Solution Structure and Stability of Tryptophan-Containing Nucleopeptide Duplexes

Dedicated to Professor M. Rico on the occasion of his 65th birthday.

Irene Gómez-Pinto,[a] Vicente Marchán,[b] Federico Gago,[c] Anna Grandas,*[b] and Carlos González*[a]

Covalently linked peptide–oligonucleotide hybrids were used as models for studying tryptophan–DNA interactions. The structure and stability of several hybrids in which peptides and oligonucleotides are linked through a phosphodiester bond between the hydroxy group of a homoserine (Hse) side chain and the 3′-end of the oligonucleotide, have been studied by both NMR and CD spectroscopy and by restrained molecular dynamics methods. The three-dimensional solution structure of the complex between Ac-Lys-Trp-Lys-Hse(p3′dGCATCG)-Ala-OH (p = phosphate, Ac = acetyl) and its complementary strand 5′dCGTAGC has been determined from a set of 276 experimental NOE distances and 33 dihedral angle constraints. The oligonucleotide structure is a well-defined duplex that belongs to the B-form family of DNA structures. The covalently linked peptide adopts a folded structure in which the tryptophan side chain stacks against the 3′-terminal guanine moiety, which forms a cap at the end of the duplex. This stacking interaction, which resembles other tryptophan–nucleobase interactions observed in some protein–DNA complexes, is not observed in the single-stranded form of Ac-Lys-Trp-Lys-Hse(p3′dGCATCG)-Ala-OH, where the peptide chain is completely disordered. A comparison with the pure DNA duplex, d(5′GCTACG3′)–(5′CGTAGC3′), indicates that the interaction between the peptide and the DNA contributes to the stability of the nucleopeptide duplex. The different contributions that stabilize this complex have been evaluated by studying other nucleopeptide compounds with related sequences.

KEYWORDS:

DNA structures · NMR spectroscopy · peptides · oligonucleotides · stacking interactions

Introduction

Molecular recognition between proteins and nucleic acids plays an essential role in living cells. The study of the basic interactions between these biomolecules is of great importance for understanding many biological processes, such as expression, replication, or repair of genetic information. The large number of structures of DNA–protein complexes now available reveal that different factors contribute to the recognition process. Hydrogen bonds, electrostatic interactions, and interactions that involve nonpolar groups are known to be involved in protein–DNA contacts. Aromatic amino acids are also engaged in stacking interactions with nucleic acid bases. This kind of interaction has been observed mainly in complexes formed by proteins and single-stranded DNA.[1–3] Although complete intercalation in canonical DNA duplexes is not a common mode of interaction between proteins and DNA, stacking interactions or partial intercalation of hydrophobic and aromatic residues are common in complexes where kinking of the DNA double helix occurs.[4–9]

The effect of a particular interaction in a large protein–DNA complex, such as stacking between an aromatic side chain and a nucleobase, can be obscured by many other contributions. For this reason, stacking interactions have been studied in short oligonucleotide fragments[10, 11] and other model compounds.[12] The existence of stacking between aromatic side chains and DNA nucleobases is indicated by evidence from a number of different techniques, such as changes in the fluorescence spectra of the aromatic residues, alterations in the circular dichroism spectra, etc.[13–18] Some models of stacking or intercalative binding between peptides containing aromatic residues and short DNA fragments have been proposed based on NMR data.[19–21] However, the structural information on peptide–DNA complexes that involve aromatic amino acid side chains is rather limited.

Among the different aromatic amino acids, tryptophan has attracted particular attention in spectroscopic studies because of its fluorescence quenching properties.[13, 22–24] In addition, its bicyclic indole ring seems to be a good candidate for inter-

[a] *Dr. C. González, I. Gómez-Pinto*
Instituto de Química Física "Rocasolano"
CSIC, C/. Serrano 119, 28006 Madrid (Spain)
Fax: (+34) 915-642-431
E-mail: cgonzalez@iqfr.csic.es

[b] *Prof. A. Grandas, Dr. V. Marchán*
Departament de Química Orgànica
Universitat de Barcelona, C/. Martí i Franquès 1
08028 Barcelona (Spain)
E-mail: agrandas@qo.ub.es

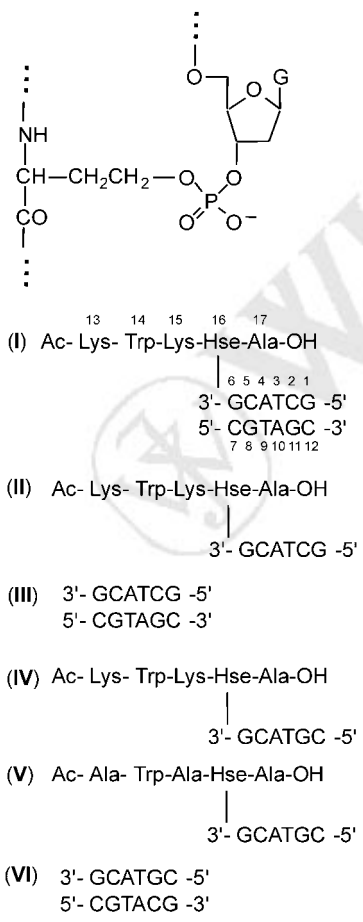
[c] *Prof. F. Gago*
Departamento de Farmacología
Universidad de Alcalá, 28871 Madrid (Spain)

  1439-4227/03/04/01 $ 20.00+.50/0

calation between DNA bases. Trp – DNA interactions may be better studied in model compounds where the peptide chain is covalently linked to the oligonucleotide. Use of these compounds greatly reduces the entropic cost of complexation. This strategy was first used by Ho and co-workers in their structural studies of an aminoacyl-capped oligonucleotide duplex.[25] In their work, the tryptophan residue was covalently attached to the 5′-end of the oligonucleotide TGCGCAC by a peptide bond. The tryptophan side chain did not intercalate between the bases, but stacked against the terminal A – T base pair to form a "cap" at the end of the double helix. However, the features observed in this structure might be affected by the lack of a flexible linker between tryptophan residue and the oligonucleotide chain.

Nucleopeptides are peptide – oligonucleotide conjugates with a covalent phosphodiester bond between the hydroxy group of an amino acid side chain and the 3′- or 5′-terminal hydroxy group of an oligonucleotide chain (Scheme 1). This kind of bond occurs in some covalently linked protein – DNA complexes, such as those induced by topoisomerases[26] or the initiation complex in the replication of some viruses,[27] thus nucleopeptides might be used as models for the design of new anticancer or antiviral drugs. In addition, these or other types of peptide – oligonucleotide conjugates may be useful as nonradioactive labels, to improve the cellular uptake of antisense oligonucleotides,[28, 29] or to investigate the molecular requirements for enzyme activity.[30] Moreover, the fact that the peptide chain is linked to the 3′-end of the oligonucleotide chain makes these molecules more resistant to exonucleases.[31] Several potential applications of peptide – oligonucleotide conjugates have been reviewed recently.[32]

In this paper, we report the structure of the complex formed by a tryptophan-containing nucleopeptide with its complementary oligonucleotide chain (**I**; see Scheme 1). The particular sequence studied in this paper was chosen according to the following criteria: 1) the peptide is long enough for the tryptophan to intercalate at different base steps (GC or AT); 2) the tryptophan is flanked by two basic lysine residues, which should favor its approach to the polyanionic oligonucleotide chain; 3) homoserine was chosen as the linking amino acid to prevent base-promoted degradation in the final steps of the synthesis.[33] The thermal stabilities of **I** and of the duplex forms of the related nucleopeptides **IV** and **V** with self-complementary oligonucleotide sequences have also been studied. Finally, we present the NMR analysis of the single-stranded form of the nucleopeptide Ac-Lys-Trp-Lys-Hse(p3′dGCATCG)-Ala-OH (**II**), and of the two pure DNA duplexes **III** and **VI**, which have been included in this study as a control.



