunitz, J. D. *Chem. Biol.* **1995**, *2*, 709.
(50) Ladbury, J. E. *Structure* **1995**, *3*, 635.
(51) Gallicchio, E.; Kubo, M. M.; Levy, R. M. *J. Am. Chem. Soc.* **1998**, *120*, 4526.
(52) van der Vegt, N. F. A.; Lee, M. E.; Trzesniak, D.; van Gunsteren, W. F. *J. Phys. Chem. B* **2006**, *110*, 12852.
(53) Yu, H. A.; Karplus, M. *J. Chem. Phys.* **1988**, *89*, 2366.
(54) van Gunsteren, W. F.; Geerke, D. P.; Oostenbrink, C.; Trzesniak, D.; van der Vegt, N. F. A. Analysis of the driving forces for biomolecular solvation and association. In *Protein Folding and Drug Design: Proceedings of the International School of Physics 'Enrico Fermi'*; Broglia, R. A., Tiana, G., Eds.; IOS Press and SIF: Amsterdam and Bologna, 2007; p 177.
(55) Freire, E. *Drug Discovery Today* **2008**, *13*, 869.
(56) Olsson, T. S. G.; Williams, M. A.; Pitt, W. R.; Ladbury, J. E. *J. Mol. Biol.* **2008**, *384*, 1002.
(57) York, D. M.; Yang, W. T.; Lee, H.; Darden, T.; Pedersen, L. G. *J. Am. Chem. Soc.* **1995**, *117*, 5001.
(58) Young, M. A.; Jayaram, B.; Beveridge, D. L. *J. Am. Chem. Soc.* **1997**, *119*, 59.
(59) Cheatham, T. E.; Young, M. A. *Biopolymers* **2000**, *56*, 232.
(60) Cheatham, T. E.; Kollman, P. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 435.
(61) Hamelberg, D.; McFail-Isom, L.; Williams, L. D.; Wilson, W. D. *J. Am. Chem. Soc.* **2000**, *122*, 10513.
(62) McConnell, K. J.; Beveridge, D. L. *J. Mol. Biol.* **2000**, *304*, 803.
(63) Štefl, R.; Koča, J. *J. Am. Chem. Soc.* **2000**, *122*, 5025.
(64) Hamelberg, D.; Williams, L. D.; Wilson, W. D. *J. Am. Chem. Soc.* **2001**, *123*, 7745.
(65) Hud, N. V.; Polak, M. *Curr. Opin. Struct. Biol.* **2001**, *11*, 293.
(66) Orozco, M.; Pérez, A.; Noy, A.; Luque, F. J. *Chem. Soc. Rev.* **2003**, *32*, 350.
(67) Cheatham, T. E. *Curr. Opin. Struct. Biol.* **2004**, *14*, 360.
(68) Kastenholz, M. A.; Hünenberger, P. H. *J. Phys. Chem. B* **2004**, *108*, 774.
(69) Ponomarev, S. Y.; Thayer, K. M.; Beveridge, D. L. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14771.
(70) Rueda, M.; Cubero, E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2004**, *87*, 800.
(71) Várnai, P.; Zakrzewska, K. *Nucleic Acids Res.* **2004**, *32*, 4269.
(72) Spitzer, G. M.; Fuchs, J. E.; Markt, P.; Kirchmair, J.; Wellenzohn, B.; Langer, T.; Liedl, K. R. *ChemPhysChem* **2008**, *9*, 2766.
(73) Kräutler, V.; Hünenberger, P. H. *Mol. Simul.* **2008**, *34*, 491.
(74) Nguyen, B.; Neidle, S.; Wilson, W. D. *Acc. Chem. Res.* **2009**, *42*, 11.
(75) Mark, A. E.; van Gunsteren, W. F. *J. Mol. Biol.* **1994**, *240*, 167.
(76) Smith, P. E.; van Gunsteren, W. F. *J. Phys. Chem.* **1994**, *98*, 13735.
(77) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047.
(78) Levy, R. M.; Gallicchio, E. *Annu. Rev. Phys. Chem.* **1998**, *49*, 531.
(79) Schafer, H.; Mark, A. E.; van Gunsteren, W. F. *J. Chem. Phys.* **2000**, *113*, 7809.
(80) Oostenbrink, C.; van Gunsteren, W. F. *J. Comput. Chem.* **2003**, *24*, 1730.
(81) Peter, C.; Oostenbrink, C.; van Dorp, A.; van Gunsteren, W. F. *J. Chem. Phys.* **2004**, *120*, 2652.
(82) Christ, C. D.; van Gunsteren, W. F. *J. Chem. Phys.* **2007**, *126*, 184110.
(83) Christ, C. D.; van Gunsteren, W. F. *J. Chem. Phys.* **2008**, *128*, 174112.
(84) Lawrenz, M.; Baron, R.; McCammon, J. A. *J. Chem. Theory Comput.* **2009**, *5*, 1106.
(85) Oostenbrink, C. *J. Comput. Chem.* **2009**, *30*, 212.
(86) Singh, N.; Warshel, A. *J. Phys. Chem. B* **2009**, *113*, 7372.
(87) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 1376.
(88) Marky, L. A.; Breslauer, K. J. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 4359.
(89) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.
(90) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction models for water in relation to protein hydration. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, The Netherlands, 1981; p 331.
(91) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300.
(92) King, P. M. Free energy via molecular simulation: a primer. In *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*; van Gunsteren, W. F., Weiner, P. K., Wilkinson, A. J., Eds.; ESCOM Science, B. V.: Leiden, The Netherlands, 1993; p 267.
(93) van Gunsteren, W. F.; Beutler, T. C.; Fraternali, F.; King, P. M.; Mark, A. E.; Smith, P. E. Computation of free energy in practice: Choice of approximations and accuracy limiting factors. In *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*; van Gunsteren, W. F., Weiner, P. K., Wilkinson, A. J., Eds.; ESCOM Science, B. V.: Leiden, The Netherlands, 1993; p 315.
(94) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(95) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(96) Heinz, T. N.; van Gunsteren, W. F.; Hünenberger, P. H. *J. Chem. Phys.* **2001**, *115*, 1125.
(97) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; Vdf Hochschulverlag AG an der ETH Zurich: Zurich, 1996.
(98) Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Kräutler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. *J. Comput. Chem.* **2005**, *26*, 1719.
(99) Soares, T. A.; Hünenberger, P. H.; Kastenholz, M. A.; Kräutler, V.; Lenz, T.; Lins, R. D.; Oostenbrink, C.; van Gunsteren, W. F. *J. Comput. Chem.* **2005**, *26*, 725.
(100) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1994**, *222*, 529.
(101) Kearsley, S. K. *Acta Crystallogr., Sect. A* **1989**, *45*, 208.
(102) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987.
(103) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33.
(104) Wellenzohn, B.; Flader, W.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R. *Biophys. J.* **2001**, *81*, 1588.
(105) Alemán, C.; Vega, M. C.; Tabernero, L.; Bella, J. *J. Phys. Chem.* **1996**, *100*, 11480.
(106) Tabernero, L.; Bella, J.; Alemán, C. *Nucleic Acids Res.* **1996**, *24*, 3458.
(107) Spitzer, G. M.; Wellenzohn, B.; Markt, P.; Kirchmair, J.; Langer, T.; Liedl, K. R. *J. Chem Inf. Model.* **2009**, *49*, 1063.

JP100483F