Structural Differences between the Free and Bound States of the DNA-Bisintercalating Peptide YSPTSPSY

Abdesslem Khiat, Martin Lamoureux, and Yvan Boulanger*

INRS-Santé, Université du Québec, 245 boul. Hymus, Pointe-Claire, Québec, Canada H9R 1G6

Received May 3, 1995[⊗]

The YSPTSPSY peptide is a DNA-bisintercalator that can adopt nonrandom conformations in solution. Strategies based on random conformational search and energy minimizations have been applied to generate populations of conformers characterizing YSPTSPSY. Subsequent analysis based on statistical methods and clustering allowed to determine the existence of four classes of conformers containing β- and/or γ-turns. NMR spectra of YSPTSPSY in solution provide evidence for such structures. Employing a Monte Carlo-based docking procedure, the YSPTSPSY peptide was docked in a DNA double-helical fragment with the sequence $[d(GACGTC)]_2$. The peptide binds on the minor groove of DNA stacking the central CG base pairs, in a manner similar to that observed in complexes of triostin A with DNA. Upon binding, the structure of the C-terminal segment is modified into a type I β-turn. Five intermolecular hydrogen bonds are observed, but the van der Waals interactions constitute the major stabilization factor for the complex. NMR chemical shifts, coupling constants, and NOESY connectivities are in agreement with the molecular model.

Introduction

The heptad repeat unit SPTSPSY of RNA polymerase II has been proposed to bind to transcription factors.[1] Extending this heptad unit with one tyrosine at the N-terminus yields the octapeptide YSPTSPSY (peptide I) which has been shown to bind to DNA[2] as a member of a peptide family containing two terminal planar rings.[3−5] This peptide contains two overlapping Ser-Pro-X-X sequences often found in gene regulatory proteins.[6] Experimental evidence based on fluorescence measurements[2] show that peptide I binds to DNA by the intercalation of tyrosine residues. In the model proposed by Suzuki,[2] both Ser-Pro-X-X sequences form type I β-turns, one being interlocked to the other. The first β-turn is formed by Ser^2-Pro^3-Thr^4-Ser^5 with the existence of two intramolecular hydrogen bonds, one between the carbonyl oxygen of Ser^2 and the amine hydrogen of Ser^5 and the other between the β-hydroxyl oxygen of Ser^2 and the amine hydrogen of Thr^4. The second β-turn is formed by the Ser^5-Pro^6-Ser^7-Tyr^8 segment and involves hydrogen bonding between the carbonyl oxygen of Ser^5 and the amine hydrogen of Tyr^8. An additional hydrogen bond was also postulated between the side chain oxygen of Ser^5 and the amine hydrogen of Ser^7. Formation of these two β-turns serves to orient the two terminal tyrosine rings in adequate position for spanning two CG base pairs of DNA.[2] A previous NMR study of the peptide in aqueous solution at pH 3.2[7] showed good evidence for the presence of type I and type II β-turns in the Ser^2-Pro^3-Thr^4-Ser^5 segment but not for the Ser^5-Pro^6-Ser^7-Tyr^8 segment. In this study, we used molecular modeling calculations to demonstrate the existence not only of the structure proposed by Suzuki[2] but also of structures comprising one β-turn and one γ-turn and structures comprising two γ-turns. These structures are very stable with the terminal tyrosine residues positioned favorably for interaction with the minor groove of DNA. NMR spectra of the peptide in solution were also recorded and support our analysis.

Triostin A is a bicyclic antibiotic with a repeated sequence of four amino acids, (D-Ser-Ala-N-MeCys-N-MeVal)$_2$, containing two planar aromatic quinoxaline rings attached to the D-serine residues.[3,8] This molecule is C-shaped with a central region containing the cyclic depsipeptide with the two quinoxaline rings perpendicular to the peptide backbone and parallel to each other. This arrangement orients the quinoxaline rings in favorable position to intercalate into DNA. The crystal and NMR structures of the triostin A−DNA complex[8−10] show that triostin A binds to the minor groove as bisintercalator around the CG base pairs. The alanine residues of the peptide are positioned so that their NH and carbonyl groups are located inside the C-shaped structure at the interface with the nucleic acid. These alanine residues stabilize the complex by forming four hydrogen bonds with the nucleic acid bases: two hydrogen bonds are formed between the amine hydrogen of the alanines and the N3 of the guanines, and two hydrogen bonds are formed between the carbonyl oxygens of the alanines and the NH_2 hydrogens of the guanines. However, the major binding interaction between the drug and the nucleic acid involves a large number of van der Waals contacts.

In this study, the mode of bisintercalation of peptide I into DNA was investigated by docking calculations as well as by NMR and found to involve a rearrangement of the C-terminal structure. Details of the structures and interactions are presented and compared to the crystal and NMR structures of the triostin A−$[d(GACGTC)]_2$ complex.

Experimental Section

Random Conformational Search. The peptide sequence of YSPTSPSY (peptide I) was built using the standard residue library of the ECEPP (empirical conformation energy program for peptides) force field.[11,12] The PEPSEA (peptidic search) method[13] was used which, instead of searching a global minimum, generates the most representative peptide population. A total of 15 000 conformers was generated by randomly

* To whom correspondence should be addressed.
⊗ Abstract published in *Advance ACS Abstracts,* May 15, 1996.

S0022-2623(95)00325-6 CCC: $12.00

changing 24 torsion angles (Φ, Ψ, χ) of the backbone and the side chains. Dihedral angles were labeled according to the IUPAC-IUB Commission on Biochemical Nomenclature.[14] The conformational energies of the resulting structures were calculated using the ECEPP force field and energy minimized using the conjugate gradients algorithm[15] for 1000 iterations with a rms gradient of 0.001 kcal·mol^{-1}·Å^{-1}. The validity of the population sample was assessed by the bell shape of the energy histogram. Individual analysis of the first 100 structures allowed a classification of the conformers into characteristic families. The calculations were performed on an IBM RISC/6000 model 520 workstation (Laboratoire de modélisation moléculaire, Université de Sherbrooke, Canada). The SYBYL molecular modeling program (Tripos Associates Inc., St. Louis, MO) was used for calculating and analyzing the molecular structures.

