Pergamon

Bioorganic & Medicinal Chemistry 8 (2000) 2643–2650

BIOORGANIC & MEDICINAL CHEMISTRY

CH/π Interactions in the Crystal Structure of TATA-Box Binding Protein/DNA Complexes†

Yoji Umezawa[a] and Motohiro Nishio[b,*]

[a]*Institute of Microbial Chemistry, 3-14-23 Kamiosaki, Shinagawa-ku, Tokyo 141-0021, Japan*
[b]*The CHPI Institute, 3-10-7 Narusedai, Machida-shi, Tokyo 194-0043, Japan*

Received 24 May 2000; accepted 14 July 2000

Abstract—Crystal structures of TATA box-binding proteins (TBP) of various sources bound to their promoter DNA (TATA box) were analyzed with use of our program CHPI. A number of short CH/C*sp*^{2} contacts have been unveiled in these complexes at the boundary of TBP and the TATA box minor groove. The result was discussed in the context of the CH/π interaction. Thus, the nature of nonpolar forces, reported in the past at the interface of the two components, has been attributed to the CH/π interaction. Furthermore, many CH/π contacts have been disclosed within the same strand of the promoter DNA. The structure of the TATA element, partially unwound and severely bent on complexation, seems to be stabilized by CH/π interactions; H2′ of the deoxyribose moiety and the *methyl group* in the thymine nucleotide play the primary role.

Introduction

The CH/π interaction is a kind of weak hydrogen bond occurring between soft acids and soft bases.[1] Previously, the present authors suggested the potential role of this attractive molecular force in stabilizing the structure of proteins.[2,3] Evidence for the CH/π interaction in protein chemistry[4] and biochemistry[5] has since been accumulated. In particular, we presented evidence, by analyzing crystal structures in the Brookhaven Protein Data Bank (PDB), that the CH/π interaction plays an important role in stabilizing the 3-D structure of proteins and their complexes.[6] High-level ab initio MO calculations supporting the CH/π theory have recently appeared.[7,8]

The TATA-box binding protein (TBP) serves as a linchpin for the assembly of pre-initiation complexes that regulate the transcription; binding of TBP is the first step where the particular base sequence of the DNA is read and recognized. On binding, TBP deforms the promoter element that contacts the protein throughout its length. As a consequence, the TATA box deviates from the canonical B-form DNA structure. It is interesting to know the interactions involved in the TBP/TATA-box complex to elucidate the mechanism of stabilizing such an unusual DNA structure. Here we present evidence that CH/π interactions play an essential role in stabilizing the crystal structure of TATA box in complex with the binding protein.

Method

A program (CHPI) was written to examine distances between CHs and π groups in proteins and nucleic acids deposited in PDB. The π-system may be an aromatic ring of phenylalanine, tyrosine, tryptophan or histidine residue or nucleic acid bases. The hydrogen may be a part of an alkyl group or a CH in an aromatic ring. An example is given for a C_6 π-system in Figure 1.

To participate in CH/π interaction, a hydrogen atom should be positioned above the π plane (region 1 in Figure 1). In order to cover other possibilities, several kinds of H/C distance and angle parameters were defined. The CH/π interatomic distances shorter than a cut-off value [D_{max} 3.05 Å (=2.9 Å (1.2 Å for C–H plus 1.7 Å for a half thickness of the aromatic molecule)[9] ×1.05)] with reasonable angle factors were considered as relevant for the presence of CH/π interaction. The numbers in the brackets in Table 1 correspond to these values. Coordinates of the hydrogen atoms from neutron diffraction studies are available for bovine pancreatic trypsin inhibitor (BPTI)[10] and ribonuclease A (RNase A).[11] Crystal data in PDB, however, do not ordinarily contain coordinates of hydrogen atoms. In

*Corresponding author. Tel.: +81-42-726-5956; fax: +81-42-726-5956; e-mail: dionisio@tim.hi-ho.ne.jp
†Information as to the CH/π interaction is available on the following website: http://www.tim.hi-ho.ne.jp/dionisio.

0968-0896/00/$ - see front matter
PII: S0968-0896(00)00197-8

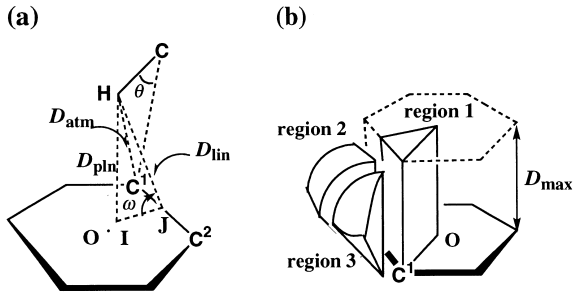

Figure 1. Method for exploring CH/π contacts with a six-membered carbon aromatic ring. O: centre of the π-plane. C^1 and C^2: nearest and second nearest sp^2-carbons, respectively, to the hydrogen H. ω: dihedral angle defined by C^1OC^2 and HC^1C^2 planes. θ: H–C–C^1 angle. D_{pln}: perpendicular distance between H and the π-plane (H/I). D_{atm}: HC^1 interatomic distance. D_{lin}: distance between H and the line C^1–C^2 (H/J). Regions to be searched. Region 1: zone where H is above the ring. Regions 2 and 3: zones where H is out of region 1 but may interact with π-orbitals. Unless otherwise noted, the program was run to search for short H/π contacts with the following conditions: $D_{max} = 3.05$ Å; $D_{pln} < D_{max}$ (region 1); $D_{lin} < D_{max}$ (region 2); $D_{atm} < D_{max}$ (region 3); $\omega_{max} = 127.5°$, $-\omega_{max} < \omega < \omega_{max}$; $\theta < 62.2°$.

these cases, hydrogens were generated on non-hydrogen atoms and their positions were optimized.[12] The CH/π distances were compared, in our earlier work, with those obtained by the neutron studies. Agreements of the two sets of data were satisfactory in the cases of BPTI[13] and RNase A,[14] at least for the purpose of surveying CH/π interactions in the structure of biopolymers. Our recent study on the peptide crystal structure[15] supports this. In analyzing the TBP complexes we chose more accurate data when available; however, the proton positions calculated on the basis of low-resolution data are not accurate enough in discussing the interatomic distances precisely. In view of the limitation to the present methodology we do not discuss the interatomic distances in detail.

