Short Peptide Fragments Derived from HMG-I/Y Proteins Bind Specifically to the Minor Groove of DNA†

Bernhard H. Geierstanger,‡,§ Brian F. Volkman,‡,∥ Werner Kremer,‡,∥ and David E. Wemmer*,‡,∥

Graduate Group in Biophysics and Department of Chemistry, University of California, Berkeley, California 94720, and the Structural Biology Division, Lawrence Berkeley Laboratory, 1 Cyclotron Road, Berkeley, California 94720

Received December 20, 1993; Revised Manuscript Received February 28, 1994®

ABSTRACT: Short peptides derived from chromosomal proteins have previously been proposed to bind specifically to the minor groove of A,T-rich DNA [for a review, see M. E. A. Churchill and A. A. Travers (1991) *Trends Biochem. Sci. 16*, 92–97]. Using NMR spectroscopy, we investigated the DNA binding of SPRKSPRK, which is one such A,T-specific motif. Under the conditions studied SPRKSPRK interacts only nonspecifically with d(CGCAAAAAAGGC)·d(GCCTTTTTTGCG). The peptides TPKRPRGRPKK, PRGRPKK, and PRGRP derived from the non-histone chromosomal protein HMG-I/Y, however, bind specifically to the central A,T sites of $d(CGCAAATTTGCG)_2$ and $d(CGCGAATTCGCG)_2$. 2D NOE measurements show that the RGR segment of each peptide is in contact with the minor groove. The arginine side chains and the peptide backbone are buried deep in the minor groove, in a fashion generally similar to the antibiotic netropsin. Under the same conditions the peptide PKGKP does not interact with the same oligonucleotide duplexes, indicating that the arginine guanidinium groups are major determinants of the A,T specificity.

In eukaryotic cells, DNA is packaged into chromatin by association with histones and other proteins. Many of these proteins have been shown to preferentially bind to A,T-rich DNA, and it has been suggested that some chromosomal proteins recognize the same structural features of DNA as minor groove binding ligands such as distamycin [for a review, see Churchill and Travers (1991)].

Suzuki (1989a) proposed such a netropsin/distamycin-like binding motif consisting of SPXY repeats, where X and Y are arginine or lysine. SPXY repeats are found in histones H1 and H2B from sea urchin sperm, proteins which are essential for chromatin assembly and organization. The SPXY sequence is also frequently found in regulatory DNA binding proteins with the serine residue being a possible phosphorylation site (Suzuki, 1989b; Hill et al., 1990). According to footprinting studies, $(SPRK)_6$ and SPRKSPRK bind specifically to an $A_6 \cdot T_6$ site, presumably in the minor groove (Churchill & Suzuki, 1989).

The peptide TPKRPRGRPKK from the DNA binding domain of the non-histone chromosomal protein HMG-I/Y was shown to have the same DNA binding characteristics as the intact protein (Reeves & Nissen, 1990). Footprinting results and competition assays with the minor groove binding ligands Hoechst 33258 (Reeves & Nissen, 1990; Radic et al., 1992) and distamycin (Radic et al., 1992) indicated that TPKRPRGRPKK binds to the minor groove of A,T-rich DNA also. On the basis of these results Reeves and Nissen (1990) proposed TPKRPRGRPKK as a novel DNA binding motif ("A,T-hook"): a crescent-shaped peptide with a TPKR turn which would bind to the minor groove in a fashion similar to netropsin and distamycin. The backbone amides of the PRGRPKK segment were suggested to hydrogen bond with DNA acceptor groups.

Distamycin, netropsin, and the dye Hoechst 33258 bind preferentially to the narrow minor groove of A,T-tract DNA sequences (Zimmer & Wähnert, 1986). These crescent-shaped ligands fit tightly into the minor groove. In the case of netropsin and distamycin the amide protons of the peptide bonds linking the pyrrole rings point into the minor groove, forming hydrogen bonds with acceptor atoms of thymine and adenine bases [Kopka et al., 1985; Coll et al., 1987; Tabernero et al., 1993; for a review, see Wemmer et al. (1994)]. The charged end groups of these ligands interact electrostatically with the DNA, and evidence for specific hydrogen bonding to DNA bases has also been presented (Rhee et al., 1993).

We set out to test the proposal that SPRKSPRK and TPKRPRGRPKK bind to the minor groove of DNA in a fashion similar to netropsin and distamycin by using NMR spectroscopy to directly examine the complexes. In contrast to footprinting studies (Churchill & Suzuki, 1989), our experiments yielded no evidence for the formation of a specific complex of SPRKSPRK with d(CGCAAAAAAGGC)·d(GCCTTTTTTGCG). However, TPKRPRGRPKK did form rapidly exchanging but specific complexes with A,T sites in oligonucleotide duplexes. After initial NOE studies had suggested that the primary region of contact was in the PRGRP sequence, the peptides PRGRPKK and PRGRP were synthesized and studied. All three PRGRP-type peptides show similar NOE contacts with DNA protons exposed in the minor groove. Titration data indicate that the binding is relatively weak (association constants in the millimolar range), with bound and free peptide near the fast-exchange limit.

The peptide–DNA complexes have been studied at stoichiometries above 1:1 to try to maximize the fraction of complex present, and we estimate >70% of the DNA is complexed under the conditions used. Under these conditions

† We are grateful to the National Institutes of Health (GM-43129) and to the Office of Energy Research, Office of Health and Environmental Research, Health Effects Research Division of the U.S. Department of Energy, under Contract No. DE-AC03-76SF00098 for research support, and to the Deutschen Forschungsgemeinschaft for a postdoctoral fellowship (to W.K.). Instrumentation grants from the U.S. Department of Energy, DE FG05-86ER75281, and the National Science Foundation, DMB 86-09035 and BBS 87-20134, are also gratefully acknowledged.

* To whom correspondence should be addressed.
‡ Lawrence Berkeley Laboratory.
§ Graduate Group in Biophysics.
∥ Department of Chemistry.

® Abstract published in *Advance ACS Abstracts,* April 1, 1994.

0006-2960/94/0433-5347$04.50/0

there is an equilibrium between free and bound peptide, in fast exchange and thus giving rise to just one set of NMR resonances. Intermolecular NOEs must arise from the complex, while free and bound forms will contribute to intramolecular NOEs. However, since the molecular weight of the complex is much higher than that of the peptide alone, intramolecular NOEs in the peptide arise predominantly from the bound state, as is well known in exchange transferred NOE studies in which a large excess of ligand is used [Clore & Gronenborn, 1982; also reviewed in Sykes (1993)]. Assuming a single binding mode, the NOE data can be used to derive semiquantitative distance restraints to determine a model of the peptide–DNA complex, although the uncertainties are somewhat larger than in studies of nonexchanging complexes. Molecular modeling using such restraints shows that the RGR motif of the PRGRP-type peptides sits deep in the minor groove of a 4-base-pair A,T site, with the arginine side chains resembling the guanidinium and amidinium groups of netropsin and determining the specificity of the peptides for A,T sequences.

MATERIALS AND METHODS

Synthesis and Purification of Oligonucleotides and Peptides. Oligonucleotides were synthesized and purified as previously described (Pelton & Wemmer, 1989). Peptides were synthesized on an Applied Biosystems 430A synthesizer using standard Fmoc chemistry. All peptides were amidated at the C-terminus, and HPLC purified on a C18 reverse-phase column using a 0.1% trifluoroacetic acid/H_2O/acetonitrile gradient. Purity and identity of the peptides were confirmed by mass spectroscopy and NMR.