Docking Procedure. Docking studies were carried out on the Silicon Graphics workstation using the software modules INSIGHTII and AUTODOCK[16] from Biosym Technologies, San Diego, CA. The Biosym consistent valence force field (CVFF) was used to build the peptide, and the B-DNA structure, determined by NMR in the presence of triostin A,[10] was obtained from the Protein Data Bank (Brookhaven National Laboratory, Upton, NY). This DNA sequence was chosen because triostin A and peptide I belong to the same family of structures.[2] Peptide I was docked in the minor groove of the DNA fragment using the Monte Carlo docking procedure. This procedure uses a high-temperature dynamics calculation. A key component of the procedure is the Monte Carlo-driven insertion of peptide conformers into the DNA structure based on the random positioning within the DNA molecule followed by a calculation of the interaction energy with the affinity grid. Displacing the peptide results in either an increased or a decreased affinity energy. When the affinity energy is increased, the positioning is rejected and a new random positioning is chosen. When the affinity energy is decreased, the positioning is accepted and displacement is pursued in the same direction until the interaction energy is practically constant. The resulting complex, which is often in a highly energetic state, is then energy minimized using the conjugate gradients algorithm until the rms gradient is less than 0.001 kcal·mol^{-1}·Å^{-1}.

Synthesis. The nucleotide hexamer GACGTC was purchased from the University Core DNA Services, University of Alberta, Calgary, Canada. Peptide I (YSPTSPSY) was synthesized by the solid phase method using the BOP (benzotriazol-1-yl-*N*-oxytris(dimethylamino)phosphonium hexafluorophosphate) coupling reagent and Fmoc (9-fluorenylmethyloxycarbonyl) amino acids with appropriate side chain protection.[17] Cleavage from the resin and deprotection was accomplished by a 60 min treatment at 273 K with liquid hydrofluoric acid containing anisole as scavenger (10:1). The crude peptide was finally purified by preparative C_{18} reverse phase HPLC. The final product was found to be more than 97% pure.

NMR Spectroscopy. NMR spectra were recorded for peptide I in DMSO-d_6 (20 mM) and for peptide I (31 mM), the oligonucleotide d(GACGTC)$_2$ (8.3 mM), and the peptide I:d(GACGTC)$_2$ (1:1) (8.3 mM) in 50 mM phosphate buffer at pH 6.8. All ^{1}H NMR spectra were acquired on a Bruker AMX2 500 spectrometer (Institut Armand-Frappier, Université du Québec, Laval, Canada) at a frequency of 500.13 MHz and at a temperature of 278 or 300 K. Double-quantum-filtered correlation spectroscopy (DQFCOSY), total correlation spectroscopy (TOCSY, mixing times 60−90 ms), and nuclear Overhauser enhancement spectroscopy (NOESY, mixing times 150−300 ms) spectra were collected using the time proportional phase incrementation mode with 1024 data points and 256 t_1 increments and with 32, 16, and 64 transients, respectively. Processing of NMR spectra was performed on a Silicon Graphics Indigo R4000 XZ workstation using the program FELIX 2.30 (Biosym Technologies, Inc., San Diego, CA). Prior to Fourier transformation, two-dimensional spectra were multiplied by a 30° shifted sinebell function. Data were zero-filled to yield a 2048 × 2048 matrix with a digital resolution of 3.0 Hz/point. Sequential assignments were performed using conventional methods.[18]

Results and Discussion

Conformational Domains of Peptide I. Conformations of peptide I generated by the random conformational search have been classified in accordance with the conventional letter-coded regions of the (Φ,Ψ) maps.[19] The most stable conformers are distributed into four major families presented in Figure 1. Table 1 presents a list of the structured conformers among the first 100 calculated structures divided in families with their identification number, conformational code, type of β-turn, conformational energy, frequency of occurrence, and presence of hydrogen bonds. All hydrogen bonds were shorter than 2.9 Å and their XH···Y angles larger than 100°.

The first family, which comprises the majority of structures, is characterized by a β-turn in the N-terminal segment Ser^2-Pro^3-Thr^4-Ser^5 (Table 1). Most conformers of this family show the presence of hydrogen bonding between the carbonyl oxygen of Ser^5 and the amine hydrogen of Ser^7, forming a γ-turn in the C-terminal segment. An important number of conformers of this family are stabilized with a third hydrogen bond between the side chain hydroxyl oxygen of Ser^2 and the amine hydrogen of Thr^4. Conformers of the second family are characterized by two hydrogen bonds, one between the carbonyl oxygen of Ser^2 and the amine hydrogen of Thr^4, the other between the carbonyl oxygen of Ser^5 and the amine hydrogen of Ser^7. These hydrogen bonds give rise to two γ-turns, one in each extremity of the peptide chain. The third family comprises conformers containing a β-turn in the C-terminal segment Ser^5-Pro^6-Ser^7-Tyr^8, along with the presence of hydrogen bonding between the carbonyl oxygen of Ser^5 and the amine hydrogen of Tyr^8. Most conformers of this family display a second hydrogen bond between the carbonyl oxygen of Ser^2 and the amine hydrogen of Thr^4. Some conformers of this family are stabilized with a third hydrogen bond between the side chain oxygen of Ser^5 and the amine hydrogen of Ser^7. Finally, the fourth family is characterized by two β-turns located respectively in the N- and C-terminal segments. Figure 2 shows the superimposition of structures belonging to the four families. Several structures belonging to none of the four families remained. These presented other hydrogen-bonding patterns or were extended structures.