Results

TBP has a phylogenetically conserved 180-residue carboxy terminal portion (TBPc) which binds to the consensus sequence TATAWAWN (W: A or T; N: any base) with high affinity, recognizing the minor groove determinants and inducing a severe DNA deformation. Crystal structures of TBPs in complex with their respective TATA element have been determined (human,[16] yeast,[17] plant[18] and bacteria[19]). Here, we analyzed human TBP (hTBP), yeast TPB (yTBP) and a plant TBP (aTBP), bound to the sequence 5′-CTGCTATAAAAGGCTG-3′, 5′-TGTATGTATATAAAAC-3′ and 5′-GCTATAA AAGGGCA-3′, respectively. These sequences contain a common TATA element TATAAAA.

CH/π interaction in proteins. Table 1 lists parts of our CHPI analyses of hTBP (PDB code 1CDW, resolution

Table 1. CH/π interactions in proteins of TBP/TATA-box complexes. (a) Human TBP; (b) yeast TBP; (c) plant TBP. Only phenylalanine residues interacting with TATA-box are listed. Interactions conserved among these species are printed in bold

IDRD	RES	VATM	IDRD	RES	VATM	DATM	DPLN	DLIN	Omega	Theta	RG
(a) hTBPc (1CDW, molecule A)											
A193	PHE	CD1	A194	ALA	HN	2.93	2.80	[2.90]	105.52	42.43	2
A193	PHE	CD2	A195	ALA	HN	[2.87]	2.71	—	109.29	31.95	3
A210	**PHE**	**CE2**	**A195**	**ALA**	**HB2**	**2.63**	**[2.60]**	**2.62**	**82.30**	**11.50**	**1**
A284	**PHE**	**CD1**	**A285**	**PRO**	**HD1**	**2.69**	**2.63**	**[2.69]**	**101.20**	**20.82**	**2**
A284	**PHE**	**CD2**	**A286**	**GLY**	**HN**	**[2.47]**	**2.37**	—	**105.99**	**29.58**	**3**
A301	**PHE**	**CE2**	**A286**	**GLY**	**HA1**	**[3.00]**	**2.95**	—	**100.21**	**46.07**	**3**
A301	PHE	CZ	A299	LEU	HB1	2.94	2.32	[2.87]	126.19	27.45	2
(b) yTBPc (1YTF, molecule A)											
A116	**PHE**	**CE2**	**A101**	**ALA**	**HB2**	**[2.67]**	**2.56**	—	**106.03**	**28.62**	**3**
A190	**PHE**	**CD1**	**A191**	**PRO**	**HD1**	**[2.54]**	**2.47**	—	**103.05**	**18.99**	**3**
A190	**PHE**	**CD2**	**A192**	**GLY**	**HN**	**[2.54]**	**2.43**	—	**106.87**	**29.33**	**3**
A190	PHE	CE2	A207	PHE	HE1	3.08	2.68	[3.04]	118.08	41.29	2
A207	**PHE**	**CE1**	**A192**	**GLY**	**HA1**	**[2.68]**	**2.48**	—	**112.23**	**38.92**	**3**
A207	PHE	CD1	A192	GLY	HA2	[3.02]	2.65	—	118.65	56.38	3
A207	PHE	CZ	A205	LEU	HB1	3.07	2.53	[3.04]	123.55	35.34	2
(c) aTBPc (1QNE, molecule A)											
A74	**PHE**	**CE2**	**A59**	**ALA**	**HB2**	**[2.49]**	**2.40**	—	**105.53**	**28.43**	**3**
A148	**PHE**	**CD1**	**A149**	**PRO**	**HD1**	**[2.72]**	**2.67**	—	**101.79**	**22.35**	**3**
A148	**PHE**	**CD2**	**A150**	**GLY**	**HN**	**[2.49]**	**2.45**	—	**101.06**	**29.58**	**3**
A148	PHE	CZ	A165	PHE	HE2	2.98	2.64	[2.95]	116.49	39.71	2
A165	**PHE**	**CE2**	**A150**	**GLY**	**HA1**	**[2.85]**	**2.79**	—	**101.29**	**43.54**	**3**

Labeling of the Phe residues corresponding to each other in hTBPc, yTBPc and a TBPc

hTBPc	yTBPc	aTBPc
193F	99F	57F
210F	116F	74F
284F	190F	148F
301F	207F	165F

IDRD: residue ID; RES: residue name; VATM: atom involved; DATM: D_{atm}; DPLN: D_{pln}; DLIN: D_{lin}; omega: ω; theta: θ; RG: region.

1.9 Å), yTBP (1YTF, 2.5 Å) and aTBP (1QNE, 1.9 Å) bound to the target DNA, showing intramolecular interactions in the proteins. Only phenylalanine residues interacting with TATA-box are listed. It is well known that the TBP/TATA-box minor groove interface is dominated by nonpolar interactions. In particular, two pairs of the phenylalanine residues (in hTBP, Phe284 and Phe301 on one hand, Phe193 and Phe210 on the other) are reported to kink the helix by wedging into the outermost two base pairs of the TATA element.

CH/π interactions at the TBP/DNA boundary. Table 2 gives parts of the CHPI analysis of hTBP, yTBP and aTBP in complex with the target DNA, showing interactions at the protein/DNA boundary. Figure 2 is a global view of the human TBPc in complex with adenovirus major late promoter (AdMLP).