Sample Preparation. Dry oligonucleotide samples were dissolved in 0.5 mL of 10 mM sodium phosphate or 10 mM deuterated Tris buffer. These solutions were adjusted to either pH 7.0 or pH 5.6 as indicated and lyophilized to dryness. For experiments in D_2O the solid was redissolved in 0.5 mL of 99.96% D_2O (Cambridge Isotope Laboratories), while for experiments in H_2O a 90% H_2O/10% D_2O mixture (0.5 mL) was used. The concentrations of double-stranded DNA samples were between 0.46 and 3.3 mM as indicated on the basis of UV absorbance measurements at 80 °C using the following extinction coefficients, ϵ_{260}: 1.24×10^5 M^{-1} cm^{-1} for d(CGCAAAAAAGGC), 1.00×10^5 M^{-1} cm^{-1} for d(GCCTTTTTTGCG), 1.13×10^5 M^{-1} cm^{-1} for d(CGCAAATTTGCG), and 1.11×10^5 M^{-1} cm^{-1} for d(CGCGAATTCGCG) (Warshaw & Cantor, 1970). For the titrations, peptide solutions of 12.5–55 mM in 90% H_2O/10% D_2O and 10 mM buffer solutions were prepared. Peptide concentrations are based on the weight of the lyophilized material after HPLC purification.

NMR Experiments and Signal Assignments. NMR experiments were performed at 600 MHz on a Bruker AMX-600 spectrometer or at 500 MHz on a GE Omega 500. Peptides were titrated into the NMR sample containing duplex DNA in approximately 0.2 or 0.25 molar equiv per addition. 1D spectra in H_2O were acquired with 8192 complex points and a spectral width of 13 514 Hz (600 MHz) or 10 000 Hz (500 MHz). One hundred twenty-eight scans were averaged for each spectrum. 2D NOESY spectra in H_2O were acquired by collecting 2048 complex points in t_2 using a spectral width of 13 514 Hz (600 MHz) or 10 000 Hz (500 MHz). Suppression of the solvent signal was achieved by using a 1–1 jump and return sequence (Sklenar et al., 1987) for the last pulse. 2D TOCSY and DQF-COSY spectra in H_2O were recorded using a low-power presaturation pulse for solvent suppression. 2D NOESY, TOCSY, and DQF-COSY spectra in D_2O were collected with 1024 complex points in t_2 using a spectral width of 6024 Hz (600 MHz) or 5000 Hz (500 MHz). Solvent suppression was achieved with presaturation. The mixing time for all NOESY experiments was 200 ms, while a 60-ms MLEV-17 spin-lock pulse (Davis & Bax, 1985) or a modification of it (Griesinger et al., 1988) was used for isotropic mixing in the TOCSY or Clean-TOCSY experiments. For NOESY spectra in H_2O 486–576 t_1 experiments with 48–96 scans were recorded and zero-filled to 2K complex points. For experiments in D_2O 332–606 t_1 experiments with 16–80 scans were recorded and zero-filled to 1K complex points. All 2D spectra were acquired in phase-sensitive mode using TPPI. The recycle delay for all spectra was 2 s. The data were processed with FELIX (Hare Research) on a Silicon Graphics workstation. Shifted skewed sine-bell functions were used for apodization of the free-induction decays. 2D NOESY, TOCSY, and DQF-COSY spectra of the peptide–DNA complexes in D_2O and H_2O allowed the assignment of the DNA and peptide resonances using standard sequential methods (Hare et al., 1983; Wüthrich, 1986). Geminal peptide protons were labeled according to their chemical shift; e.g., for arginine R8 Cβ protons, R8b1 is downfield of R8b2.

Distance Restraints. Models of the PRGRP peptide with $d(CGCAAATTTGCG)_2$ and $d(CGCGAATTCGCG)_2$ were calculated on the basis of the NMR data. Distance restraints were generated from the volume integrals of the cross peaks in the D_2O and H_2O NOESY spectra acquired at a mixing time of 200 ms. Volumes from H_2O spectra were corrected for the 1–1 excitation profile as described previously (Dwyer et al., 1992). Volumes of intermolecular peptide–DNA cross peaks were multiplied by 1.25 to correct for partial occupancy of the DNA site (approximately 80%). Distance restraints were calculated relative to the volume integrals of the cytosine H5–H6 cross peak volumes and classified as strong (1.8–3.0 Å), medium (2.5–4.0 Å), and weak (3.5–5.5 Å). Because many restraints involved pseudoatoms, and to accommodate uncertainties due to partial occupancy of the complex, the upper bounds of the restraints were set higher than usual. In all, 12 intermolecular peptide–DNA restraints and 13 intramolecular restraints of the peptides were used for the model of PRGRP bound to $d(CGCGAATTCGCG)_2$ presented in Figure 7. (Listings of the intermolecular peptide–DNA and intramolecular peptide restraints are available as supplementary material.)

Molecular Modeling. The complexes of PRGRP with $d(CGCAAATTTGCG)_2$ and $d(CGCGAATTCGCG)_2$ were modeled using restrained energy minimization and molecular dynamics. Models of the DNA duplexes in standard B-form and the peptide molecule were constructed using the Biopolymer module of Insight II (Biosym). The peptide was first energy minimized using the intramolecular peptide restraints only. The peptide was then manually docked into the minor groove with the help of the Builder module of Insight II on a Silicon Graphics workstation. This procedure assures that unfavorable peptide–DNA contacts in the early phases of minimization will not prevent the peptide molecule from reaching a conformation consistent with the intramolecular peptide restraints. Energy minimizations on the peptide–DNA complexes were then performed using the Discover module of Insight II. Hydrogen bonds for standard Watson–Crick base pairing were included as NOE restraints using a force constant of 200 (kcal/mol)/Å^2, while a force constant of 25 (kcal/mol)/Å^2 was used for the experimentally derived NOE restraints. The cutoff distance for nonbonded interac-