Scheme 1. *The nucleopeptide linking site and the different compounds studied in this paper. Hse = homoserine, Ac = acetyl.*

Results

Preliminary model-building studies

The peptide sequence in compound **I** has been chosen to allow the exploration of different modes of interaction between tryptophan and the oligonucleotide. Preliminary model-building studies were carried out with the program SYBYL to verify this choice. Different models of the nucleopeptide in which the indole ring of tryptophan intercalates in different positions of the DNA duplex were constructed. As can be seen in Figure S1 of the Supporting Information, the length of the linker between tryptophan and the 3′-end of the oligonucleotide is sufficient to allow for intercalation into the center of the DNA duplex (into the 5′-TA-3′ step). Intercalation at the 5′-AC-3′ or 5′-CG-3′ steps is clearly also possible. The linker region (which includes a segment of the main peptide chain), the homoserine side chain, and the connecting phosphate group, allow the tryptophan ring to adopt a variety of possible orientations, either as a "cap" at the end of the duplex, or inserted into any of the grooves.

NMR assignment

Non-exchangeable protons in both oligonucleotide strands were assigned by using established techniques for right-handed, double-stranded nucleic acids and DQF-COSY, TOCSY, and 2D NOESY spectra.[34, 35] Sugar spin systems were identified in the DQF-COSY and TOCSY spectra and connected with their own base and their 5′-neighbor in the NOESY spectra. These assignment pathways could be followed in the base – H1′ (see Figure 1) and in the base – H2′/H2″ regions for both strands. All exchangeable DNA protons could be assigned with the NOESY spectra

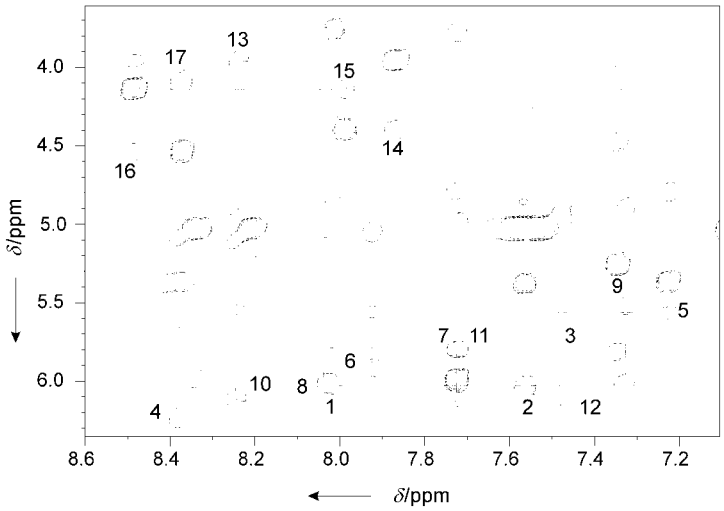


Figure 1. A region of the NOESY spectruma of [Ac-Lys-Trp-Lys-Hse(p3′dGCATCG)-Ala-OH] – [p5′dCGTAGC] in H_2O (100 mM NaCl, T = 5 °C, pH = 7, τ_m = 200 ms). Sequential assignment pathways are shown for the peptide and both oligonucleotide strands. Intraresidual Hα – HN and H1′ – H6/H8 cross-peaks are labeled with the residue number (see Scheme 1 for numbering).

recorded in H_2O, except for the amino resonances of guanine bases. The cross-peak patterns observed for the exchangeable protons indicate that all bases form Watson – Crick pairs throughout the duplex (see Figure S2 in the Supporting Information).

Spin systems in the peptide chain were identified in TOCSY and COSY spectra and connected by the sequential Hα →HN NOEs observed in H_2O. The assignment pathway for the peptide chain is shown in Figure 1. Almost all peptide resonances were assigned (see Table S1 in the Supporting Information). Stereospecific assignments for the Hβ and Hγ protons of the homoserine residue were based on the relative intensities of the intraresidual NOE cross-peaks at low mixing times and the pattern of *J*-coupling constants. No other stereospecific assignments of the peptide resonances could be obtained.

Chemical shift comparison with the control duplex

An almost complete resonance assignment was carried out for the control duplex 5′d(GCTACG) – 5′d(CGTAGC) (**III**) by following the same methodology as for the nucleopeptide complex. An assignment list of the proton resonances is given in the Supporting Information. The nucleotides affected by the peptide chain in the nucleopeptide can be identified by comparing their proton chemical shifts with those of the control duplex. The proton with the largest shift difference is the imino proton of G6, which is shifted by 0.4 ppm. The amino resonances of C7 and C5 are also shifted by 0.13 and 0.06 ppm, respectively (for the hydrogen-bonded protons). The non-exchangeable protons are also affected, mainly in residues 5 and 6, as can be seen in Figure 2. These chemical shift differences indicate that the peptide chain only affects one end of the DNA duplex. The chemical shift changes are not consistent with intercalation of the tryptophan side chain into the center of the duplex.

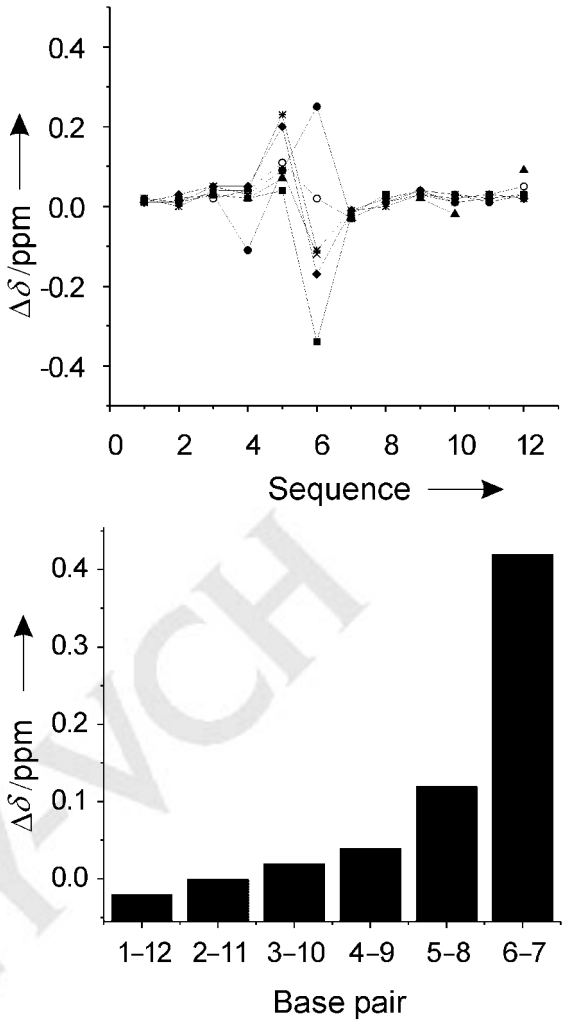


*Figure 2. Chemical shift differences between [Ac-Lys-Trp-Lys-Hse(p3′dGCATCG)-Ala] – [p5′dCGTAGC] and the control duplex. Top: non-exchangeable protons (○, H6/H8; ▲, H5/H2/Met; ●, H1′; *, H2′; ◆, H2″; ■, H3′; x, H4′). Bottom: imino protons.*