In Figure 2, we see every side-chain group of the above phenylalanines to be CH/π-bonded to the TATA element. In Tables 1 and 2, interactions involving phenylalanines conserved among these three species are intensified. Figure 3 is a diagrammatic illustration of the CH/π networks in the human TBPc in complex with the promoter DNA (Table 1(a) and Table 2(a)). Figure 4 gives stereo views of the interactions.

Table 2. Intermolecular CH/π interactions in hTBP, yTBP and aTBP complexes. CH/π interactions conserved among the 3 species are printed in bold

IDRD	RES	VATM	IDRD	RES	VATM	DATM (Å)
(a) hTBP complexed with CTGCTATAAAAGGCTG/CAGCCTTTTATAGCAG (1CDW, complex ABC). Sequence of the base-pair oligonucleotide bound to TBP is shown						
		0 1 2 3 4 5 6 7 8 9				
		5′-C-T-G- C - T - A - T - A - A - A - A - G - G -C-T-G-3′				
		3′-G-A-C- G - A - T - A - T - T - T - T - C - C -G-A-C-5′				
		0′ 1′ 2′ 3′ 4′ 5′ 6′ 7′ 8′ 9′				
2	**A**	**C2**	**284**	**PHE**	**HZ**	**2.99**
7	A	N3	210	PHE	HZ	2.85
1′	**A**	**C2**	**284**	**PHE**	**HE1**	**2.83**
193	PHE	CG	7′	T	H4′	3.12[a]
210	**PHE**	**CD1**	**7**	**A**	**H2′1**	**2.69**
210	**PHE**	**CE1**	**7**	**A**	**H1′**	**2.66**
210	PHE	CG	8	G	H5′1	3.06
301	PHE	CE2	2′	T	H2′1	2.89
301	**PHE**	**CE1**	**2′**	**T**	**H1′**	**2.51**
301	**PHE**	**CD2**	**1′**	**A**	**H5′1**	**2.76**
(b) yTBP complexed with TGTATGTATATAAAAC/GTTTTATATACATACA (1YTF, complex AEF). Sequence of the base-pair oligonucleotide bound to TBP is shown						
		1 2 3 4 5 6 7 8				
		5′-A- T - A - T - A - A - A - A - C -3′				
		3′-T- A - T - A - T - T - T - T - G -5′				
		1′ 2′ 3′ 4′ 5′ 6′ 7′ 8′				
2	**A**	**C2**	**A190**	**PHE**	**HZ**	**3.02**
3′	A	N3	A213	VAL	HG11	2.80
1′	**A**	**C2**	**A190**	**PHE**	**HE1**	**2.73**
A99	PHE	CE2	7′	T	H1′	3.16
A116	**PHE**	**CE2**	**7**	**A**	**H2′1**	**3.08**
A116	**PHE**	**CZ**	**7**	**A**	**H1′**	**2.65**
A207	**PHE**	**CE2**	**2′**	**T**	**H1′**	**2.53**
A207	**PHE**	**CG**	**1′**	**A**	**H5′1**	**2.87**
(c) aTBP complexed with GCTATAAAAGGGCA/TGCCCTTTTATAGC (1QNE, complex ACD). Sequence of the base-pair oligonucleotide bound to TBP is shown						
		1 2 3 4 5 6 7 8				
		5′-C- T - A - T - A - A - A - A - G -G-3′				
		3′-G- A - T - A - T - T - T - T - C -C-5′				
		1′ 2′ 3′ 4′ 5′ 6′ 7′ 8′				
2	**A**	**C2**	**A148**	**PHE**	**HZ**	**3.06**
3	T	C2	A163	LEU	HD11	3.02
6	A	N3	A29	VAL	HG11	2.86
7	A	C2	A57	PHE	HZ	2.96
1′	**A**	**N1**	**A148**	**PHE**	**HE1**	**2.79**
A74	**PHE**	**CE1**	**7**	**A**	**H1′**	**2.52**
A74	**PHE**	**CD1**	**7**	**A**	**H2′2**	**2.64**
A74	PHE	CG	8	G	H5′1	2.72
A165	**PHE**	**CE1**	**2′**	**T**	**H1′**	**2.56**
A165	PHE	CE1	2′	T	H2′2	3.09
A165	**PHE**	**CD2**	**1′**	**A**	**H5′1**	**2.73**

IDRD: residue ID; RES: residue name; VATM: atom involved; DATM: D_{atm}. [a]Phe193 (H/C distance 3.12 Å) is included in Figures 2–4.

At the opening base step T1/A1′–A2/T2′, Phe284 and Phe301 are involved in the interaction with nucleic acid bases. Pro285, Gly286 (NH/π, CH/π) and Leu299 are also involved in the network. Phe193 and Phe210 are involved in the other network at the base step A7/T7′–G8/C8′, whereby Ala194 (NH/π) and Ala195 (NH/π, CH/π) are also contributing to stabilize the network. In Phe284, Hζ and Hε of the side-chain group interact with the aromatic ring of A2 and A1′ bases, respectively, while the aromatic part of Phe301 (Cδ, Cε) interacts with hydrogens in the deoxyribose moiety of A1′ and T2′. Also, the aromatic ring of Phe193 (Cγ) is close to T7′. The Phe210 side-chain plays both parts: H-donor (Hζ/A7N3) as well as π-donor (Cδ/A7H2′, Cε/A7H1′ and Cγ/G8H5′).

In other complexes also, the 4 phenylalanine side-chains are CH/π-bonded to DNA (Tables 2(b) and 2(c)). Notice that most CH/π interactions involving the four Phe residues are conserved among these three species.

CH/π interactions in the TATA elements. Table 3 lists CH/π contacts unveiled in the TATA element.