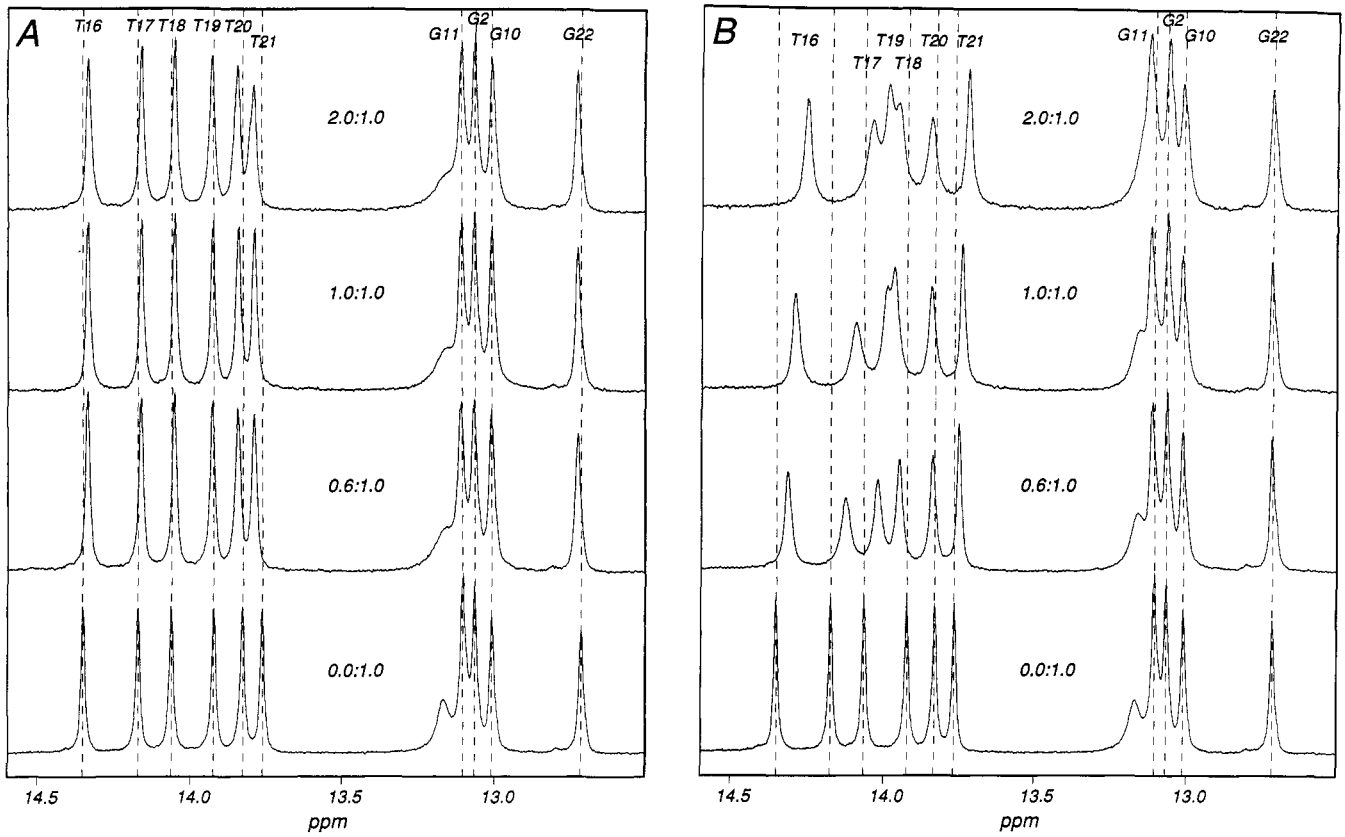

FIGURE 1: Imino proton region of ^{1}H NMR spectra acquired at several points in titrations of d(CGCAAAAAAGGC)·d(GCCTTTTTTGCG) (2.05 mM in 10 mM phosphate buffer at pH 7) with (A) SPRKSPRK and with (B) TPKRPRGRPKK at 15 °C (600 MHz). Approximate peptide to DNA ratios are indicated. Chemical shifts are referenced to a residual HDO line at 4.91 ppm.

tions was set at 15 Å with a switching distance of 2 Å. A distance-dependent dielectric of the form $\epsilon = R$ was used to account for solvent effects. The energy of the complexes was initially minimized using 100 steps of a steepest descents algorithm, and a final rms derivative of <0.001 (kcal/mol)/Å^2 was achieved in fewer than 20 000 steps of conjugate gradient minimization. To allow for rearrangement of the peptide side chains, a low-temperature (100 K) dynamics run was carried out followed by a final energy-minimization step with conjugate gradient. Including the dynamics step in this procedure, as in the model of PRGRP bound to d(CGCGAATTCGCG)$_2$ presented in Figure 7, made few qualitative differences to the models obtained with just energy minimization.

RESULTS

Binding of SPRKSPRK. Substoichiometric amounts of SPRKSPRK were added to millimolar NMR samples of d(CGCAAAAAAGGC)·d(GCCTTTTTTGCG). The imino protons of the DNA duplex were monitored during the titration with peptide at 15 °C (Figure 1A). The chemical shifts of all DNA protons had changed by less than 0.03 ppm after the addition of 2 molar equiv of SPRKSPRK to d(CGCAAAAAAGGC)·d(GCCTTTTTTGCG), suggesting that the peptide has low affinity for the A_6 site. The largest spectral changes are seen for the thymidine imino protons, which may indicate that what little binding occurs does occur in this region. The titration was repeated at 5 °C, yielding the same conclusions (data not shown). When the titration was continued to a 6:1 molar ratio of peptide to DNA, precipitation occurred in the NMR tube, and the signal-to-noise ratio of the NMR spectra decreased steadily (data not shown). This behavior is qualitatively similar to that of polyamines, which bind nonspecifically and lead to precipitation through charge neutralization. The analysis of 2D NOESY spectra of the sample at a 2:1 molar ratio of SPRKSPRK to A_6 DNA revealed no evidence for any contacts between DNA and peptide protons (data not shown). Identical results were obtained when SPRKSPRK was added to d(CGCAAATTGGC)·d(GCCAATTTGCG) (data not shown), further supporting the conclusion that SPRKSPRK interacts only nonspecifically with DNA under the conditions studied.

Binding of TPKRPRGRPKK. Addition of substoichiometric amounts of TPKRPRGRPKK to d(CGCAAAAAAGGC)·d(GCCTTTTTTGCG) at 15 °C caused shifting and significant line broadening of the DNA imino proton resonances (Figure 1B). The presence of only one set of resonances indicates that the kinetic off rate of the peptide is near the fast-exchange limit. Even at a 2:1 peptide:DNA molar ratio the resonances continue to shift, indicating that saturation of binding sites has not been reached. NOE contacts between the peptide and the adenine C2 protons as well as sugar C1′ protons verify binding of TPKRPRGRPKK to the minor groove of the A_6 site (data not shown). These contacts included protons of both arginine R6 and arginine R8 side chains and glycine G7 protons. The NOE contacts can only be explained if TPKRPRGRPKK binding occurs in both orientations on the DNA. Due to exchange between orientations, and because of severe spectral overlap, binding studies of TPKRPRGRPKK were continued with the symmetric oligonucleotide duplex d(CGCAAATTTGCG)$_2$ as discussed in detail in the following section.

Binding of TPKRPRGRPKK, PRGRPKK, PRGRP, and PKGKP to d(CGCAAATTTGCG)$_2$. In addition to binding of TPKRPRGRPKK, binding of the peptides PRGRPKK, PRGRP, and PKGKP to d(CGCAAATTTGCG)$_2$ was also