The largest number of structures belongs to family 1 which includes the most stable conformers, energetically and thermodynamically (frequency of occurrence). On the other hand, conformers with two β-turns (family 4) are the least stable energetically and the least abundant (Table 1). These results also indicate that the C-terminal portion of peptide I is less structured than the N-terminal portion, in agreement with the NMR results.[7]

Analysis of the structures in each family allows a characterization of the geometrical parameters, especially the distances between the β-carbons and hydroxyl oxygens of the terminal tyrosine residues. In most conformers, the aromatic rings of both tyrosines are located on the same side of the molecule and their Cβ−Cβ and hydroxyl O−O distances are comparable to the reported values of 11.1 and 10.2 Å, respectively.[2] Superimposition of one structure of each family with the

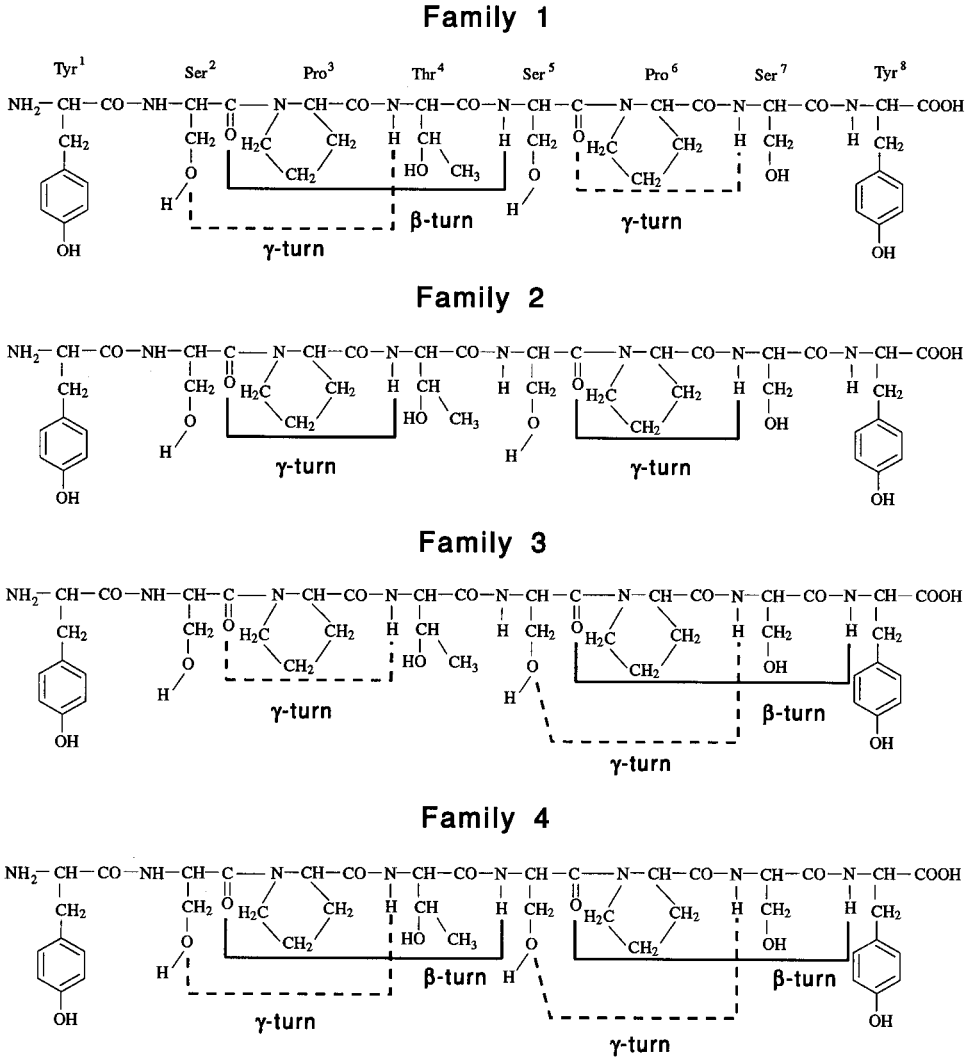

Figure 1. Chemical structures of the four families of peptide I obtained following random conformational search and energy minimization. The locations of hydrogen bonds leading to the formation of β- or γ-turns existing in all structures and the formation of possible γ-turns are given by continuous and dotted lines, respectively.

structure proposed by Suzuki[2] is illustrated in Figure 3 and demonstrates that the presence of two β-turns is not necessary to position the aromatic rings on the same side of the molecule with Tyr^1−Tyr^8 distances suitable for intercalation into DNA. In fact, conformers of families 1−3 containing γ-turn structures present their terminal tyrosine rings in positions very similar to the Suzuki model.[2]

Our approach with multiconformational analysis allows us to underline two important points. First, clear evidence is obtained that the structure of unbound peptide I is not a random coil but that the conformational space is populated with structures in which the Ser-Pro-X-X segments form mostly type III β-turns and occasionally type I or type II β-turns (the structures of type I and type III β-turns are very similar[18]). Second, comparing with Suzuki's model,[2] we have been able to show (Figure 3) that structures with a N-terminal β-turn and a C-terminal γ-turn (family 1), a N-terminal γ-turn and a C-terminal β-turn (family 3), or two γ-turns (family 2) are very stable presenting in adequate position the terminal tyrosine rings for intercalation into DNA.