A number of CH/π contacts have been found within the same strand (Table 3). The interactions are shown between nucleotides adjacent to each other: C0/T1 (2 contacts), A2/T3, A4/T3, A5/A4, A6/A5, A7/A6, G9/G8 (Table 3(a), Fig. 5(a)), C8′/T7′, T7′/T6′, T6′/T5′ (2 contacts), T4′/T5′, A3′/T4′ (2 contacts) and G0′/A1′ (Table 3(a), Fig. 5(b)). They are formed largely between the aromatic ring of the bases and the hydrogens (H2′) of the sugar moiety. Further, it is remarkable that the intra-strand CH/π interactions involve the methyl group of thymine: C0(C^6)/T1CH_3, A2(N^9)/T3CH_3, C8′(C^6)/T7′CH_3, T7′(C^5)/T6′ CH_3 and T6′(C^6)/T5′CH_3.

Predominance of the CH/π interactions involving deoxyribose H2′ and thymine methyl group has been shown

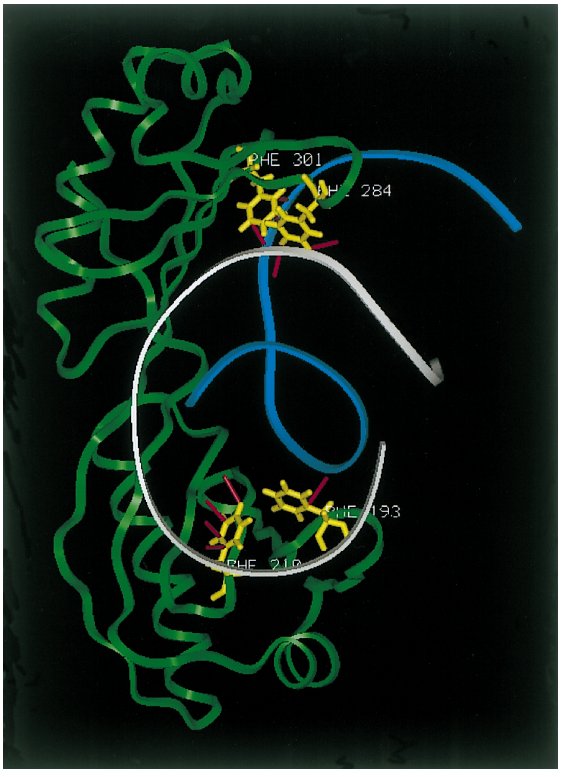

Figure 2. Global view of hTBPc/TATA-box (AdMLP) complex illustrating the CH/π interactions: hTBPc in green, TATA element in white and blue ribbons. Red sticks indicate intermolecular contacts involving the 4 phenylalanine residues (yellow).

Table 3. Intra-strand interaction in the TATA-box. Only interactions involving the central base-pair oligonucleotide are shown. Asterisks indicate the positions where the phenylalanine residues are inserted. Interactions involving thymine methyl group are shown in bold

IDRD	RES	VATM	IDRD	RES	VATM	DATM (Å)
(a) Human (1CDW). Adenovirus major late promoter AdMLP: 5′-CT*ATAAAA*GG-3′ (strand B: 0–9)/3′-GA*TATTTT*CC-5′ (strand C: 0′–9′)						
0	**C**	**C6**	**1**	**T**	**H5M2**	**2.79**
1	T	C6	0	C	H2′2	2.55
2	**A**	**N9**	**3**	**T**	**H5M3**	**3.04**
4	A	C8	3	T	H2′2	2.80
5	A	C8	4	A	H2′2	3.01
6	A	C8	5	A	H2′2	2.71
7	A	C8	6	A	H2′2	2.74
9	G	C8	8	G	H1′	2.92
8′	**C**	**C6**	**7′**	**T**	**H5M1**	**2.59**
7′	**T**	**C5**	**6′**	**T**	**H5M1**	**2.93**
6′	**T**	**C6**	**5′**	**T**	**H5M2**	**3.02**
5′	T	C6	6′	T	H2′2	2.93
4′	T	C6	5′	T	H2′2	2.93
3′	A	N9	4′	T	H2′1	2.92
3′	A	N7	4′	T	H2′2	2.79
0′	G	C8	1′	A	H1′	3.00
(b) Yeast (1YTF). 5′-AT*ATAAAA*C-3′ (strand E: 0–8)/3′-TA*TATTTT*G-5′ (strand F: 0′–8′)						
0	**A**	**N7**	**1**	**T**	**H5M3**	**3.00**
2	**A**	**C5**	**3**	**T**	**H5M3**	**3.13**
4	A	N7	3	T	H2′2	2.77
5	A	N9	4	A	H2′2	2.98
6	A	C8	5	A	H2′2	2.62
7	A	C8	6	A	H2′2	2.76
7	A	N7	8	C	H5	2.82
8′	**G**	**N7**	**7′**	**T**	**H5M1**	**2.85**
8′	**G**	**C8**	**7′**	**T**	**H5M2**	**2.75**
7′	**T**	**C5**	**6′**	**T**	**H5M2**	**2.93**
6′	**T**	**C6**	**5′**	**T**	**H5M2**	**2.86**
5′	T	C6	6′	T	H2′1	2.84
3′	A	N9	4′	T	H2′1	2.93
3′	A	C8	4′	T	H2′2	2.84
1′	**A**	**C8**	**0′**	**T**	**H5M2**	**2.98**
(c) Plant (1QNE). 5′-CT*ATAAAA*GG-3′ (strand C: 0–9)/3′-GC*TATTTT*CC-5′ (strand D: 0′–9′)						
1	A	C5	0	C	H2′2	2.68
2	**A**	**C8**	**3**	**T**	**H5M1**	**2.89**
4	A	C8	3	T	H2′1	2.96
5	A	N9	4	A	H2′1	2.92
6	A	C8	5	A	H2′1	2.64
7	A	C8	6	A	H2′1	2.58
9	G	C8	8	G	H1′	2.88
8′	**C**	**C6**	**7′**	**T**	**H5M2**	**2.49**
7′	**T**	**C5**	**6′**	**T**	**H5M2**	**2.79**
6′	**T**	**C6**	**5′**	**T**	**H5M1**	**2.81**
4′	T	C6	5′	T	H2′1	2.85
3′	A	N7	4′	T	H2′1	2.67
3′	A	N9	4′	T	H2′2	2.86
0′	G	C8	1′	A	H1′	3.04

IDRD: residue ID; RES: residue name; VATM: atom involved; DATM: D_{atm}.