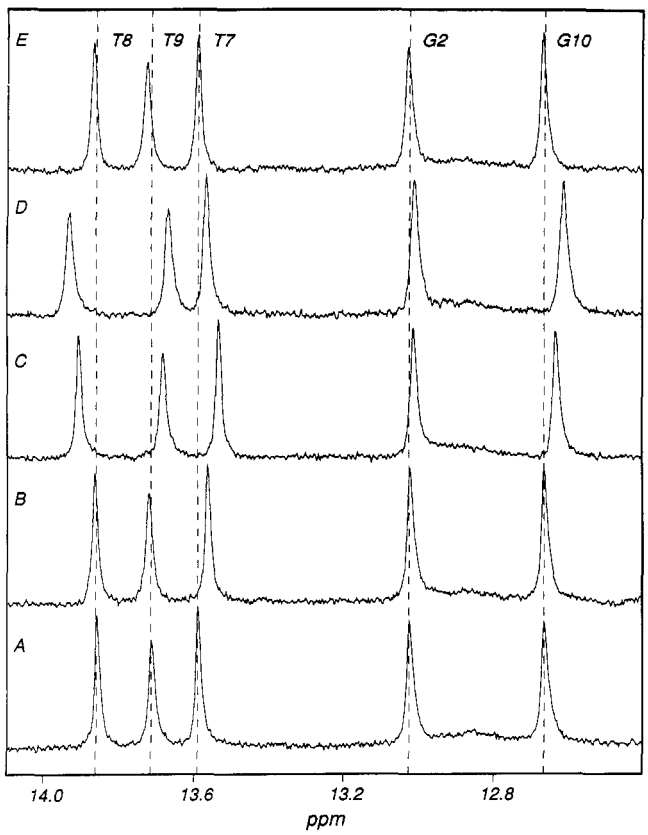

FIGURE 2: Imino proton region of ^{1}H NMR spectra of d(CGCAAATTTGCG)$_2$ (0.46 mM in 10 mM phosphate buffer at pH 5.6) after the addition of 2 molar equiv of peptide per DNA duplex: (A) free DNA, (B) PRGRP, (C) PRGRPKK, (D) TPKRPRGRPKK, and (E) PKGKP at 25 °C (500 MHz). Samples contain approximately 0.025 mM EDTA. Chemical shifts are referenced to TSP standard at 0.00 ppm.

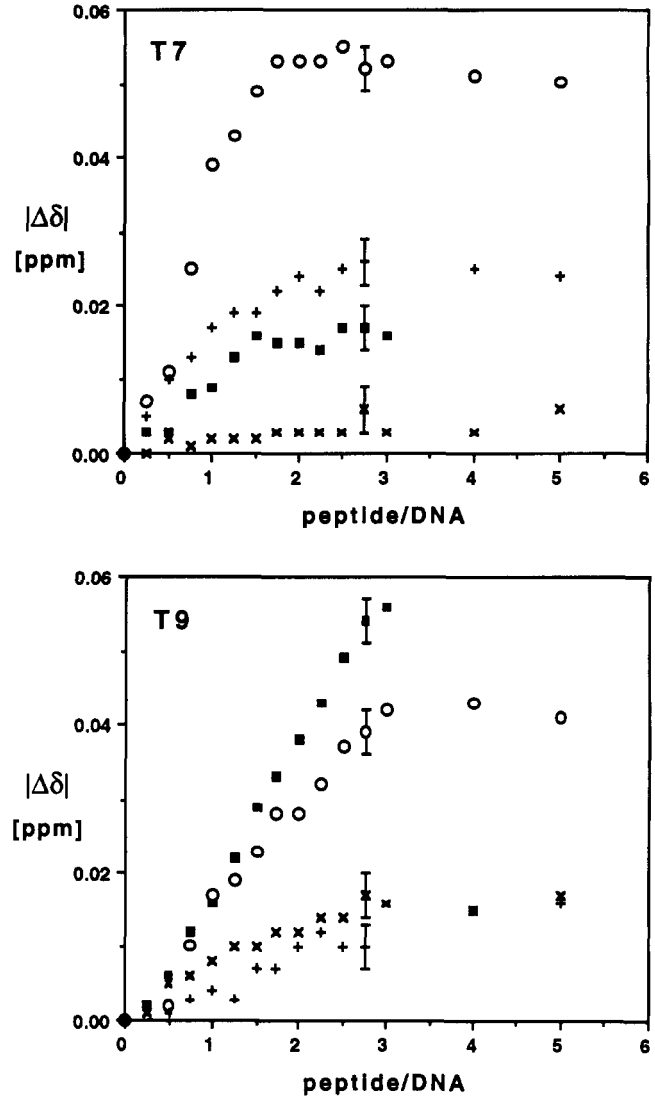

FIGURE 3: Chemical shift changes (absolute values) of selected imino protons of d(CGCAAATTTGCG)$_2$ titrated with TPKRPRGRPKK (■), PRGRPKK (○), PRGRP (+), and PKGKP (×) at 25 °C (500 MHz) as a function of the peptide:DNA ratio. Conditions are the same as in Figure 2.

investigated. Figure 2 shows the imino proton resonances of d(CGCAAATTTGCG)$_2$ after addition of 2 molar equiv of PRGRP, PRGRPKK, TPKRPRGRPKK, and PKGKP at 25 °C and pH 5.6. The imino protons of all A·T base pairs change similarly in the PRGRPKK and TPKRPRGRPKK titrations (Figure 2C,D), while in the case of PRGRP only the T7 imino line changes significantly (Figure 2B). The resonance line of the guanine imino closest to the A_3T_3 site is shifted upfield only upon binding of PRGRPKK and TPKRPRGRPKK. This suggests that the PRGRP part of the peptides binds specifically to the A_3T_3 region of the oligonucleotide duplex, while parts of the longer peptides are also associated with the flanking G·C base pairs.

The absolute chemical shift changes of the T7 and T9 imino resonances are compared in Figure 3. Different asymptotic values are reached at high peptide:DNA ratios for the titration with TPKRPRGRPKK, PRGRPKK, and PRGRP, suggesting small differences in the local interactions between peptides and DNA. Because of electrostatic attraction, lysine and arginine side chains will contribute nonspecifically to the binding affinity. One would therefore expect the longer peptides to bind with higher affinity and that the asymptotic values corresponding to the chemical shifts of the peptide–DNA complex should be reached at lower peptide:DNA ratios. The uncertainties of the titrations make this conclusion unclear. The chemical shifts of the imino protons of T7 and T8 (not shown) at the center of the A_3T_3 site change linearly up to a peptide:DNA ratio of approximately 1.25:1 and start to level off at higher ratios. The T9 and G10 imino protons (not shown) do not follow this trend, indicating a more complicated binding behavior at high peptide:DNA ratios.

To test the basis for specific interaction of the peptide side chains and DNA, we replaced the arginine residues of the PRGRP peptide with lysines to give PKGKP. Upon titration with the lysine-containing peptide the imino resonance lines of the A_3T_3 oligomer do not shift significantly (Figures 2E and 3). This suggests that the binding affinity of PKGKP for DNA must be much lower than that of the arginine-containing peptides. The absence of any NOE cross peaks between DNA protons and peptide protons in 2D NOESY spectra reinforces this interpretation (data not shown). Although the lysine and arginine side chains are both positively charged at pH 5.6, there are significant differences in both shape and hydrogen-bonding capability between the amino and guanidinium groups which are likely responsible for the difference in binding specificity and affinity.