NMR Structural Data for Peptide I. The results of the multiconformational analysis are supported by the NMR data of peptide I in solution. As illustrated in the TOCSY spectrum in Figure 4A,B, the presence of major and minor peaks confirms the existence of multiple structures. Analysis of the major structure allows to identify NOESY connectivities between the δ protons of Pro^3 and the NH proton of Thr^4 as well as between the δ protons of Pro^6 and the NH proton of Ser^7 (Figure 4C). These $d_{\delta N}$ connectivities are consistent with type I and type III β-turns or γ-turns where the shortest interproton distances are 2.7, 2.7, and 3.0 Å, respectively. In type II β-turns, the $d_{\delta N}$ distance is of the order of 5.6 Å and would therefore be unobservable by NMR. The coupling constants $^3J_{NH\alpha}$ for Thr^4 and Ser^7 were 8.6 and 7.6 Hz, respectively, in DMSO-d_6 at 300 K and 7.7 and 7.4 Hz, respectively, in phosphate buffer at 278 K. Within experimental error, such values are consistent with a type I β-turn (calculated value 7.9 Hz), a γ-turn (calculated value 6.4 Hz), or a random coil structure. For a type III β-turn, the calculated $^3J_{NH\alpha}$ value is of the order of 4.2 Hz. Therefore, molecular modeling calculations favor type III β-turns, whereas the NMR results favor the presence of type I β-turns. Since the structures of type I and type III β-turns are very similar, this difference could be explained by the fact that molecular modeling calculations do not take into account solvent effects.

Structure of the Peptide I−DNA Complex. Fig-

Table 1. Characteristics of Structured Conformers of Peptide I among the First 100 Conformers Generated by Random Conformational Search

conformer no.	Zimmerman code[a]	type of β-turn	energy (kcal/mol)	frequency of occurrence	H-bonds[b]	conformer no.	Zimmerman code[a]	type of β-turn	energy (kcal/mol)	frequency of occurrence	H-bonds[b]
				Family 1: N-Terminal β-Turn							
1	F*EAA	III	−41.01	1	3	36	FFAA	III	−18.56	1	2
2	FAAA	III	−39.39	2	2, 3	39	B*EAA	III	−16.99	1	
3	EAAE	III	−38.41	2		41	FFAA	III	−16.23	1	3
4	EAAE	III	−36.56	1		42	GDAB	I	−15.67	1	2
5	GDAA	III	−35.36	1	2, 3	45	BFAA	III	−13.82	2	3
6	GDAA	III	−34.09	1		47	D*FAA	III	−12.02	2	3
7	FA*AA	III	−33.79	1		50	FA*CA*	II	−11.29	1	3
8	FAAE	III	−32.47	2	2, 3	51	FFAA	III	−10.77	1	2
10	EA*AA	III	−30.12	1	3	52	AA*AA	III	−10.10	1	2
12	BEAA	III	−29.88	2		53	BCAB	I	−9.31	1	3
16	H*FCA	I	−28.04	1	3	54	B*EAD	III	−8.90	2	3
23	FECA	I	−25.01	1	2, 3	57	CFFA*	II	−8.20	1	
25	B*EAA	III	−24.22	2		59	FF*AA	III	−6.56	2	3
26	D*A*AA	III	−24.00	1		60	FA*GG	III	−5.02	1	2
31	FA*AA	III	−21.11	2	3	61	BA*AA	III	−4.41	1	3
34	B*EAB	I	−19.92	2	3						
				Family 2:[c] Two γ-Turns							
9			−31.41	1		40			−16.44	2	
14			−29.39	2		44			−14.21	1	
17			−27.98	1		46			−13.59	1	
22			−25.91	1		49			−11.57	1	
27			−22.70	1		58			−7.29	1	
33			−20.13	1		62			−3.07	2	
38			−17.36	1		63			−1.66	1	
				Family 3: C-Terminal β-Turn							
11	FAAA*	III	−30.07	2	1, 4	21	FAAB	III	−26.20	2	1
13	F*ARE	I	−29.65	1	4	24	EABA	I	−24.32	1	
15	ECD*E	II	−29.00	1	1, 4	28	A*AAF*	III	−22.22	1	4
18	DAAC	III	−27.88	1	1, 4	30	A*CA*E	II	−21.84	1	
19	EAAG	III	−27.01	2	1, 4	37	DCGC*	III	−17.57	1	1
20	DCGE	III	−26.73	1	1	43	FCA*E	II	−14.52	1	
				Family 4: Two β-Turns							
29	FAADAAC	III, III	−22.14	2	2	35	FAAFCA*E	III, II	−19.60	1	
32	DABEABA	I, I	−20.69	1	2, 4						

[a] Zimmerman codes of the amino acids in the β-turn segment:[19] residues Ser^2-Pro^3-Thr^4-Ser^5 in the N-terminal segment and Ser^5-Pro^6-Ser^7-Tyr^8 in the C-terminal segment. [b] Hydrogen bonds correspond to 1, between the carbonyl oxygen of Ser^2 and the amine hydrogen of Thr^4; 2, between the β-hydroxyl oxygen of Ser^2 and the amine hydrogen of Thr^4; 3, between the carbonyl oxygen of Ser^5 and the amine hydrogen of Ser^7; and 4, between the β-hydroxyl oxygen of Ser^5 and the amine hydrogen of Ser^7. [c] All conformers of this family displayed two γ-turns, one formed by the hydrogen bonding between the carbonyl oxygen of Ser^2 and the amine hydrogen of Thr^4 and the second between the carbonyl oxygen of Ser^5 and the amine hydrogen of Ser^7.

ure 5 displays the peptide I−$[d(GACGTC)]_2$ complex as viewed in the major groove (Figure 5A) and sideways (Figure 5B). As can be seen, peptide I binds to the minor groove with the two tyrosine rings stacking on each side of the CG base pairs. Different degrees of ring overlap are observed between the tyrosine and nucleotide bases. A comparison of the torsion angles indicates that the geometry of the oligonucleotide, which had been determined by NMR in the presence of triostin A,[10] is only slightly changed in the presence of peptide I.