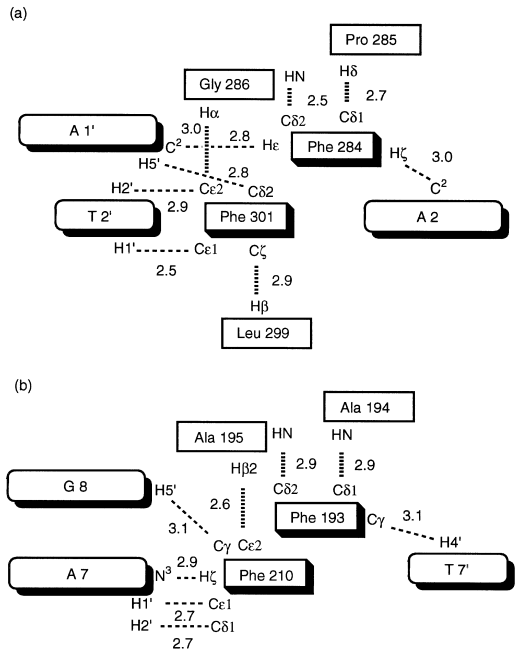

Figure 3. Schematic illustrations of the CH/π networks disclosed in the hTBPc/TATA-box complex. (a) At base step T1/A1′–A2/T2′ and (b) A7/T7′–G8/C8′. Dotted lines and hashed lines indicate intermolecular and intramolecular CH/π contacts, respectively. Numbers refer to the distances (Å) between atoms.

for the other TATA elements: yeast DNA (Table 3(b)) as well as *Arabidopsis thaliana* DNA (Table 3(c)). Involvement of H2′ and thymine methyl group is not limited to the binding region of the promoter. For instance, we see a network of such interactions in an archaeal promoter sequence (1D3U, 2.4 Å, Table 4). Figure 6 is a part of the network including thymine nucleotides.

Discussion

To summarize, a number of short CH/π distances have been shown in TBP/TATA-box complexes. It seems reasonable to suggest that an appreciable portion of the nonpolar interactions reported at the boundary of TBP and the minor groove of TATA-box is attributed to a more specific molecular force, the CH/π interaction. Furthermore, many intra-strand CH/π contacts have been disclosed within the TATA-box. Nishinaka et al. reported on the importance of this type of interaction (H2′ versus adjacent nucleotide base) in stabilizing the structure of DNA, which is unwound on binding to RecA protein.[20] Chou et al. suggested a role of the CH/π interaction (H2′ and/or H4′ versus nucleotide base) in stabilizing the unusual hairpin DNA structure of human centromeres and telomeres.[21] In the present study, hydrogens of deoxyribose (mostly H2′) play an important

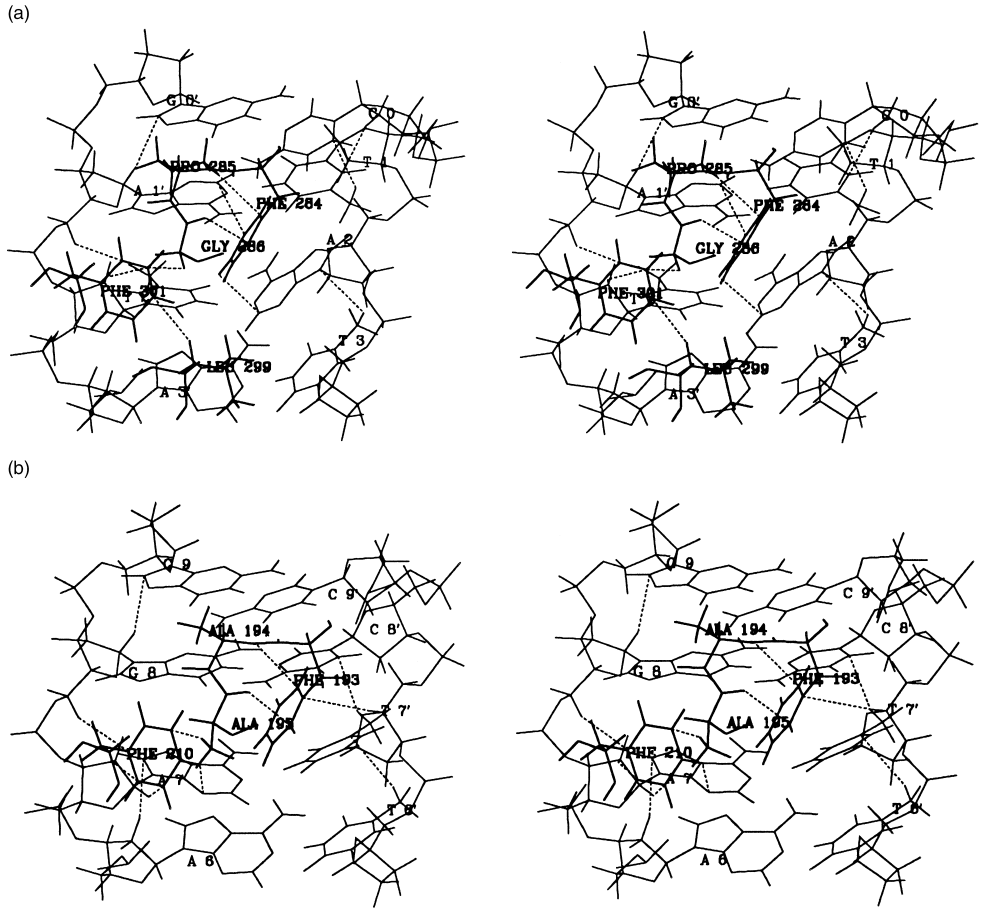

Figure 4. Stereo diagrams of the first and the last base-pair steps in the TATA-box showing CH/π interactions (1CDW). (a) Minor groove view into the cleft of step C0/G0′–T3/A3′. (b) Minor groove view of the base step A6/T6′–G9/C9′. Thick lines: protein; thin lines: DNA. Dotted lines indicate CH/π contacts. Intermolecular as well as intramolecular interactions are shown.