NMR spectra of the single-stranded form of Ac-Lys-Trp-Lys-Hse(p3′dGCATCG)-Ala-OH

The NMR spectra of the single-stranded Ac-Lys-Trp-Lys-Hse(p3′dGCATCG)-Ala-OH (**II**) were also analyzed. Sequential assignment of the peptide was carried out following the same procedure as for **I**. However, a complete resonance assignment of the oligonucleotide moiety could not be carried out in this case because the exchangeable protons of the oligonucleotide moiety were not observed and many sequential connections between spin systems of the non-exchangeable protons were missing. In spite of this, some signals from the non-exchangeable protons could be identified (see Figure S3 and Table S2 in the Supporting Information). Interestingly, no NOE cross-peaks were observed between the tryptophan residue and the oligonucleotide, which indicates that the stacking interaction between the tryptophan residue and guanine 6 only occurs when the complementary strand is present and the duplex is formed.

Melting transitions

The melting behavior of [Ac-Lys-Trp-Lys-Hse(p3′dGCATCG)-Ala-OH] – [p5′dCGTAGC] and of the equivalent duplex formed by pure oligonucleotide strands was monitored by NMR spectroscopy. Melting temperatures for both molecules were estimated by following the changes in the methyl resonances of the thymine residues (see Figure 3 and Table S3). Under the experimental conditions used in the NMR experiments, the nucleopeptide duplex melts at 46 °C, while the pure oligonucleotide duplex melts at 40 °C. This enhanced stability of the nucleopeptide is not unexpected since the peptide chain contains two lysine residues and the number of cationic residues has been shown to be one of the main factors in the stabilization of peptide – oligonucleotide conjugate duplexes.[36]

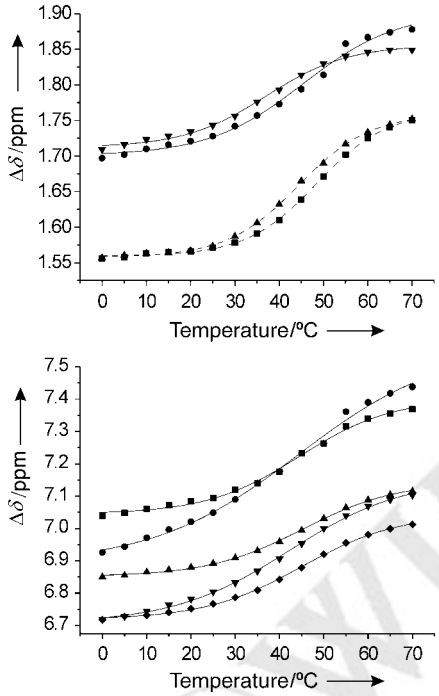


Figure 3. Chemical shift variation versus temperature. Top: the methyl resonances of the thymine residues of [Ac-Lys-Trp-Lys-Hse(p3′dGCATCG)-Ala-OH] – [p5′dCGTAGC] (●, T3; ■, T9), and the control duplex (▼, T3; ▲, T9). Bottom: the aromatic protons of the tryptophan residue (●, HE3; ■, HZ2; ▲, HH2; ▼, HD1; ◆, HZ3).

The melting behavior of the peptide moiety itself is more interesting. Melting curves monitored by NMR spectroscopy have the advantage that the melting behavior of different parts of the molecule can be followed independently. The profile of chemical shift variation for the aromatic protons of the tryptophan residue is very similar to that of the methyl protons in the nucleopeptide duplex (Figure 3), which indicates that the transition from the folded peptide to a random coil structure occurs at approximately the same temperature as duplex dissociation.

To gain an insight into the different contributions to the stability of **I**, the melting behavior of the related compounds [Ac-Lys-Trp-Lys-Hse(p3′dGCATGC)-Ala-OH]$_2$ (**IV**) and [Ac-Ala-Trp-Ala-Hse(p3′dGCATGC)-Ala-OH]$_2$ (**V**) has been studied by both NMR and CD spectroscopy. In both cases, the oligonucleotide sequence is self-complementary, which facilitates the interpretation of the melting curves. In the case of **IV**, the peptide sequence is identical to that of **I**, but in **V** the lysine residues have been replaced by alanine residues to reduce the electrostatic interactions between the peptide and the DNA. Analysis of the NMR spectra of the duplex forms of these nucleopeptides indicates that in all cases the tryptophan residue adopts a similar conformation, stacked against the terminal GC base pair. NMR and CD melting curves were obtained at five different nucleopeptide concentrations and thermodynamic parameters were obtained from the variation of the melting temperature with the concentration[37] (see Figure S4 in the Supporting Information). This analysis was also carried out for the reference duplex d(GCATGC)$_2$. The results are shown in Table 1. As in the case of **I**, the presence of the peptide stabilizes the duplex. The difference in ΔG values at 25 °C between the nucleopeptide with cationic residues **IV** and the control duplex is 4 kJ mol^{-1}, while this difference is only 2 kJ mol^{-1} for nucleopeptide **V**.

Relaxation time measurements

The T_1 and T_2 relaxation times for the aromatic protons of the tryptophan residue and the base protons of the oligonucleotide are listed in Table S4 in the Supporting Information. The difference between the relaxation times of the tryptophan protons and those of the protons in the DNA is clearly within the dispersion of T_1 and T_2 values of the DNA protons. Therefore, there is no reason to suppose that isotropic tumbling is not a correct approximation for this molecule, as it is for double-stranded DNA duplexes. Consequently, a single correlation time can be estimated from T_1 and T_2 by following the procedure of Suzuki et al.[38] The average τ_c value obtained by this method is 1.9 ± 0.3 ns. In the subsequent relaxation matrix calculations, global correlation times of 1.5, 2.0, and 2.5 ns were used.

Table 1. *Thermodynamic parameters for duplex-to-random-coil transitions.*[a]

compound	ΔH^0 [kJ mol^{-1}]	ΔS^0 [kJ mol^{-1} K]	ΔG^0_{298} [kJ mol^{-1}]	T_m [°C][b]
[Ac-Lys-Trp-Lys-Hse(p3′dGCATGC)-Ala-OH]$_2$	−193 ± 2	−551 ± 7	−29 ± 2	34
[Ac-Ala-Trp-Ala-Hse(p3′dGCATGC)-Ala-OH]$_2$	−207 ± 2	−607 ± 13	−27 ± 1	31
d(GCATGC)$_2$	−185 ± 2	−535 ± 10	−25 ± 1	28

[a] Measurements were taken at 100 mM NaCl, 25 mM phosphate, pH = 7. [b] Melting temperatures at a 100-mM strand concentration.

Distance constraints

The total number of meaningful experimental distance constraints was 276. There are 238 interproton distances in the DNA moiety, which gives an average number per base pair of 39, with both exchangeable and non-exchangeable protons included. A summary of the experimental distance constraints used is shown in Table 2. Eighteen NOEs were observed between the peptide and the terminal residues of the oligonucleotide chain. These peptide–DNA distance constraints are indicated schematically in Figure 4 and listed in Table S5.