Analysis of the 2D NMR Spectra of TPKRPRGRPKK, PRGRPKK, and PRGRP Complexes with d(CGCAAATTTGCG)$_2$. 2D NOESY, DQF-COSY, and TOCSY spectra in H_2O and D_2O were acquired for TPKRPRGRPKK, PRGRPKK, and PRGRP complexes with d(CGCAAATTTGCG)$_2$ at a peptide:DNA ratio of 1.25:1. Analysis of 2D NOESY data shows that PRGRP, PRGRPKK, and

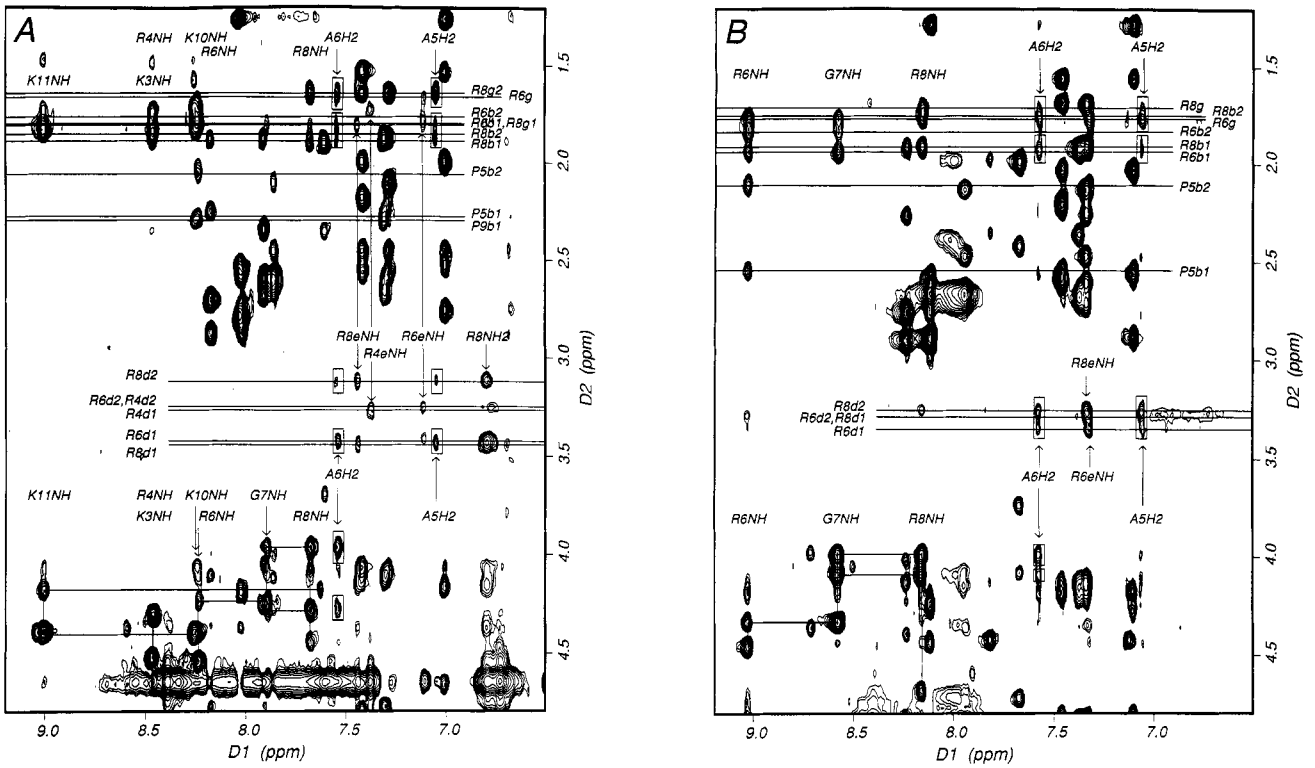

FIGURE 4: Parts of the NOESY spectra of d(CGCAAATTTGCG)$_2$ titrated with (A) TPKRPRGRPKK (35 °C, 600 MHz, 1 mM DNA duplex in 10 mM phosphate buffer, pH 5.6; peptide:DNA ratio, 1.25:1; 9:1 H_2O/D_2O; τ_{mix} = 200 ms) and (B) PRGRP (15 °C, 600 MHz, 3.3 mM DNA duplex in 10 mM Tris buffer, pH 7; peptide:DNA ratio, 1.25:1; 9:1 H_2O/D_2O; τ_{mix} = 200 ms). The sequential CαH to NH connectivities for the peptides are shown with solid lines. Intermolecular peptide–DNA cross peaks are boxed. The approximate positions of selected aliphatic peptide protons are shown as horizontal lines (CβH = b, CδH = d, and CγH = g). Exchangeable peptide protons (amide = NH, arginine ϵ = eNH, and arginine guanidinium protons = R8NH2) and adenine C2 protons (C2H = H2) are labeled.

TPKRPRGRPKK have similar intermolecular peptide–DNA contacts (Figures 4 and 5) and intramolecular NOEs (Figure 6) when bound to d(CGCAAATTTGCG)$_2$. (Listings of the intermolecular peptide–DNA NOEs and the sequential intrapeptide NOEs, as well as complete chemical shift assignments, are available for all complexes as supplementary material.) Intermolecular peptide–DNA NOEs are only observed between peptide protons of residues R6, G7, and R8 of the PRGRP sequence and DNA protons exposed in the minor groove. These contacts place the side chains and the backbone of the RGR part of the peptides into the minor groove of the central A_3T_3 site of the DNA duplex (Figure 5). The arginine side chains of R6 and R8 must point in opposite directions with the glycine G7 residue close to the AT step. No NOE contacts from the TPKR residues or the C-terminal lysine residues of TPKRPRGRPKK or PRGRPKK to DNA were observed.

Intramolecular NOEs (Figure 6) indicate that the peptides are in an extended conformation when bound to DNA. Strong NOE cross peaks between proline Cδ and the Cα protons of the preceding residue confirm that all proline residues are in the *trans* conformation. Additional spin systems detected in the TOCSY and DQF-COSY spectra of the TPKRPRGRPKK and PRGRPKK complexes suggest a small population of peptides with *cis*-proline isomers. *cis*-Proline isomer resonances were observed for PRGRP and PKGKP free in solution (25% *cis*-X-Pro) but not for the longer peptides. No NOEs were observed which would be indicative of a turn structure in the TPKR sequence as proposed for the "A,T-hook motif" (Reeves & Nissen, 1990).

Sequential intra-DNA NOE patterns are indicative of standard B-form DNA (Hare et al., 1983). The aromatic C6H/C8H chemical shifts do not change significantly upon peptide binding (less than 0.05 ppm difference from free DNA). Small but significant changes (0.06–0.18 ppm) are observed for the C1′ protons of A6, T7, T8, and T9 in the peptide binding site. The adenine C2 protons of A5 and A6 do not change significantly, while A4 C2H changes by approximately 0.1 ppm, possibly reflecting the proximity of the charged arginine groups at the ends of the A_3T_3 site. These data suggest that peptide binding does not significantly perturb the DNA conformation.

Modeling of the PRGRP Complex with d(CGCAAATTTGCG)$_2$. Intermolecular peptide–DNA and intrapeptide restraints derived from NOESY data were used to model the PRGRP complex with d(CGCAAATTTGCG)$_2$. The NOE restraints were inconsistent with peptide binding to a single site. This observation suggested that PRGRP may bind to different parts of the 6-base-pair A_3T_3 site and/or slide rapidly between these subsites. In order to improve the model for the interaction of PRGRP with the minor groove of DNA, we therefore investigated the binding of PRGRP to d(CGCGAATTCGCG)$_2$ as well.