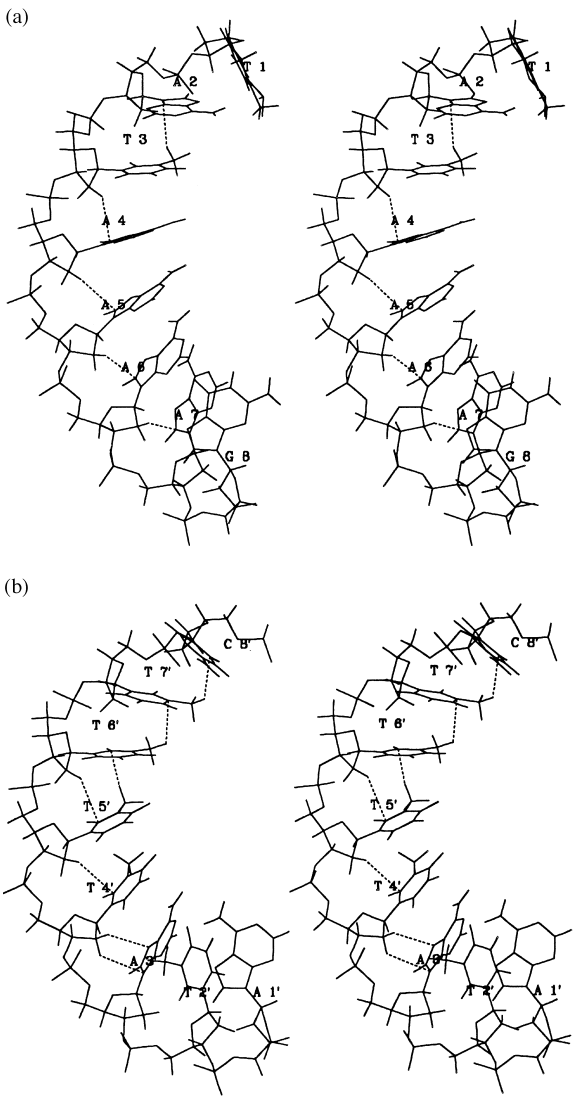

Figure 5. Intra-strand CH/π interactions in a TATA element (1CDW). (a) 5′-TATAAAAG-3′ (strand B: 1–8); (b) 3′-ATATTTTC-5′ (strand C: 1′–8′). Dotted lines indicate CH/π contacts (stereo view).

Table 4. Intra-strand CH/π interactions in the TATA element of *Pyrococcus woesei* (1D3U, complex ACD). 5′-AGAGTAAAGT*T-TAAAT*ACTTATAT-3′ (strand C: 1401–1424)/3′-TCTCATTT-CA*AATTTA*TGAATAT-5′ (strand D: 1448–1426). Asterisks indicate the positions where the phenylalanine residues are inserted. Residues depicted in Figure 6 are underlined

IDRD	RES	VATM	IDRD	RES	VATM	DATM (Å)
C1403	A	C8	C1402	G	H1′	2.96
C1404	G	C8	C1405	T	H5M1	2.86
C1404	G	N7	C1405	T	H5M3	2.79
C1409	G	N7	C1410	T	H5M2	2.57
C1410	T	C5	C1411	T	H5M1	2.98
C1411	T	C5	C1412	T	H5M1	2.95
C1412	T	C6	C1411	T	H2′2	2.72
C1414	A	C8	C1413	A	H2′2	2.95
C1415	A	C8	C1414	A	H2′1	2.76
C1415	A	N7	C1414	A	H2′2	2.57
C1416	T	C5	C1415	A	H2′2	2.70
C1417	A	C8	C1416	T	H2′2	2.73
C1418	C	C5	C1419	T	H5M2	2.92
C1418	C	C6	C1419	T	H5M3	2.60
C1419	T	C6	C1420	T	H5M1	2.73
C1421	A	N7	C1422	T	H5M1	2.80
C1421	A	C8	C1422	T	H5M2	2.86
C1423	A	C8	C1422	T	H1′	2.91
C1423	A	C8	C1424	T	H5M1	2.96
D1427	A	C8	D1426	T	H2′1	2.35
D1427	A	N7	D1426	T	H2′2	2.49
D1427	A	C8	D1428	T	H5M1	2.91
D1427	A	C8	D1428	T	H5M3	2.84
D1431	G	C8	D1432	T	H5M3	2.55
D1432	T	C4	D1433	A	H62	2.83
D1433	A	C8	D1434	T	H5M3	2.99
D1435	T	C6	D1434	T	H2′1	2.74
D1437	A	N9	D1436	T	H2′1	2.92
D1437	A	N7	D1436	A	H2′2	2.93
D1438	A	C8	D1437	A	H2′2	2.63
D1439	A	C8	D1438	A	H2′2	3.05
D1440	C	C6	D1439	A	H1′	2.96
D1440	C	C6	D1441	T	H5M2	2.75
D1441	T	C6	D1442	T	H5M2	2.72
D1442	T	C6	D1443	T	H5M2	3.06
D1444	A	C8	D1443	T	H2′1	2.95
D1445	C	C6	D1446	T	H5M3	2.61
D1447	C	N1	D1448	T	H5M1	3.01
D1447	C	C5	D1448	T	H5M3	3.05

IDRD: residue ID; RES: residue name; VATM: atom involved; DATM: D_{atm}.

role in stabilizing the structure of TATA elements.[22] The most interesting, among others, is that the methyl group in thymine is involved. This type of interaction seems to play a role in the stabilization of unwound DNA, which is destabilized by loosening the canonical B-DNA structure. Interactions involving the thymine *methyl group* may well be a chemical basis why every promoter DNA has a T-rich sequence. It is fascinating to speculate that for this reason nature chose thymine but not uracil as the partner of adenine in the duplex DNA.