Table 2. *Experimental constraints and calculation statistics.*

Experimental distance constraints		
intra-residue	118	
sequential	82	
inter-strand	38	
peptide	20	
peptide-DNA	18	
total number DNA	238	
total number	276	
RMSD [Å]		
DNA (all heavy atoms)	0.9 ± 0.2	
DNA backbone	1.0 ± 0.2	
DNA bases	$0.5 \pm 0.$	
peptide (all heavy atoms)	2.1 ± 0.6	
peptide (well-defined region)[a]	0.7 ± 0.2	
Residual violations	Average	Range
sum of violation [Å]	23.7	22.6–26.5
max. violation [Å]	0.5	0.3–0.6
NOE energy [Kcal mol^{-1}]	149	135–160

[a] Heavy atoms for residues 14 and 16 and backbone atoms of residue 15

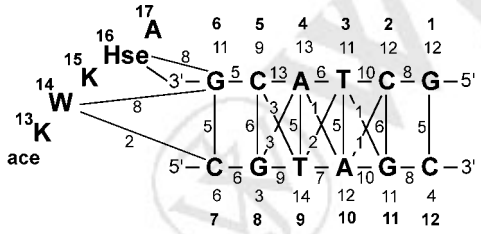


Figure 4. *Distribution of distance constraints. Bold numbers indicate the sequence order. Other numbers indicate the number of intraresidue, sequential, and interstrand distance constraints.*

With the exception of Hse16, no intraresidual distance constraints were considered in the peptide chain since these NOE intensities could be contaminated by a minor population of random-coil structures. In the case of the side-chain protons of homoserine, the pattern of intraresidual NOEs observed at short mixing times and all the $J_{\alpha\beta}$ and $J_{\beta\gamma}$ coupling constants indicate that this side chain adopts a well-defined conformation ($\chi_1 \approx -60^\circ$, $\chi_2 \approx -60^\circ$). In this particular case, intraresidual constraints were included in the calculation.

J-coupling analysis

J-coupling constants for deoxyriboses were estimated from DQF-COSY experiments. In most cases, the *J*-couplings are consistent with sugar puckers predominantly in the *S* domain, with the exception of the couplings of C12. In this residue, the H2′ and H2″ protons are degenerated, and *J*-coupling analysis could not be carried out. For all other residues, dihedral angle constraints were included in the structure calculation to keep the deoxyriboses in an *S*-type conformation.

Structure calculation

Distance and torsion angle constraints were used to calculate the structure of this molecule by restrained molecular dynamics. Structures were calculated with the program DYANA 1.5[39] and further refined with the SANDER module of the molecular dynamics package AMBER 5.1.[40] Two different calculations were carried out with the AMBER force field. In the first refinement, calculations were carried out in vacuo according to a high-temperature annealing protocol. At the end of this calculation, all structures converged to a well-defined region of the conformational space. The average value of the mutual root mean square deviations (RMSDs) between the 10 final structures is around 1 Å, which indicates that reasonable convergence has been reached after the annealing procedure.

It has recently been shown that the inclusion of explicit solvent and a better evaluation of the electrostatic term improves the quality of the final NMR-derived structures.[41] Consequently, we have refined every final structure from in vacuo calculations by means of 220-ps NMR-restrained molecular dynamics trajectories that include explicit solvent, and by the use of the Particle Mesh Ewald (PME) method. In the resulting structures, final distance restraint violations, and E_{NOE} terms are small (see Table 2), which shows the ability of these structures to fulfill the experimental constraints. The 10 structures that resulted from the final refinement in water are displayed in Figure 5. As can be seen in this figure (and also in Table 2), the ensemble of structures is well defined. As is usual in NMR-derived structures, the bases are better defined than the sugar–phosphate backbone. Both modified and unmodified strands show similar RMSD values.

Description of the structures

The conformation of the DNA in duplex **I** belongs to the B family of DNA structures. All glycosidic angles are *anti* and the pseudorotation phase angles of the deoxyribose moeities are generally in the *S* domain, with the exception of C12, which has an average pseudorotation phase angle of around 90° (see Figure 6). This value probably reflects an enhanced flexibility of the 3′-terminal residue and not a real *E*-type sugar conformation. Such values are usually observed in dynamic regions of oligonucleotide fragments, where deoxyribose units undergo a rapid exchange equilibrium between *S*- and *N*-type conformations with similar populations. Interestingly, this effect is not observed at the other 3′-end, where the peptide is attached,

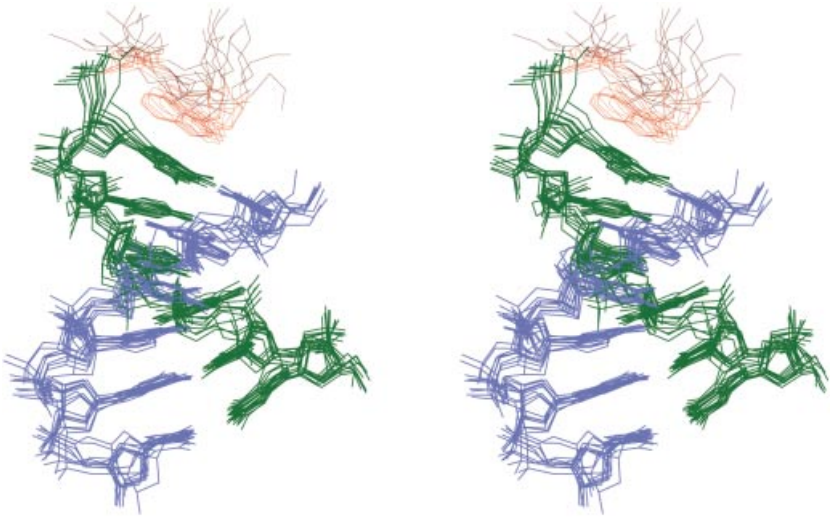


Figure 5. Stereoscopic view of the superposition of the 10 refined structures of [Ac-Lys-Trp-Lys-Hse(p3′dGCATCG)-Ala-OH] – [p5′dCGTAGC]. Each of the oligonucleotide strands is shown in a different color. The peptide moiety is shown in red. Only the backbone (indicated in a darker color) and the well-defined side chains are shown.

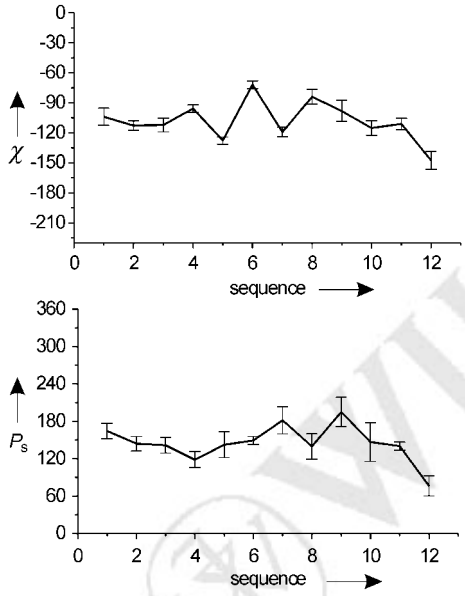


Figure 6. Glycosidic (χ) and pseudorotation phase (P_s) angles versus sequence position in the structure of [Ac-Lys-Trp-Lys-Hse(p3′dGCATCG)-Ala-OH] – [p5′dCGTAGC].

which indicates that the peptide reduces the flexibility in the terminal region of the oligonucleotide chain.