A major stabilizing factor for the peptide I−DNA complex is provided by the existence of a large number of van der Waals interactions, the calculated energy of interaction being −94.03 kcal/mol. A total of 36 van der Waals distances shorter than 3.5 Å are present between the atoms of peptide I and those of the DNA fragment. An important stabilizing factor is the aromatic ring stacking which is maximal when the inter-ring distance is 3.4 Å. The Tyr^1 ring largely overlaps with the G^4 ring at a distance of 3.40 Å and partially with the A^8 ring at a distance of 3.50 Å. The aromatic ring of Tyr^8 overlaps strongly with the G^{10} ring at a distance of 3.52 Å and with the A^2 ring but at the longer distance of 4.00 Å. In addition, van der Waals interactions exist between Ser^2, Ser^5, and the central CG base pairs. Five intermolecular hydrogen bonds are formed as follows: between the amine hydrogen of Ser^2 and the NH_2 nitrogen of G^4, between the carbonyl oxygen of Ser^2 and the NH_2 hydrogens of G^{10}, between the side chain hydroxyl hydrogen of Ser^2 and the NH_2 nitrogen of G^4, between the side chain hydroxyl hydrogen of Ser^5 and the NH_2 nitrogen of G^8, and between the carbonyl oxygen of Ser^7 and the N3 of A^2. This suggests that Ser^2 and Ser^5 are the key residues for the binding of peptide I to the central CG base pairs.

Comparison of Peptide I−DNA and Triostin A−DNA Complexes. Similarities are found between the peptide I−DNA complex and the triostin A−DNA complex. In both complexes, strong van der Waals contacts and four hydrogen bonds are found between the central CG base pairs and the serine or alanine residues, for peptide I or triostin A, respectively. The stacking interactions between the tyrosine rings and the base pair rings in the peptide I−DNA complex are similar to those observed between the six-membered ring of quinoxaline in triostin A and the base pair rings in the triostin A−DNA complex. In both complexes, multiple van der Waals interactions constitute the major component of the stabilization.[8−10]

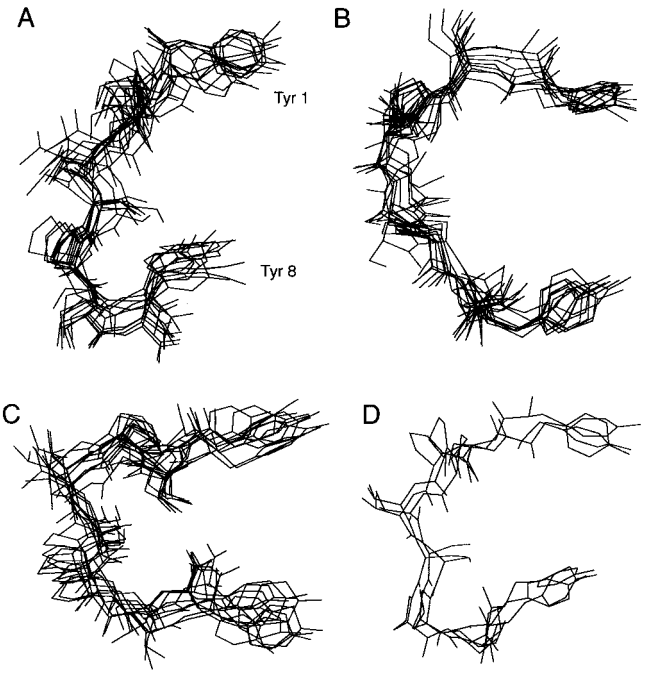

Figure 2. Superimposition of structures generated by random conformational search and energy minimization for (A) family 1, (B) family 2, (C) family 3, and (D) family 4 of peptide I.

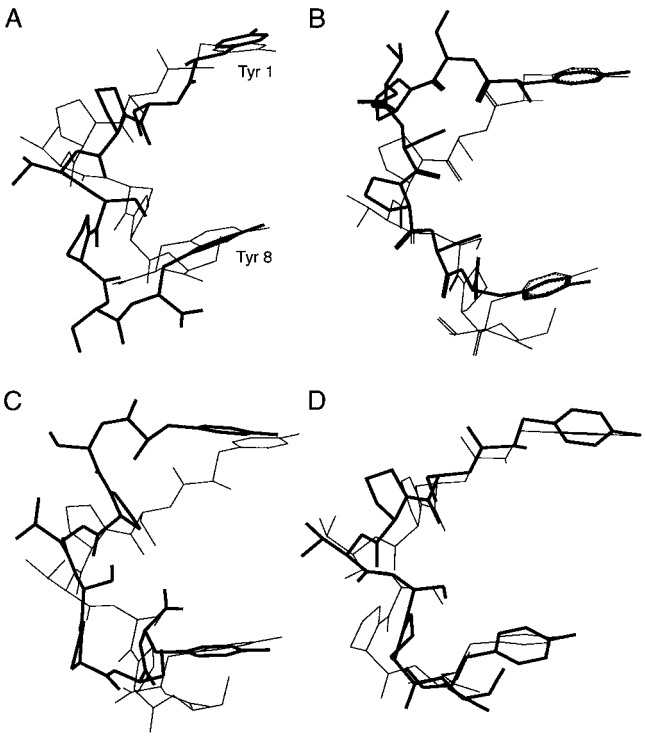

Figure 3. Superimposition of peptide I structures (thick lines) belonging to (A) family 1, (B) family 2, (C) family 3, and (D) family 4 with the peptide I structure proposed by Suzuki (narrow lines).[2]

Superimposition of the intercalated structures of peptide I and triostin A (Figure 6) shows strong similarities between the two molecules. Both tyrosine rings in the case of peptide I and both quinoxaline rings in the case of triostin A are perpendicular to the main chain, parallel together and positioned favorably for intercalation. Both peptides are C-shaped, and Thr[4] in peptide I seems to play the same structural role as the disulfide bridge in triostin A. In the peptide I structure, both terminal tyrosine residues are diagonally opposed