Knowledge of weak intermolecular forces is crucial in structural biology. We previously demonstrated that the CH/π interaction is important in stabilizing the 3-D structure of proteins and their complexes with specific substrates. The present finding that CH/π interactions are ubiquitously found in the crystal structure of TBP/promoter complexes supports our thesis that the CH/π interaction is indispensable in deeper understanding of molecular biology. According to theory[23] and experimental data, we believe that the CH/π interaction is originated by charge transfer process from the π-system to the antibonding orbital of the C−H bond.[24] Also important is the contribution from the dispersion mechanism.[8,25] The Coulomb force, though not very significant, also contributes.[8,26] We conclude that a substantial part of the interactions, broadly attributed in the past to nonspecific molecular interactions such as the van der Waals force or the so-called 'hydrophobic effect',[27] should now be re-examined in the context of a new paradigm.[28]

Acknowledgements

We thank Dr. Ichiro Hirao (ERATO, JST) for useful comments, Drs. Tomio Takeuchi and Shinichi Kondo (Institute of Microbial Chemistry) for encouragement.

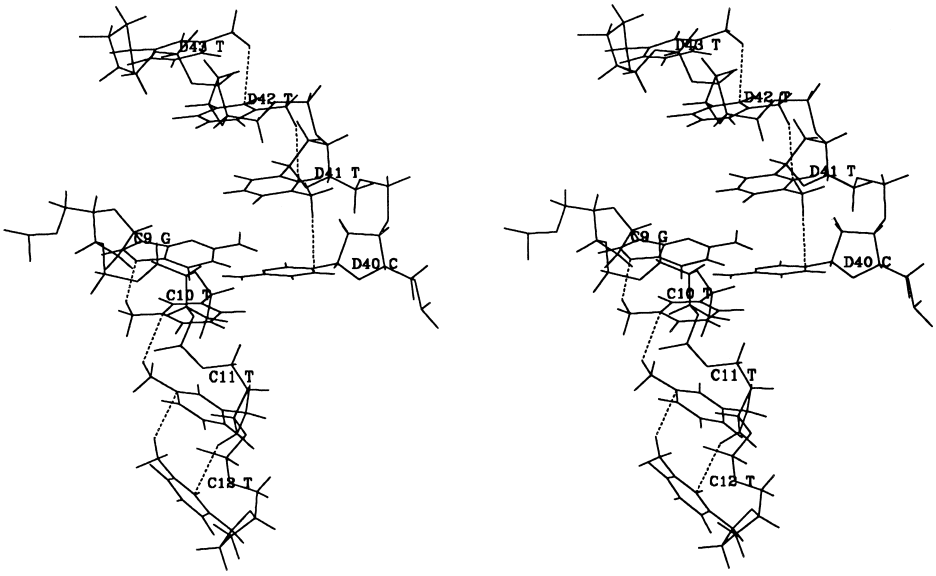

Figure 6. Intra-strand CH/π interactions in a promoter sequence (1D3U, strand C: 1409–1412; strand D: 1440–1443 in PDB file). Dotted lines indicate CH/π contacts (stereo view).