Characterization and Modeling of the PRGRP Complex with d(CGCGAATTCGCG)$_2$. Titration of d(CGCGAATTCGCG)$_2$ with PRGRP causes chemical shift changes comparable to PRGRP binding to the A_3T_3 oligonucleotide duplex (data not shown). The PRGRP peptide contacts with the A_2T_2 site are similar in nature to those observed with A_3T_3 (Figure 5). Distance restraints derived from NOESY spectra in D_2O and H_2O were used to model the complex of PRGRP with d(CGCGAATTCGCG)$_2$. A model consistent with the NMR data is shown in Figure 7. The peptide lies deep in the minor groove and spans all four base pairs of the A_2T_2 site. Hydrogen bonds were assigned by the Insight program between the glycine G7 amide and the O2 of T8

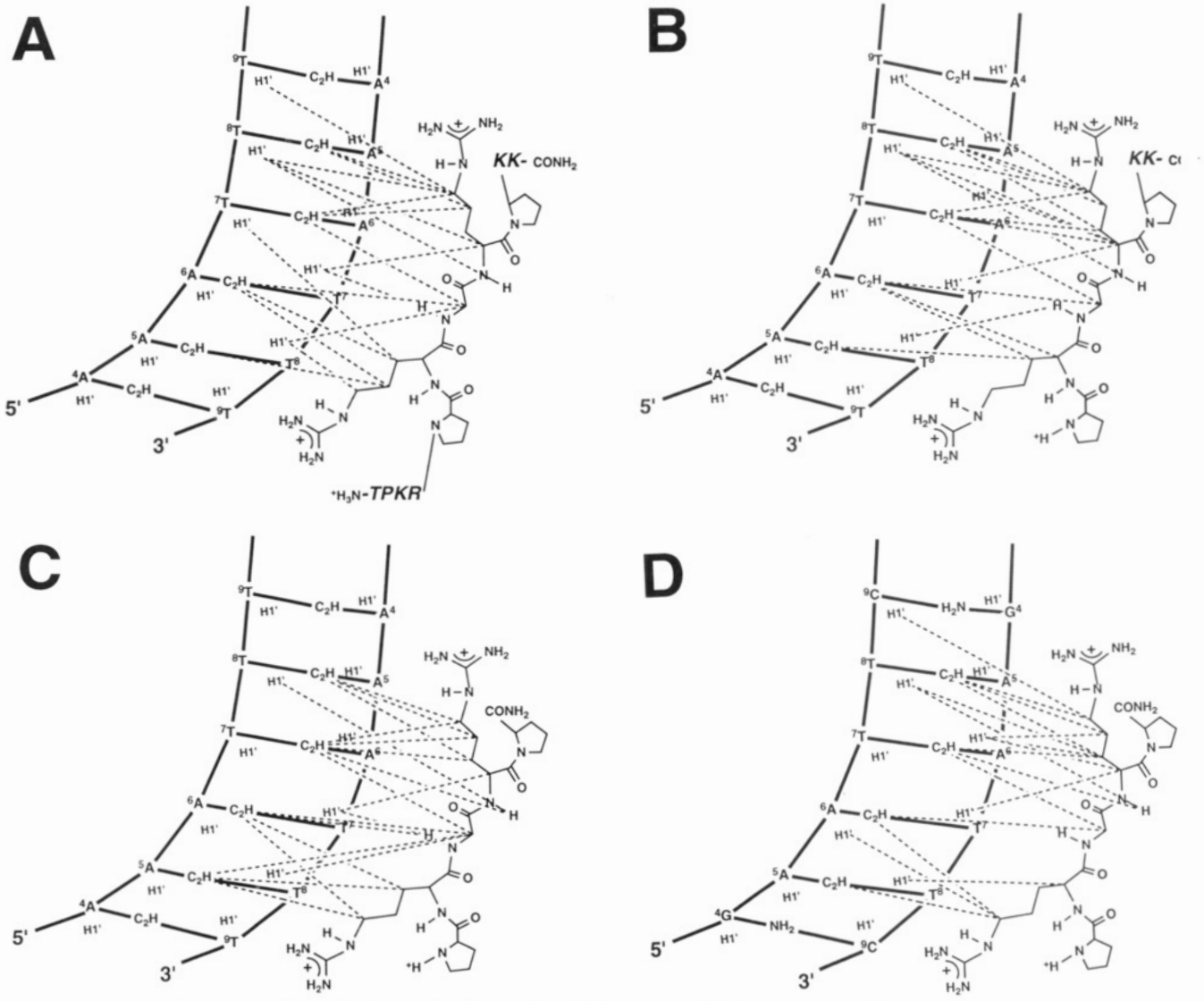

FIGURE 5: Schematic representation of the intermolecular peptide–DNA contacts of (A) TPKRPRGRPKK, (B) PRGRPKK, and (C) PRGRP bound to d(CGCAAATTTGCG)$_2$ and (D) PRGRP bound to d(CGCGAATTCGCG)$_2$ as determined from ^{1}H NOESY spectra in H_2O and D_2O. Only intermolecular peptide–DNA contacts that could be assigned unambiguously are shown. Listings of all intermolecular NOEs are available as supplementary material. For clarity, contacts are drawn to the peptide carbon centers.

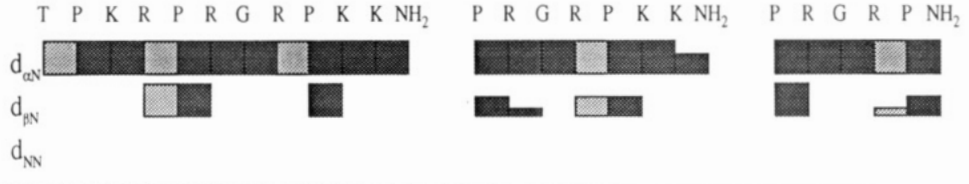

FIGURE 6: Sequential NOE diagram for the peptides TPKRPRGRPKK, PRGRPKK, and PRGRP bound to d(CGCAAATTTGCG)$_2$. The intensity of the interresidue NOE is indicated by the height of the shaded block. Lighter shading indicates NOEs involving proline Cδ protons in place of backbone amide protons.

exposed in the minor groove of the central A_2T_2 sequence. The guanidinium protons of the arginine side chain of R6 form hydrogen bonds to N3 of adenine A5. The arginine side chains point in opposite directions with the positively charged guanidinium groups presented to phosphate groups on opposite strands as well as to the bases A5 and C9. Positions and orientations of the guanidinium groups varied slightly in different energy minimizations and during molecular dynamics runs. This suggests that there are probably several different, energetically nearly equivalent positions for the ends of the side chains. The position of the peptide backbone in different runs was quite reproducible. The orientations of the arginine side chains resemble those of the charged end groups of netropsin (Figure 8). As in the case of netropsin, electrostatic interactions of the side chains and hydrogen bonding of side-chain protons and backbone amides stabilize the peptide–DNA complex. van der Waals interactions between the glycine and arginine methylene protons and the wall of the DNA groove also seem to contribute to the binding affinity of the PRGRP peptide.