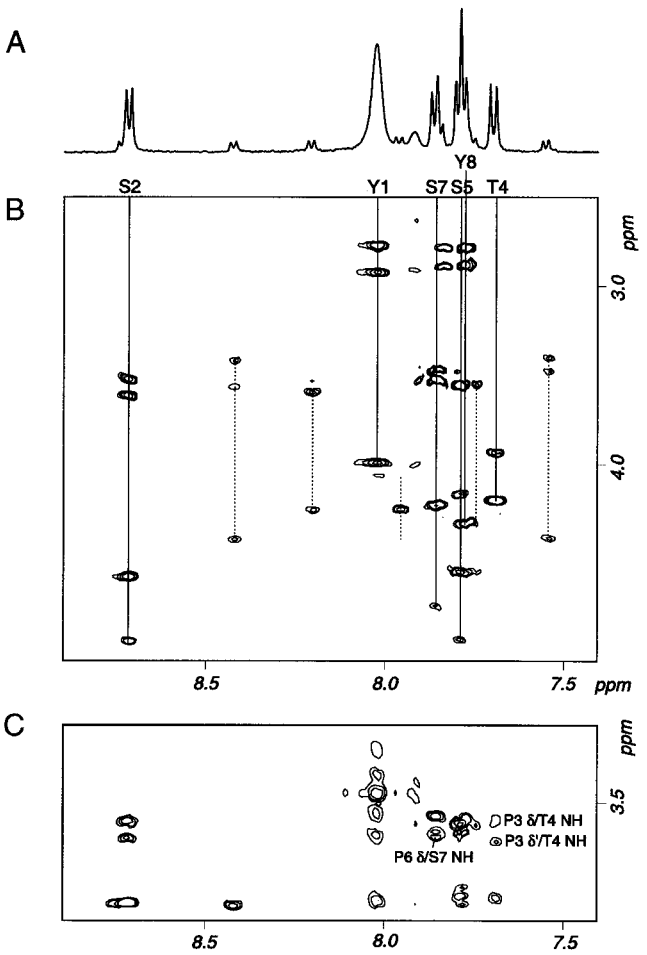

Figure 4. Sections of the ^{1}H NMR spectra of peptide I in DMSO-d_6 at 300 K: (A) NH region of the one-dimensional spectrum, (B) NH-α and NH-aliphatic regions of the TOCSY spectrum, and (C) NH-aliphatic region of the NOESY spectrum recorded with a mixing time of 300 ms. Dotted lines correspond to minor structures which have not been assigned.

to Thr[4], whereas Ser[2] and Ser[5] are pointing toward the nucleic acid and positioned to play a crucial role in complex formation. A similar situation is observed for triostin A where the serine residues are diagonally opposed to the disulfide bridge and the two alanine residues are pointing toward the nucleic acid in a favorable position for bisintercalation.

Conformation of Peptide I in the Complex. The lowest energy conformation of peptide I, belonging to family 1 (conformation 1 in Table 1), has been docked into the DNA fragment. This conformation was chosen because, according to our results, the majority of conformers belongs to family 1. It contains both a N-terminal β-turn (residues Ser[2]-Pro[3]-Thr[4]-Ser[5]) and a C-terminal γ-turn (residues Ser[5]-Pro[6]-Ser[7]). The docking results indicate that in the complex of peptide I with the central CG base pairs, the major stabilization factor is due to the interaction between the N-terminal part of peptide I and the CG base pairs. These interactions are only possible if the peptide adopts a conformation in which the N-terminal segment is ideally distanced from the CG base pairs, which occurs only when peptide I is folded to form a β-turn in its N-terminal part. As illustrated in Figure 7, conformers of family 2 or family 3 with a N-terminal γ-turn cannot bind to DNA, even though their tyrosine rings can intercalate. Atoms from the N-terminal segment remain far from the CG base

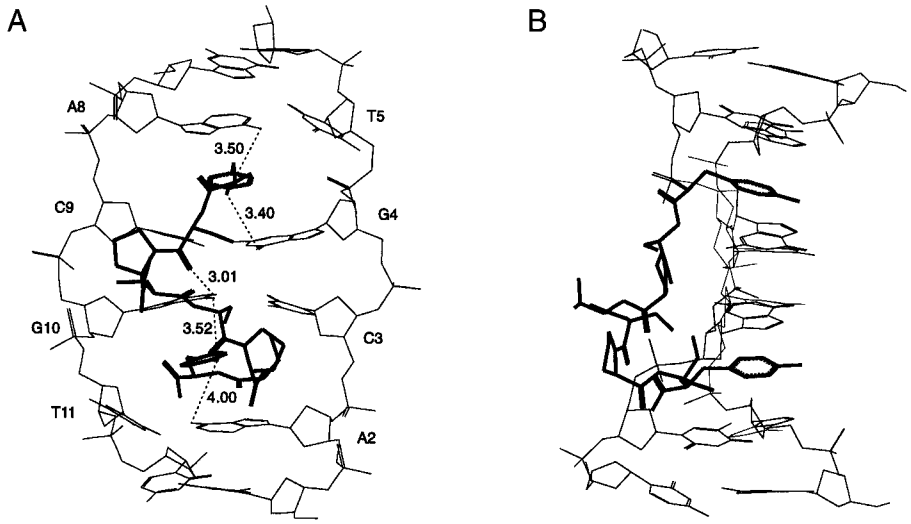

Figure 5. Structure of the complex formed by peptide I (thick lines) intercalated into the $[d(GACGTC)]_2$ oligonucleotide (narrow lines) following docking calculations: (A) front view into the major groove and (B) side view. The terminal nucleic acid base pairs are not shown. Several intermolecular distances (in Å) are given in part A.

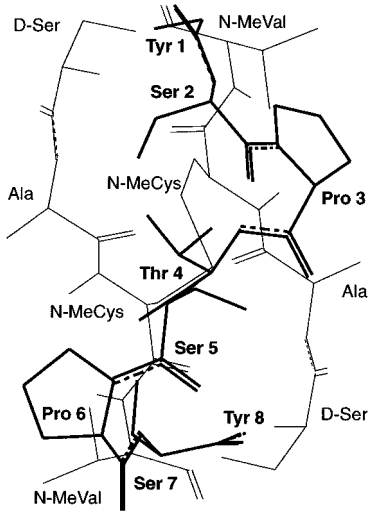

Figure 6. Comparison of the structures of peptide I (thick lines) and triostin A (narrow lines)[10] when bound to the same oligonucleotide. Structural similarities are evident, Thr[4] in peptide I overlapping with the disulfide bridge in triostin A.