The backbone of the peptide chain adopts a well-defined structure between residues Trp14 and Hse16, whilst the rest of the chain is completely disordered. The RMSD for all peptide heavy atoms is 2.1 Å, while that for the well-defined part of the chain (including side chains) is only 0.7 Å. Among the side chains, only Trp14 and Hse16 present well-defined conformations, with χ_1 and χ_2 values of $32 \pm 22^\circ$ and $-120 \pm 20^\circ$ for tryptophan, and $-78 \pm 6^\circ$ and $-81 \pm 8^\circ$ for homoserine, respectively. Interestingly, most of the torsion angles in the linker segment between O3′ of G6 and the tryptophan ring present well-defined conformations. This peptide segment fold approaches the G – C pair from its major groove side and places the tryptophan residue on top of the guanine base (see Figure 7). The two aromatic groups are almost parallel to each other with very little tilt, which maximizes the stacking interaction between them.

Discussion

Structure

The structural analysis of the duplex formed by nucleopeptide Ac-Lys-Trp-Lys-Hse(p3′dGCATCG)-Ala-OH and its complementary oligonucleotide strand shows that the covalently linked peptide does not give rise to large changes in the structure of the DNA. As expected, the most significant distortions in the structure of the oligonucleotide with respect to a B-type duplex occur in the neighborhood of the peptide moiety, where the glycosidic angle of G6 has an unusual value of around -80°. Also, the rise value of 2.8 Å, observed between the two terminal base pairs (C5G8 – G6C7), is more common in A-type structures. Some distortion of the canonical B-like double helix is also observed at the T3A10 – A4T9 step. The large roll and tilt between these two base pairs produces a slight bending of the helix towards the major groove. This curvature was also apparent when the distance restraints

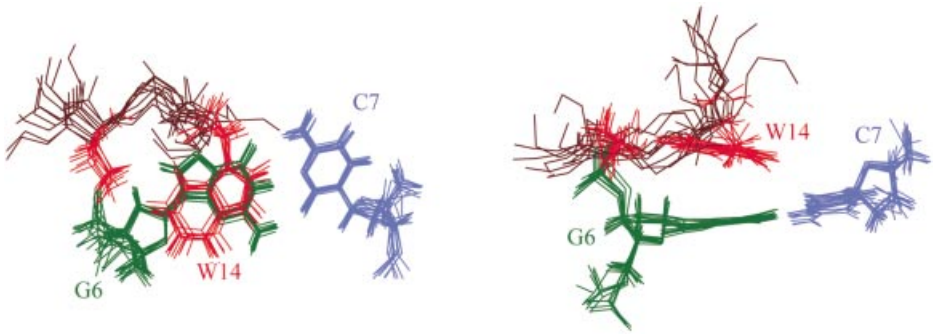


Figure 7. Detail of the interaction between the terminal G – C base pair and the peptide moiety (the same color code is used as in Figure 5).

were removed (data not shown) and could be related to the fact that both the TpA and CpG steps have a strong tendency to roll, which causes bending toward the major groove.[42] DNA bending is difficult to determine with precision in NMR-derived structures since the experimental information obtained from NMR methods (NOEs and *J*-couplings) is mainly local and consequently, global geometric parameters such as bending are poorly defined. It is tempting to correlate the orientation of Lys15 with the slight bending observed in the DNA, although the ensemble of structures presented here may not be sufficiently defined to determine the global axis with enough precision to confirm this idea. No specific contacts are observed between the lysine residue and the oligonucleotide but in seven out of ten structures the lysine side chain points towards the major groove of the DNA helix. The most negative region of the molecular electrostatic potential (MEP) of a G – C pair is found in the region of the N7 and O6 atoms of guanine in the major groove.[42] This negative MEP can attract the positively charged side chain of Lys15 to O6 of G6. The approach of this cationic residue towards the major groove might additionally be responsible for the observed bending, in line with the idea that electrostatic effects are the main driving force for DNA bending.[43]

The most prominent feature of the three-dimensional structure of **I** is the stacking interaction between the tryptophan side chain and the terminal G – C base pair of the duplex. "Capping" interactions with the terminal base pairs of a DNA duplex have also been observed in the complex between 4′,6-diamidino-2-phenylindole (DAPI) and the DNA hexamer $(CGATCG)_2$, where the indole derivative stacks on the terminal G – C base pair,[42] and in the structure of the aminoacyl – DNA duplex $(W\text{-}dTGCGCAC)_2$, obtained by Ho and co-workers.[25] A comparison of Ho's structure with that of **I** indicates that the capping interaction occurs when tryptophan is linked to either the 3′- or the 5′-end of the oligonucleotide chain. However, the relative position of the tryptophan side chain with respect to the terminal base pair is different in the two structures. In the case of $(W\text{-}dTGCGCAC)_2$, the indole group lies on the centre of the terminal A – T base pair, but in our structure the tryptophan is located over the adjacent guanine base (see Figure 8), which maximizes the overlap with the purine base. This different orientation may be the result of the shorter and more rigid linker between the tryptophan residue and the 5′-terminal phosphate group in $(W\text{-}dTGCGCAC)_2$ but may also reflect an intrinsic preference of tryptophan for interaction with guanine bases. Interestingly, a similar orientation has been found in crystallographic studies of complexes formed between tryptophan and 7-methylguanines.[44]

It could be argued that the orientation adopted by the tryptophan ring on top of the terminal G – C base pair is partially induced by the presence of lysine residues or an intrinsic folding tendency of the peptide chain. However, the peptide moiety of **I** only adopts a folded structure in the complex with its complementary DNA strand. The same behavior was observed in nucleopeptides **IV** and **V**. We conclude that the folding of the peptide is induced in all cases by interactions between the aromatic groups of Trp14 and G6.

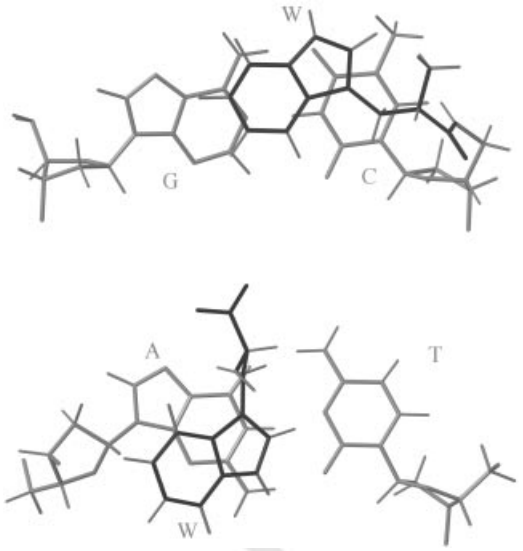


Figure 8. *Comparison of the tryptophan interaction with the terminal base pair in the structure of the aminoacyl-capped duplex $(W\text{-}TGCGCAC)_2$[25] (top) and in the nucleopeptide [Ac-Lys-Trp-Lys-Hse(p3′dGCATCG)-Ala-OH] – [p5′dCGTAGC] (bottom). The tryptophan residue is shown in black and the nucleobases in gray.*