DISCUSSION

Nonspecific Binding of SPRKSPRK. On the basis of footprinting studies at high peptide to binding site ratios Churchill and Suzuki (1989) reported specific binding of SPRKSPRK to A_6 sites. Our NMR titration data provide little evidence for the formation of a specific complex of

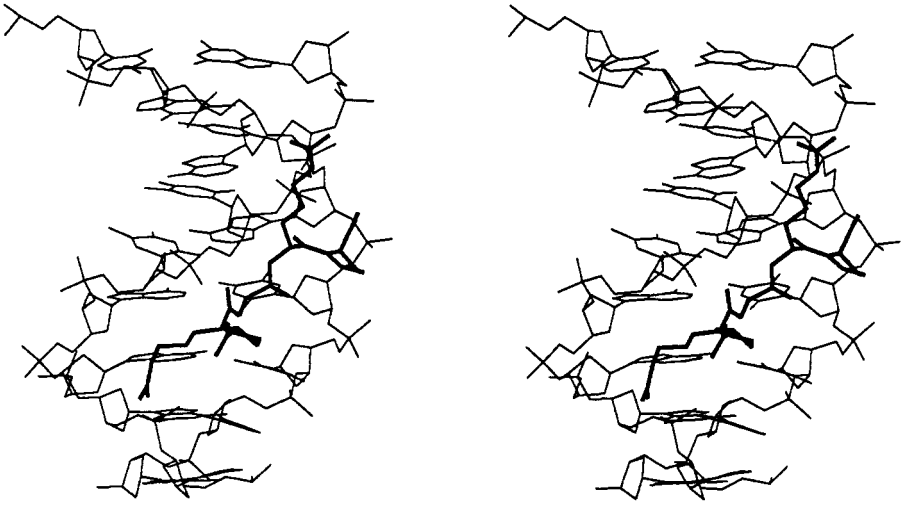

FIGURE 7: Stereo diagram of a molecular model consistent with the NOESY data of PRGRP bound to d(CGCGAATTCGCG)$_2$. All protons are omitted for clarity. Only the central eight base pairs are shown. The peptide is shown in thick solid lines.

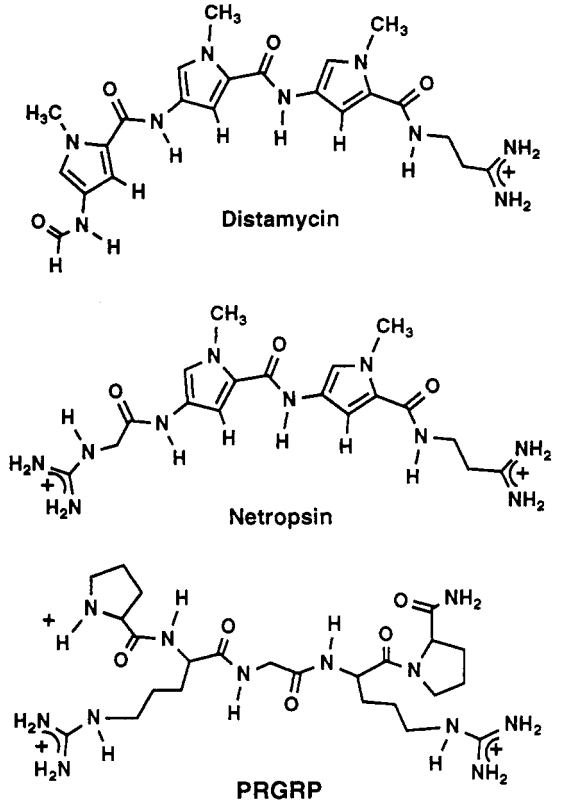

FIGURE 8: Comparison of the structures of distamycin, netropsin, and PRGRP.

SPRKSPRK with d(CGCAAAAAAGGC)·d(GCCTTTTTTGCG). Very little line broadening or shifting of resonance lines is observed. Addition of excess SPRKSPRK results in precipitation of the DNA oligomer, presumably due to screening of the negatively charged phosphate backbone by the +5 charge of the peptide. The high peptide to binding site ratios used for footprinting were therefore not accessible to NMR titrations. No NOE cross peaks between peptide and DNA protons were found in the 2D NOESY spectra of oligomer d(CGCAAATTGGC)·d(GCCAATTTGCG) or d(CGCAAAAAAGGC)· d(GCCTTTTTTGCG) combined with 2 molar equiv of SPRKSPRK. Suzuki, Gerstein, and Johnson (1993) mention that their collaborator (Dr. Molinali, Milano, Italy) recently made similar observations. In summary, we conclude that SPRKSPRK interacts nonspecifically with DNA at least under the conditions studied. In the context of a larger protein, the SPRKSPRK sequence could, however, contribute to A,T binding preference, if most of the affinity results from other parts of the protein.

Specific Binding of TPKRPRGRPKK and Derivatives to A,T Sites. Competition experiments with minor groove binding ligands such as Hoechst 33258 and distamycin led Reeves and Nissen (1990) to propose that the 11 amino acid fragment TPKRPRGRPKK derived from mammalian HMG-I/Y proteins binds to the minor groove of A,T-rich sites. NMR results of the present study provide direct physical evidence for minor groove binding: NOE contacts of peptide protons to the adenine C2 protons and the C1′ protons of the deoxyribose rings exposed in the minor groove unambiguously place the peptide in the minor groove of the A,T sequences of d(CGCAAAAAAGGC)·d(GCCTTTTTTGCG) and d(CGCAAATTTGCG)$_2$. NOE contacts of TPKRPRGRPKK to the A_6 site indicate binding in both directions. Observation of only one set of DNA lines at any stoichiometry shows that the peptide exchanges rapidly between bound and free forms. The line broadening observed suggests that the exchange rate of the peptide–DNA complex is close to the intermediate exchange regime with a residual lifetime for the complex of the order of milliseconds.

NOE data from the TPKRPRGRPKK complex with the symmetric A_3T_3 site indicate that only the PRGRP part of the peptide is in contact with the DNA minor groove at the central A,T sequence. The shorter peptides PRGRPKK and PRGRP show similar contacts. PKGKP did not show any NOE contacts to DNA. In addition, the peptide resonances from PKGKP shifted much less upon addition to DNA than did those of PRGRP. This comparison shows that the PRGRP sequence, and in particular the arginine residues, determines the sequence preference of the TPKRPRGRPKK peptide for A,T sites.

In the complexes of TPKRPRGRPKK, we were unable to identify any contacts between DNA and the TPKR sequence or the C-terminal lysine residues of the peptide. The intensity of the sequential intrapeptide NOEs of the amide protons to Cα and Cβ protons of the previous amino acid residue indicate that the peptide is in an extended conformation when bound to DNA. In addition, no evidence for the proposed conformation of the A,T-hook motif was observed (Reeves & Nissen, 1990). The NOESY data suggest that the PRGRP sequence "anchors" these peptides specifically into the minor groove.

All three PRGRP-type peptides exchange rapidly between the bound and free forms. The peptide with the highest binding affinity (lowest off rate) will spend more time in the bound form, and the chemical shifts of both peptide and DNA will reflect this fact. The positively charged side chains of the longer peptides were expected to increase the binding affinities of these peptides, but due to the uncertainty of the titration data (Figure 3), this could not be confirmed. However, changes of the imino resonance line of the guanine closest to the A_3T_3 site are only observed upon binding of PRGRPKK and TPKRPRGRPKK. This suggests that residues other then PRGRP are loosely associated with base pairs flanking the specific binding site. The titration data suggest a more complicated binding behavior at high peptide:DNA ratios than a simple one-peptide-per-oligonucleotide duplex, possibly the binding of a second peptide at an unoccupied site of lower affinity. The local interactions of the three PRGRP-type peptides with DNA may be slightly different as suggested by the different asymptotic chemical shift values reached at high peptide:DNA ratios in the titrations.