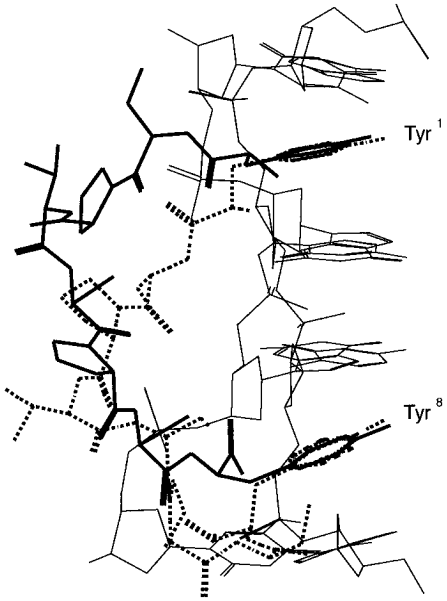

Figure 7. Partial structure of the peptide I−$[d(GACGTC)]_2$ complex showing the difference between a peptide I conformer from family 2 (thick lines) and the energy-minimized docked conformer (dotted lines). An important conformational difference is observed in the N-terminal segment which results in the impossibility of binding for a family 2 conformer with two γ-turns.

pairs, and therefore, interactions essential to the formation of the complex cannot occur. In our attempts to dock a conformer of family 2 containing two γ-turns, the peptide was always expelled from the DNA binding site.

Following docking and energy minimization of the complex, the structure of the family 1 conformer rearranged itself inside the complex. Formation of an intermolecular hydrogen bond between the carbonyl oxygen of Ser[2] and the NH_2 nitrogen of G[4] provoked the rupture of the intramolecular hydrogen bond between the carbonyl oxygen of Ser[2] and the amine hydrogen of Ser[5]. This leads to favorable interactions between the N-terminal segment of peptide I and the central CG base pairs. Concomitantly, the structure of the C-terminal segment rearranged itself in a more folded conformation to allow a better insertion of the Tyr[8] ring into the A^2-T^{11} and C^3-G^{10} base pairs. Analysis of the torsion angles of peptide I (Table 2) in the final complex structure indicates that the values of the Φ and Ψ angles of Pro[3] and Thr[4] are very similar to the angles characterizing a type I β-turn, except for the Ψ angle of Thr[4] which deviates by approximately 40° from the ideal value. The major structural change, however, occurs in the C-terminal segment of the peptide which changes from a γ-turn to a type I β-turn better defined than in the N-terminal segment. This new geometry of peptide I complexed to the DNA fragment is relatively similar to the structure proposed by Suzuki[2] solely on the basis of a comparison between peptide I and triostin A. It is particularly interesting to note the similarity of the interatomic distances between the terminal aromatic rings. The distances between the β-carbons and the hydroxyl oxygens of Tyr[1] and Tyr[8] were respectively of 11.1 and 10.2 Å in Suzuki's model and 11.07 and 10.08 Å in our model. Comparison of the peptide I conformations before and after the docking procedure suggests that the conformation of the peptide is modified due to bisintercalation and that the structures of peptide I free and bound to DNA are different.

Table 2. ^{1}H NMR Chemical Shifts for Peptide I in Solution and in the Complex

	chemical shift (ppm)					
residue	NH	αH	βH	γH	δH	ϵH
			In Solution[a]			
Tyr[1]	8.19	4.15	2.98, 3.04		7.04	6.73
Ser[2]	8.51	4.70	3.68, 3.73			
Pro[3]		4.33	1.89, 2.12	1.67	3.63, 3.73	
Thr[4]	8.21	4.25	4.16	1.12		
Ser[5]	8.30	4.70	3.73, 3.78			
Pro[6]		4.34	2.12, 2.27	1.92	3.54, 3.63	
Ser[7]	8.35	4.30	3.67			
Tyr[8]	8.16	4.56	2.89, 3.06		7.04	6.73
			In the Complex[a]			
Tyr[1]	7.80	4.18	2.94, 3.05		7.04	6.74
Ser[2]	8.51	4.71	3.69, 3.75			
Pro[3]		4.37	1.93, 2.15	1.72	3.66, 3.75	
Thr[4]	8.24	4.27	4.15	1.13		
Ser[5]	8.32	4.73	3.71, 3.83			
Pro[6]		4.37	1.96, 2.29	1.96	3.56, 3.68	
Ser[7]	8.42	4.33	3.72			
Tyr[8]	7.80	4.33	2.83, 3.02		7.04	6.74

[a] Spectra were recorded at 278 K in 50 mM phosphate buffer at pH 6.8.

Table 3. Comparison of Peptide I−Oligonucleotide Distances Calculated by Molecular Modeling and NOE Connectivities Observed by NMR

NOE connectivity[a]	range of distances in model[b] (Å)
Tyr[1] NH−G[4] 4′H	2.7−6.2
Ser[2] αH−C[3] 1′H	2.2−4.9
Tyr[8] αH−A[2] 1′H	3.3−5.7
Tyr[8] δH−G[4] 3′H	1.7−4.8
Tyr[8] ϵH−G[4] 3′H	2.7−5.2

[a] All connectivities were weak (3−5 Å) for a NOESY spectrum acquired with a mixing time of 300 ms. [b] Range of distances measured in the 10 lowest energy structures of the complex.