Thermodynamics

The stability of the duplexes **I** and **IV** is clearly enhanced by the presence of two positively charged lysine residues. The two contributions to the interaction (tryptophan stacking and lysine electrostatic interactions) can be estimated separately from the thermodynamic data shown in Table 1. The difference between the ΔG^0_{298} values for nucleopeptide **V** and those of the control duplex is 2 kJ mol^{-1}, which indicates that each tryptophan – guanine interaction stabilizes the duplex by about 1 kJ mol^{-1}. The origin of this stabilization is entirely enthalpic ($\Delta\Delta H =$ 11 kJ mol^{-1}). The $\Delta\Delta S$ value between **V** and the control duplex is not favorable because of the entropic cost of folding the peptide chain. On the other hand, by comparing the thermodynamic data of **IV** and **V** and assuming that the two stabilizing effects are independent, it can be concluded that the presence of two lysine residues in the peptide chain stabilizes the duplex by 2 kJ mol^{-1} (1 kJ mol^{-1} per peptide chain). This increased stability with respect to nucleopeptide **V** is entropic in nature. In this case, the entropic cost of folding the peptide chains is partially compensated by the release of counterions when the lysine side chains approach the DNA. This effect has been observed in other complexes of oligonucleotides with lysine-containing peptides.[22, 45]

The stabilization induced by the tryptophan capping interaction in our nucleopeptide compounds is not very high but is larger than that observed in the aminoacyl – DNA duplex $(W\text{-}dTGCGCAC)_2$.[25] Again, this could either reflect a preference of the tryptophan to interact with G – C rather than with A-T base pairs or it may be the result of a more favorable orientation of the indole ring with respect to the terminal base pair in our compounds. The larger conformational freedom of the peptide linker allows the tryptophan residue to adopt a conformation that optimizes the interaction of the indole ring with the

oligonucleotide. Examination of the relative orientation of the dipole moments of both the guanine moiety and the indole rings (data not shown), which has been suggested to be useful to visually assess electrostatic complementarity in stacking interactions,[46] reveals a suboptimal alignment that is probably dictated by the hydrophobic effect, an effect that tends to bury the largest ring surface area away from the solvent. A small stabilizing contribution, however, can be expected from the fact that the indole NH group is directed towards the most negative region of the pairing cytosine residue. Overall, our results are consistent with other studies, which show that the stability of a DNA duplex is increased by linking large hydrophobic or aromatic groups at the end of the helix.[47, 48] This effect will be very useful in the design of more stable peptide – oligonucleotide conjugates.

Biological implications

There are some interesting similarities between the stacking interaction observed in our structure and other tryptophan – guanine interactions found in two complexes of viral nucleocapside proteins with single-stranded DNA.[49, 50] In both viral complexes a tryptophan side chain stacks on top of a guanine base. The relative orientation between these two residues is very similar to that found between Trp14 and G6 in **I**. Also, the same interaction is found in the complex of an HIV-1 nucleocapsid protein and an RNA loop.[51]

Another interesting feature of our nucleopeptide structure is that tryptophan approaches the G – C base pair on the major groove side of the duplex. In most of the DNA – protein complexes in which partial intercalation takes place the amino acid residue inserts into the DNA from the minor groove side.[52] A different mode of interaction has been observed recently in the structure of the complex of the very short repair endonuclease (Vsr) with a DNA duplex containing a TG mismatch.[53] In this complex, several aromatic residues, which include two tryptophan residues, approach the DNA from the major groove side and disrupt the base stacking on the side of the mismatch, which produces a large bend in the double helix. The two tryptophan residues, which are conserved in the Vsr family, stack against the TG base pair in a way that largely resembles the capping interaction observed in our nucleopeptide model.

Although the nucleopeptide models studied in this paper have been designed to favor intercalation, the indole ring of the tryptophan residue prefers to stack at the end of the duplex. This result may reflect an intrinsic preference of tryptophan, since in many protein – DNA complexes in which so-called "partial intercalation" takes place, the aromatic residues interact with DNA in regions where base stacking is disrupted and nucleobases would otherwise be exposed. This result may cast some doubts on many studies on peptide – oligonucleotide complexes in which intercalation was only supported by fluorescence quenching or changes in NMR chemical shifts. Instead, these effects may be the result of interactions with terminal base pairs or exposed bases in loop regions.

Methods

Synthesis and purification of nucleopeptides and oligonucleotides: Nucleopeptides were assembled by following previously described stepwise solid-phase procedures.[54, 55] Boc-Hse(DMT)-O$^-$ HEt_3N^+ , Boc-Lys(Tfa)-OH, and Boc-Trp(For)-OH (Boc = *tert*-butoxycarbonyl; DMT = 4,4′-dimethoxytrityl; For = formyl; Tfa = trifluoroacetyl) were used to assemble the peptide moiety, and the standard commercially available (2-cyanoethyl-*N*,*N*-diisopropyl)-phosphoramidites were used for the elongation of the oligonucleotide. Crude nucleopeptides were obtained after overnight treatment with concentrated aqueous ammonia/dioxane (1:1) solution at 55 °C and purified by reversed-phase medium-pressure column chromatography (gradient: 0 – 30 % B, where A = 0.05 M ammonium acetate and B = acetonitrile/H_2O 1:1). The nucleoside composition of nucleopeptide **II** after enzymatic digestion with snake venom phosphodiesterase and alkaline phosphatase was: dC 2.10, dG 1.20, dT 1.17, dA 1.07. The following values were obtained from the MALDI-TOF MS characterization (negative mode, 2,4,6-trihydroxyacetophenone) of the different nucleopeptides: Compound **II**, *m/z*: found: 2526.3 $[M-H]^-$; average calcd: 2529.0; Compound **IV**: *m/z*: found: 2526.9; Compound **V**: found: 2413.1 $[M-H]^-$; average calcd: 2414.1.

5′dCGTAGC3′ and 5′dGCTAGC3′ were assembled on controlled pore glass on a 10-μmol scale according to standard procedures and were deprotected by reaction with concentrated aqueous ammonia at 55 °C. Both molecules were purified as described above. MALDI-TOF MS (negative mode, 2,4,6-trihydroxyacetophenone): 5′dCGTAGC3′: *m/z*: 1792.5, $[M-H]^-$; 5′dGCTACG3′: *m/z*: 1791.7 $[M-H]^-$; average calcd: 1792.2.

NMR spectroscopy: All NMR samples were suspended in either D_2O or H_2O/D_2O (9:1, 500 μL). The resulting solution was buffered (10 mM phosphate, 100 mM NaCl, pH = 7.0) and had a duplex concentration of about 2 mM. NMR spectra were acquired in a Bruker AMX spectrometer operating at 600 MHz and processed with the UXNMR software. DQF-COSY, TOCSY,[56] and NOESY[57] experiments were recorded in D_2O and in H_2O/D_2O (9:1). All 2D experiments were carried out at 5 °C. The NOESY spectra were acquired with mixing times of 100 and 250 ms in D_2O, and 200 ms in H_2O. TOCSY spectra were recorded with an 80-ms mixing time. A jump-and-return pulse sequence[58] was employed in 1D H_2O experiments to observe the rapidly exchanging protons. Water suppression was achieved in 2D experiments in H_2O by the inclusion of a WATERGATE[59] module in the pulse sequence prior to acquisition. The spectral analysis program XEASY[60] was used for the assignment of the NOESY cross-peaks.

Longitudinal relaxation (T_1) experiments were carried out by using the inversion recovery method with a 180° composite pulse.[61] The intensities of the isolated peaks were fitted to the expression $S(t) = A + Be^{-t/T_1}$. Spin – spin relaxation times (T_2) were estimated by using the spin-echo method. The data were fitted to a single exponential function: $S(t) = Ae^{-t/T_2}$.