Restrained molecular modeling, using distance restraints derived from NMR data of the PRGRP complex with $d(CGCAAATTTGCG)_2$ and $d(CGCGAATTCGCG)_2$, positions the RGR part of the peptide deep in the minor groove (Figure 7). This part of the peptide follows the curvature of the minor groove. While the PRGRP peptide can bind to and/or slide between subsites in the A_3T_3 site, it is placed firmly in the center of the A_2T_2 site. The arginine side chains point away from one another, with the peptide backbone and the bridging glycine residue at the bottom of the minor groove next to the central AT step. An amino acid containing a side chain at the central position would have severe van der Waals conflicts with the DNA groove. Molecular modeling suggests specific hydrogen bonds between the guanidinium protons of the arginines and the bases and possibly also the phosphate oxygens. However, the details of such interactions are suspect in calculations done in the absence of explicit solvent, and the NOESY data do not provide direct information supporting such interactions. The importance of specific interactions of the arginine side chains with the bases and phosphate groups of DNA for the A,T specificity of the PRGRP-type peptides is supported by the observation that PKGKP does not interact in the same fashion with the same oligonucleotide duplex under identical conditions.

The geometry and arrangement of the cationic side chains of PRGRP resemble those of the antibiotic netropsin (Figure 8). Netropsin also binds to the minor groove of A,T sequences. A combination of electrostatic and van der Waals interactions, as well as specific hydrogen bonds between netropsin amides and DNA acceptor groups, stabilizes the netropsin–DNA complex. Recently, NMR studies (Rhee et al., 1993) suggested the presence of specific hydrogen bonds between the guanidinium and amidinium groups with the DNA bases also. The arginine guanidinium groups of PRGRP appear to interact similarly. The higher binding affinity of netropsin ($K_b \geq 10^7$ M^{-1}) (Rentzeperis & Marky, 1993) compared to PRGRP is presumably due to better van der Waals interaction of the pyrrole rings with the walls of the minor groove and the more rigid conformation of the netropsin backbone.

Conclusions and Implications for Protein Binding to DNA. The peptides TPKRPRGRPKK, PRGRPKK, and PRGRP derived from the non-histone chromosomal protein HMG-I/Y bind specifically to the minor groove of A,T sequences. In contrast to a previous proposal (Reeves & Nissen, 1990) and in accordance with an earlier suggestion (Lund et al., 1987), it is the PRGRP part of the peptides that interacts specifically with DNA. On the basis of NOESY data we propose a model in which the RGR part mimics netropsin. While electrostatic interactions may be important for binding affinity, the arginine side chains and their interactions determine the sequence specificity. HMG-I/Y was recently shown to function as a transcription factor (Thanos & Maniatis, 1992; Du et al., 1993) and contains three repeats of the PRGRP motif spaced by approximately 30 residues. It will be interesting to see whether in the context of the whole protein these PRGRP motives are used as binding modules similar to our model. Binding of many other A,T-specific proteins to DNA has been attributed mainly to electrostatic interactions between basic amino acid residues and the phosphate backbone of DNA. Characterization of the PRGRP-type peptides bound to the minor groove of DNA suggests that specific interactions between arginine side chains and the DNA bases as well as the phosphate backbone may be important for the preference of many of these proteins for A,T-rich DNA.

SUPPLEMENTARY MATERIAL AVAILABLE

Listings of the intermolecular peptide–DNA NOEs in all complexes, the sequential intrapeptide NOEs in all complexes, and the intermolecular peptide–DNA and intramolecular peptide restraints used to derive the molecular model of the PRGRP complex with $d(CGCGAATTGCGC)_2$ and achieved distances; tables of chemical shift assignments of peptide resonances and selected DNA resonances in all complexes and uncomplexed in aqueous solution (17 pages). Ordering information is given on any current masthead page.

REFERENCES

Churchill, M. E. A., & Suzuki, M. (1989) *EMBO J. 8*, 4189–4195.

Churchill, M. E. A., & Travers, A. A. (1991) *Trends Biochem. Sci. 16*, 92–97.

Clore, G. M., & Gronenborn, A. M. (1982) *J. Magn. Reson. 53*, 423–442.

Coll, M., Frederick, C. A., Wang, A. H.-J., & Rich, A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8385–8389.

Davis, D. G., & Bax, A. (1985) *J. Magn. Reson. 65*, 355–360.

Du, W., Thanos, D., & Maniatis, T. (1993) *Cell 74*, 887–898.

Dwyer, T. J., Geierstanger, B. H., Bathini, Y., Lown, J. W., & Wemmer, D. E. (1992) *J. Am. Chem. Soc. 114*, 5911–5919.

Griesinger, C., Otting, G., Wüthrich, K., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870–7872.

Hare, D. R., Wemmer, D. E., Chou, S.-H., Drobny, G., & Reid, B. R. (1983) *J. Mol. Biol. 171*, 319–336.

Hill, C. S., Packman, L. C., & Thomas, J. O. (1990) *EMBO J. 9*, 805–813.

Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., & Dickerson, R. E. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1376–1380.

Lund, T., Dahl, K. H., Mørk, E., Holtlund, J., & Laland, S. G. (1987) *Biochem. Biophys. Res. Commun. 148*, 725–730.

Pelton, J. G., & Wemmer, D. E. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 5723–5727.

Radic, M. Z., Saghbini, M., Elton, T. S., Reeves, R., & Hamkalo, B. A. (1992) *Chromosoma 101*, 602–608.

Reeves, R., & Nissen, M. S. (1990) *J. Biol. Chem. 265*, 8573–8582.

Rentzeperis, D., & Marky, L. A. (1993) *J. Am. Chem. Soc. 115*, 1645–1650.

Rhee, Y., Wang, Ch., Gaffney, B. L., & Jones, R. A. (1993) *J. Am. Chem. Soc. 115*, 8742–8746.

Sklenar, V., Brooks, B. R., Zon, G., & Bax, A. (1987) *FEBS Lett. 216*, 249–252.
Suzuki, M. (1989a) *EMBO J. 8*, 797–804.
Suzuki, M. (1989b) *J. Mol. Biol. 207*, 61–84.
Suzuki, M., Gerstein, M., & Johnson, T. (1993) *Protein Eng. 6*, 565–574.
Sykes, B. D. (1993) *Curr. Opin. Biotechnol. 4*, 392–396.
Tabernero, L., Verdaguer, N., Coll, M., Fita, I., van der Marel, G. A., van Boom, J. H., Rich, A., & Aymami, J. (1993) *Biochemistry 32*, 8403–8410.
Thanos, D., & Maniatis, T. (1992) *Cell 71*, 777–789.
Warshaw, M., & Cantor, C. (1970) *Biopolymers 9*, 1079–1103.
Wemmer, D. E., Geierstanger, B. H., Fagan, P. A., Dwyer, T. J., Jacobsen, J. P., Pelton, J. G., Ball, G. E., Leheny, A. R., Chang, W.-H., Bathini, Y., Lown, J. W., Rentzeperis, D., Marky, L., Singh, S., & Kollman, P. (1994) in *Proceedings of the Eighth Conversation on Biomolecular Stereodynamics* (Sarma, R. H., & Sarma, M. H., Eds.) Adenine Press, Guilderland, NY (in press).
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, J. Wiley & Sons, New York.
Zimmer, C., & Wähnert, U. (1986) *Prog. Biophys. Mol. Biol. 47*, 31–112.