NMR Structure of the Peptide I−DNA Complex. NMR spectra recorded for the peptide I−DNA complex provide data in support of the calculated model. One of the major conclusions of the molecular modeling results is the fact that the conformation of the peptide is modified in the C-terminal part. A comparison of the chemical shifts of the peptide in solution and in the presence of the DNA fragment (Table 2) shows differences mostly in the C-terminal fragment, the main differences being observed between the NH protons of Ser[7] and Tyr[8] and between the α-proton of Tyr[8] and the β-protons of Pro[6] and Tyr[8]. A significant difference in chemical shift is also observed for the NH protons of Tyr[1], a residue involved in the interaction (Figure 5). The existence of type I β-turns in the complex is supported by the observation of $^3J_{NH\alpha}$ values of 7.4 and 7.6 Hz, respectively, for Thr[4] and Ser[7], the calculated value being 7.9 Hz. Finally, at least five unambiguous intermolecular NOESY connectivities could be identified between peptide I and d[GACGTC]$_2$. These values are listed in Table 3 and are shown to agree with the range of distances calculated in our molecular models. Therefore, the good agreement between our NMR measurements and the calculated molecular model demonstrates the validity of the multiconformational analysis approach.

Acknowledgment. The authors would like to acknowledge the assistance of Mr. Louis Senécal and Mr. Alain Larocque for computer programming and analysis of the NMR spectra and Dr. Serge St-Pierre for giving access to his laboratory for peptide synthesis. This work was supported by the Medical Research Council of Canada and the Institut National de la Recherche Scientifique.

References

(1) Sigler, P. B. Transcriptional Activation. Acid Blobs and Negative Noodles. *Nature* **1988**, *333*, 210−212.
(2) Suzuki, M. The Heptad Repeat in the Largest Subunit of RNA Polymerase II Binds by Intercalating to DNA. *Nature* **1990**, *344*, 562−565.
(3) *Chemistry of Antitumour Agents*; Wilman, D. E., Ed.; Blackie & Son Ltd.: Glasgow, 1990; pp 1−59; 403−436.
(4) Wakelin, L. P. G. Polyfunctional DNA Intercalating Agents. *Med. Res. Rev.* **1986**, *6*, 275−339.
(5) Waring, M. J.; Wakelin, L. P. G. Echinomycin: A Bifunctional Intercalating Antibiotic. *Nature* **1974**, *252*, 653−657.
(6) Suzuki, M. SPXX, a Frequent Sequence Motif in Gene Regulatory Proteins. *J. Mol. Biol.* **1990**, *207*, 61−84.
(7) Harding, M. M. NMR Studies of YSPTSPSY: Implications for the Design of DNA Bisintercalators. *J. Med. Chem.* **1992**, *35*, 4658−4664.
(8) Ughetto, G.; Wang, A. H.-J.; Quigley, G. J.; van der Marel, G. A.; van Boom, J. H.; Rich, A. A Comparison of the Structure of Echinomycin and Triostin A Complexed to a DNA Fragment. *Nucleic Acids Res.* **1985**, *13*, 2305−2323.
(9) Wang, A. H.-J.; Ughetto, G.; Quigley, G. J.; Hakoshima, T.; van der Marel, G. A.; van Boom, J. H.; Rich, A. The Molecular Structure of a DNA-Triostin A Complex. *Science* **1986**, *225*, 1115−1121.
(10) Addess, K. J.; Feigon, J. Sequence Specificity of Quinoxaline Antibiotics. 1. Solution Structure of a 1:1 Complex Between Triostin A and [d(GACGTC)]$_2$ and Comparison with the Solution Structure of the [N-MeCys[3],N-MeCys[7]]TANDEM-[d(GATATC)]$_2$ Complex. *Biochemistry* **1994**, *33*, 12386−12396.
(11) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. Energy Parameters in Polypeptides. VII. Geometric Parameters, Partial Atomic Charges, Nonbonded Interactions, Hydrogen Bond Interactions, and Intrinsic Torsional Potentials for the Naturally Occurring Amino Acids. *J. Phys. Chem.* **1975**, *79*, 2361−2381.
(12) Momany, F. A.; Carruthers, L. M.; McGuire, R. F.; Scheraga, H. A. Intermolecular Potentials from Crystal Data. III. Determination of Empirical Potential and Applications to the Packing Configurations and Lattice Energies in Crystals of Hydrocarbons, Carboxylic Acids, Amines, and Amides. *J. Phys. Chem.* **1975**, *78*, 1595−1620.
(13) Hassani, C. A. *La Génération et l'Analyse des Populations de Structures Peptidiques en Vue d'Étudier les Hypersurfaces d'Énergie Conformationnelle en Mécanique Classique*. Ph.D. Thesis, Université de Sherbrooke, Sherbrooke, Canada, 1991.
(14) IUPAC-IUB Commission on Biochemical Nomenclature. Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains. Tentative Rules (1969). *Biochemistry* **1970**, *9*, 3471−3479.
(15) Powell, M. J. D. Restart Procedures for the Conjugate Gradient Method. *Math. Program.* **1977**, 241−254.
(16) Goodsell, D. S.; Olson, A. J. Automated Docking of Substrates to Proteins by Simulated Annealing. *Proteins: Struct. Funct. Genet.* **1990**, *8*, 195−202.
(17) Fields, G. B.; Noble, R. L. Solid Phase Peptide Synthesis Utilizing 9−Fluorenylmethyloxycarbonyl Amino Acids. *Int. J. Pept. Protein Res.* **1990**, *35*, 161−214.
(18) Dyson, H. J.; Rance, M.; Houghten, R. A.; Lerner, R. A.; Wright, P. Folding of Immunogenic Peptide Fragments of Proteins in Water Solution I. Sequence Requirements for the Formation of a Reverse Turn. *J. Mol. Biol.* **1988**, *201*, 161−200.
(19) Zimmerman, S. S.; Pottle, M. S.; Némethy, G.; Scheraga, H. A. Conformational Analysis of the 20 Naturally Occurring Amino Acid Residues Using ECEPP. *Macromolecules* **1977**, *10*, 1−9.

JM9503254