Melting experiments: Circular dichroism spectra were collected on a Jasco J-720 spectropolarimeter fitted with a thermostated cell holder and interfaced with a Neslab RP-100 water bath. CD melting experiments were recorded at 280 nm with a heating rate of 20 °C h^{-1} from 5 °C to 90 °C. NMR melting curves were carried out by following the changes in chemical shifts of well-resolved resonances in 1D spectra recorded at different temperatures (5 – 80 °C). Both NMR and CD melting curves were fitted with Microcal Origin 5.0 software by assuming a two-state equilibrium. Thermodynamic parameters were obtained from the slope of the plots of $1/T_m$ versus ln*C* by assuming

that the melting reactions take place with no appreciable net heat capacity changes.[37]

Experimental constraints: Quantitative distance constraints were obtained from NOESY experiments by using a complete relaxation matrix analysis with the program MARDIGRAS.[62, 63] A global isotropic correlation time was considered in the MARDIGRAS calculation, except for methyl groups, for which a three-state jump model was used. To estimate the error bounds in the interprotonic distances properly, different MARDIGRAS calculations were carried out with two different initial models (standard A- and B-form duplexes), two mixing times (100 and 250 ms), and three correlation times (1.5, 2.0, and 2.5 ns). Final distances and their errors were determined by averaging upper and lower bounds in all individual runs. A loose distance limit of 5 Å was set for those weak NOEs where no quantitative analysis could be carried out.

Sums of *J*-coupling constants involving H1′, H2′, and H2′′ protons were estimated from DQF-COSY cross-peaks. The population of major *S* conformers were calculated from the sums of $J_{1'2'}$ and $J_{1'2''}$ coupling constants, according to the expression derived by Rinkel and Altona.[64] In those deoxyribose units where the population of *S*-type conformer is larger than 75%, torsion angle constraints were included. Since only the sums of coupling constants were estimated, loose values were set for these dihedral angles.

In addition to these experimentally derived constraints, Watson–Crick hydrogen bond restraints were used. Target values for distances and angles related to hydrogen bonds were set as described in crystallographic data.[65] No backbone angle constraints were employed. Distance constraints with their corresponding error bounds were incorporated into the AMBER potential energy by defining a flat-well potential term.

Structure determination: Structures were calculated with the program DYANA 1.5[39] and further refined with the SANDER module of the molecular dynamics package AMBER 5.1.[40] The eight best DYANA structures were taken as starting points for the AMBER refinement. Additional starting models of the nucleopeptide duplex were built based on the A and B canonical structures[66] by using the SYBYL graphics program (Molecular Simulations, 1998). A suitable number of hydrated sodium ions were introduced to achieve electrical neutrality. These ten structures were first minimized and then submitted to a high-temperature restrained molecular dynamics calculation that followed standard annealing protocols used in our group,[67, 68] with minor variations.

The 10 structures obtained after the annealing process were further refined by using MD simulations that included explicit water molecules and eight neutralizing ions. Periodic boundary conditions and the PME method[69] were used to account for long-range electrostatic effects. This lengthy process allowed us to relax a few local distortions in the structure of the DNA that appeared in the structures derived from the simulated annealing procedure. The structures were first minimized by using standard equilibration protocols and were then subjected to 220 ps of restrained molecular dynamics simulations at constant pressure (1 atm) and temperature (300 K). The AMBER-95 force field[70] was used for the standard DNA and amino acid residues, and water molecules were represented by the TIP3P model. The RESP strategy[71] was used to obtain electrostatic charges for atoms involved in the Hse-3′pG link by using quantum mechanically derived HF/6-31G(d) molecular electrostatic potentials.

Coordinates: Atomic coordinates have been deposited with the Protein Data Bank (accession number 1J9N).

Supporting information available: Figure S1: a model of compound **I** with the tryptophan residue intercalated between the central AT base pairs. Figure S2: imino region of the NOESY spectrum of nucleopeptide **I** in H_2O. Figure S3: fragment of the NOESY spectra of the single-stranded form of Ac-Lys-Trp-Lys-Hse(p3′dGCATCG)-Ala in H_2O. Figure S4: CD melting curves of the duplex form of Ac-Ala-Trp-Ala-Hse(p3′dGCATGC)-Ala at different concentrations, and plots of $1/T_m$ versus lnC. Tables S1 and S2: assignment lists. Table S3: melting temperatures for different protons in the nucleopeptide **I** and in the control duplex **III**. Table S4: relaxation times for different protons in compound **I**. Table S5: a list of peptide–DNA NOE contacts.

We gratefully acknowledge Dr. Doug Laurents for careful reading of the manuscript. This work was supported by the DGICYT ■■please define ■■ (Grant no. PB97-0941-C02-01/02), the Generalitat de Catalunya (Grant nos. 2000SGR18, 2001SGR49), and the Centre de Referència de Biotecnologia.

[1] G. M. Cheetham, T. A. Steitz, *Science* **1999**, *286*, 2305–2309.
[2] C. A. Brautigam, T. A. Steitz, *Curr. Opin. Struct. Biol.* **1998**, *8*, 54–63.
[3] A. Bochkarev, R. A. Pfuetzner, A. M. Edwards, L. Frappier, *Nature* **1997**, *385*, 176–181.
[4] A. Travers, *Curr. Opin. Struct. Biol.* **2000**, *10*, 102–109.
[5] J. J. Love, X. Li, D. A. Case, K. Giese, R. Grosschedl, P. E. Wright, *Nature* **1995**, *376*, 791–795.
[6] J. L. Kim, S. K. Burley, *Nat. Struct. Biol* **1994**, *1*, 638–653.
[7] Y. Kim, J. H. Geiger, S. Hahn, P. B. Sigler, *Nature* **1993**, *365*, 512–520.
[8] J. L. Kim, D. B. Nikolov, S. K. Burley, *Nature* **1993**, *365*, 520–527.
[9] M. H. Werner, J. R. Huth, A. M. Gronenborn, G. M. Clore, *Cell* **1995**, *81*, 705–714.
[10] J. J. Toulme, C. Hélène *J. Biol. Chem.* **1977**, *252*, 244–249.
[11] C. Hélène, J. L. Dimicoli *FEBS Lett.* **1972**, *26*, 6–10.
[12] S. L. McKay, B. Haptonstall, S. H. Gellman *J. Am. Chem. Soc.* **2001**, *123*, 1244–1245.
[13] S. Padmanabhan, W. Zhang, M. W. Capp, C. F. Anderson, M. T. Record, Jr. *Biochemistry* **1997**, *36*, 5193–5206.
[14] K. B. Roy, S. Kukreti, H. S. Bose, V. S. Chauhan, M. R. Rajeswari, *Biochemistry* **1992**, *31*, 6241–6245.
[15] M. R. Rajeswari, H. S. Bose, S. Kukreti, A. Gupta, V. S. Chauhan, K. B. Roy, *Biochemistry* **1992**, *31*, 6237–6241.
[16] M. R. Rajeswari, T. Montenay-Garestier, C. Hélène, *Biochemistry* **1987**, *26*, 6825–6831.
[17] J. Sartorius, H. J. Schneider, *FEBS Lett.* **1995**, *374*, 387–392.
[18] N. R. Shine, T. L. James, *Biochemistry* **1985**, *24*, 4333–4341.
[19] A. Khiat, M. Lamoureux, Y. Boulanger *J. Med. Chem.* **1996**, *39*, 2492–2498.
[20] C. Robledo-Luiggi, W. D. Wilson, E. Pares, M. Vera, C. S. Martinez, D. Santiago, *Biopolymers* **1991**, *31*, 907–917.
[21] Y. Mihara, M. Doi, T. Inohara, M. Kawamura, N. Hamanaka, T. Ishida, *Biochem. Biophys. Res. Commun.* **1997**, *240*, 803–806.
[22] D. P. Mascotti, T. M. Lohman, *Biochemistry* **1993**, *32*, 10568–10579.
[23] D. P. Mascotti, T. M. Lohman, *Biochemistry* **1997**, *36*, 7272–7279.
[24] W. Zhang, J. P. Bond, C. F. Anderson, T. M. Lohman, M. T. Record, Jr., *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 2511–2516.
[25] W. C. Ho, C. Steinbeck, C. Richert, *Biochemistry* **1999**, *38*, 12597–12606.
[26] J. C. Wang, *Annu. Rev. Biochem.* **1996**, *65*, 635–692.
[27] M. Salasm, *Annu. Rev. Biochem.* **1991**, *60*, 39–71.
[28] M. Junghans, J. Kreuter, A. Zimmerm, *Nucleic Acids Res.* **2000**, *28*, E45.
[29] W. Mier, R. Eritja, A. Mohammed, U. Haberkorn, M. Eisenhut, *Bioconjugate Chem.* **2000**, *11*, 855–860.
[30] L. Debéthune, G. Kohlhagen, A. Grandas, Y. Pommier, *Nucleic Acids Res.* **2002**, *30*, 1198–1204.
[31] C. D. Juby, C. D. Richardson, R. Brousseau, *Tetrahedron Lett.* **1991**, *32*, 879–882.
[32] C. H. Tung, S. Stein, *Bioconjugate Chem.* **2000**, *11*, 605–618.
[33] M. Beltrán, M. Maseda, J. Robles, E. Pedroso, A. Grandas, *Lett. Pept. Sci.* **1997**, *4*, 147–155.

[34] K. Wüthrich, *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York, **1986**.
[35] R. M. Scheek, N. Russo, R. Boelens, R. Kaptein, J. H. Van Boom, *J. Am. Chem. Soc.* **1983**, *105*, 1914 – 1916.
[36] J. G. Harrison, S. Balasubramanian, *Nucleic Acids Res.* **1998**, *26*, 3136 – 3145.
[37] K. J. Breslauer, *Methods Enzymol.* **1995**, *259*, 221 – 242.
[38] E. Suzuki, N. Pattabiraman, G. Zon, T. L. James, *Biochemistry* **1986**, *25*, 6854 – 6865.
[39] P. Guntert, C. Mumenthaler, K. Wüthrich, *J. Mol. Biol.* **1997**, *273*, 283 – 298.
[40] D. A. Case, D. A. Pearlman, J. W. Caldwell, T. E. Cheatham III, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, D. M. Ferguson, R. J. Radmer, G. L. Seibel, U. C. Singh, P. K. Weiner, P. A. Kollman, AMBER 5.1., 5th ed., University of California, San Francisco, **1997**.
[41] D. E. Konerding, T. E. Cheatham III, P. A. Kollman, T. L. James, *J. Biomol. NMR* **1999**, *13*, 119 – 131.
[42] E. Trotta, E. D'Ambrosio, G. Ravagnan, M. Paci, *J. Biol. Chem.* **1996**, *271*, 27 608 – 27 614.
[43] J. K. Strauss, L. J. Maher, *Science* **1994**, *266*, 1829 – 1834.
[44] T. Ishida, M. Doi, H. Ueda, M. Inoue, G. M. Scheldrick, *J. Am. Chem. Soc.* **1988**, *110*, 2286 – 2294.
[45] D. P. Mascotti, T. M. Lohman, *Biochemistry* **1992**, *31*, 8932 – 8946.
[46] F. Gago, *Methods* **1998**, *14*, 277 – 292.
[47] C. F. Bleczinski, C. Richert, *J. Am. Chem. Soc.* **1999**, *121*, 10 889 – 10 894.
[48] K. M. Guckian, B. A. Schweitzer, R. X.-F. Ren, C. J. Sheils, P. L. Paris, D. C. Tahmassebi, E. T. Kool, *J. Am. Chem. Soc* **1996**, *118*, 8182 – 8183.
[49] N. Morellet, H. Demene, V. Teilleux, T. Huynh-Dinh, H. de Rocquigny, M. C. Fournie-Zaluski, B. P. Roques, *J. Mol. Biol.* **1998**, *283*, 419 – 434.
[50] W. Schuler, C. Dong, K. Wecker, B. P. Roques, *Biochemistry* **1999**, *38*, 12 984 – 12 994.
[51] R. N. De Guzman, Z. R. Wu, C. C. Stalling, L. Pappalardo, P. N. Borer, M. F. Summers, *Science* **1998**, *279*, 384 – 388.
[52] M. H. Werner, A. M. Gronenborn, G. M. Clore, *Science* **1996**, *271*, 778 – 784.
[53] S. E. Tsutakawa, H. Jingami, K. Morikawa, *Cell* **1999**, *99*, 615 – 623.
[54] J. Robles, M. Beltrán, V. Marchán, Y. Pérez, I. Travesset, E. Pedroso, A. Grandas, *Tetrahedron* **1999**, *55*, 13 251 – 13 264.
[55] L. Debéthune, V. Marchán, G. Fábregas, E. Pedroso, A. Grandas, *Tetrahedron* **2002**, *58*, 6965 – 6978.
[56] A. Bax, D. J. Davies, *J. Magn. Reson.* **1985**, *65*, 355 – 360.
[57] A. Kumar, R. R. Ernst, K. Wüthrich, *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1 – 6.
[58] P. Plateau, M. Güeron, *J. Am. Chem. Soc* **1982**, *104*, 7310 – 7311.
[59] M. Piotto, V. Saudek, V. Sklenar, *J. Biomol. NMR* **1992**, *2*, 661 – 665.
[60] C. Bartels, T. Xia, M. Billeter, P. Güntert, K. Wüthrich, *J. Biomol. NMR* **1995**, *6*, 1 – 10.
[61] R. Freeman, S. P. Kempsell, M. H. Levitt, *J. Magn. Reson.* **1980**, *38*, 453 – 479.
[62] B. A. Borgias, T. L. James, *J. Magn. Reson.* **1988**, *79*, 493 – 512.
[63] B. A. Borgias, T. L. James, *J. Magn. Reson* **1990**, *87*, 475 – 487.
[64] L. J. Rinkel, C. Altona *J. Biomol. Struct. Dyn.* **1987**, *4*, 621 – 649.
[65] W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, **1984**.
[66] S. Arnott, D. W. Hukins, *Biochem. Biophys. Res. Commun.* **1972**, *47*, 1504 – 1509.
[67] N. Escaja, E. Pedroso, M. Rico, C. González, *J. Am. Chem. Soc.* **2000**, *122*, 12 732 – 12 742.
[68] R. Soliva, V. Monaco, I. Gomez-Pinto, N. J. Meeuwenoord, G. A. Marel, J. H. Boom, C. González, M. Orozco *Nucleic Acids Res.* **2001**, *29*, 2973 – 2985.
[69] T. E. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10 089 – 10 092.
[70] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 11 946 – 11 975.
[71] C. I. Bayly, P. Cieplak, W. D. Cornell, P. A. Kollman, *J. Phys. Chem.* **1993**, *97*, 10 269 – 10 280.

Received: June 25, 2002 [F 444]