References and Notes

1. Nishio, M.; Hirota, M.; Umezawa, Y. *The CH/π Interaction. Evidence, Nature, and Consequences*. Wiley-VCH: New York, 1998.
2. Nishio, M. *Kagaku no Ryoiki* **1977**, *31*, 998; **1979**, *33*, 422; *29th Symposium on Protein Structures*, Osaka, 1978, Abstract, 161; Kodama, Y.; Nishihata, K.; Zushi, S.; Nishio, M.; Uzawa, J.; Sakamoto, K.; Iwamura, H. *Bull. Chem. Soc. Jpn.* **1979**, *52*, 2661.
3. Nishio, M.; Umezawa, Y.; Hirota, M.; Takeuchi, Y. *Tetrahedron* **1995**, *51*, 8665.
4. O'Reilly, M.; Watson, K. A.; Schinzel, R.; Palm, D.; Johnson, L. N. *Nat. Struct. Biol.* **1997**, *5*, 405; Harata, K.; Muraki, M. *Acta Crystallogr. Sect. D* **1997**, *53*, 650; Zographos, S. E.; Oikonomakos, N. G.; Tsitsanou, K. E.; Leonidas, D. D.; Chrysina, E. D.; Skamnaki, V. T.; Bischoff, H.; Goldmann, S.; Watson, K. A.; Johnson, L. N. *Structure* **1997**, *5*, 1413; Kashima, A.; Inoue, Y.; Sugio, S.; Maeda, I.; Nose, T.; Shimohigashi, Y. *Eur. J. Biochem.* **1998**, *255*, 12; Jabs, A.; Weiss, M. S.; Hilgenfeld, R. *J. Mol. Biol.* **1999**, *286*, 291; Pal, D.; Chakrabarti, P. *J. Mol. Biol.* **1999**, *294*, 271; Samanta, U.; Pal, D.; Chakrabarti, P. *Proteins: Structure, Function, and Genetics* **2000**, *38*, 288; Muraki, M.; Harata, K.; Sugita, N.; Sato, I. *Biochemistry* **2000**, *39*, 292.
5. Kim, D.-H.; Li, Z.-H.; Soo, S. L.; Park, J.; Chung, S. J. *Bioorg. Med. Chem.* **1998**, *6*, 239; Nose, T., Fujita, T., Nakajima, M., Inoue, Y., Costa, T., Shimohigashi, Y. *J. Biochem.* **1998**, *124*, 354; Sato, T.; Kanai, Y.; Hoshino, T. *Biosci. Biotech. Biochem.* **1998**, *62*, 407; Shimohigashi, Y.; Nose, T.; Yamauchi, Y.; Maeda, I. *Biopolymers* **1999**, *51*, 9; Schoepfer, J.; Fretz, H.; Gay, B.; Furet, P.; Garcia-Echeverria, C.; End, N.; Caravatti, G. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 221; Barreca, M. L.; Carotti, A.; Carrieri, A.; Chimirri, A.; Monforte, A. M.; Pellegrini-Calace, M.; Rao, A. *Bioorg. Med. Chem.* **1999**, *7*, 2283.
6. (a) Umezawa, Y.; Nishio, M. *Bioorg. Med. Chem.* **1998**, *6*, 493; (b) Umezawa, Y.; Nishio, M. *Bioorg. Med. Chem.* **1998**, *6*, 2507.
7. Fan, M.-F.; Lin, Z.; McGrady, J. E.; Mingos, D. M. P. *J. Chem. Soc., Perkin Trans. 2* **1996**, 563; Samanta, U.; Chakrabarti, P.; Chandrasekhar, J. *J. Phys. Chem. A* **1998**, *102*, 8964; Novoa, J. J.; Mota, F. *Chem. Phys. Lett.* **2000**, 318, 345.
8. Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Phys. Chem. A* **1999**, *103*, 8265; idem. *J. Am. Chem. Soc.* **2000**, *122*, 3746.
9. Pauling, L. *The Nature of the Chemical Bonds*; Cornell Univ. Press: Ithaca, New York, 1960, p. 260.
10. Wlodawer, A.; Walter, J.; Huber, R.; Sjölin, L. *J. Mol. Biol.* **1984**, *180*, 301.
11. Wlodawer, A.; Sjölin, L. *Biochemistry* **1983**, *22*, 2720.
12. The CVFF force field within the modeling package DISCOVER (MSI, San Diego, CA) was used.
13. Ref 1, pp 178–181; Tables 11.1 and 11.2.
14. Ref 6a, Tables 1 and 2.
15. Umezawa, Y.; Tsuboyama, S.; Takahashi, H.; Uzawa, J.; Nishio, M. *Bioorg. Med. Chem.* **1999**, *7*, 2021.
16. Nikolov, D. B.; Chen, H.; Haley, E. D.; Hoffman, A.; Roeder, R. G.; Burley, S. K. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 4862 (PDB code 1CDW, 1.9 Å); Juo, Z.-S.; Chiu, T.-K.; Leiberman, P. M.; Baikalov, I.; Berk, A. J.; Dickerson, R. E. *J. Mol. Biol.* **1996**, *261*, 239 (1TGH, 2.9 Å); Tsai, F. T. F.; Sigler, P. B. *EMBO J.* **2000**, *19*, 25 (1C9B, 2.65 Å).
17. Kim, Y.; Gieger, J. H.; Hahn, S.; Sigler, P. B. *Nature* **1993**, *365*, 512 (1YTB, 1.9 Å); Tan, S.; Hunziker, Y.; Sargent, D. F.; Richmond, T. J. (1YTF, 2.5 Å, not published).
18. Kim, J. L.; Nikolov, D. M.; Burley, S. K. *Nature* **1993**, *365*, 520 (2.25 Å); Nikolov, D. B.; Chen, H.; Haley, E. D.; Usheva, A. A.; Hisatake, K.; Lee, D.-K.; Roeder, R. G.; Burley, S. K. *Nature* **1995**, *377*, 119 (1VOL, 2.7 Å); Kim, J. L.; Burley, S. K. *Nat. Struct. Biol.* **1994**, *1*, 638 (1QNE, 1.9 Å).
19. Kosa, P. F.; Ghosh, G.; DeDecker, B. S.; Sigler, P. B. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 6042 (1AIS, 2.1 Å); Littlefield, O.; Korkhin, Y.; Sigler, P. B. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13668 (1D3U, 2.4 Å).
20. Nishinaka, T.; Ito, Y.; Yokoyama, S.; Shibata, T. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 6623; Nishinaka, T.; Shinohara, A.; Ito, Y.; Yokoyama, S.; Shibata, T. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 11071.
21. Chou, S.-H; Tseng, Y.-Y. *J. Mol. Biol.* **1999**, *285*, 41. See also Hirao, I.; Kawai, G.; Yoshizawa, S.; Nishimura, Y.; Ishido, Y.; Watanabe, K.; Miura, K. *Nucleic Acids Res.* **1994**, *22*, 576; Zhu, L.; Chou, S.-H.; Reid, B. R. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 12159; Yoshizawa, S.; Kawai, G.; Watanabe, K.; Miura, K.; Hirao, I. *Biochemistry* **1997**, *36*, 4761; Chou, S.-H; Tseng, Y.-Y.; Wang, S.-W. *J. Mol. Biol.* **1999**, *287*, 301.

22. This may explain why deoxyribose is used as the sugar part of DNA. In the sugar moiety of RNA (ribose), H2′2 (more prone to interact with a sp^2 carbon) is replaced by OH.
23. Ref 1, Chapter 3; Nishio, M.; Hirota, M. *Tetrahedron* **1989**, *45*, 7201.
24. Takagi, T.; Tanaka, A.; Matsuo, S.; Maezaki, H.; Tani, M.; Fujiwara, H.; Sasaki, Y. *J. Chem. Soc., Perkin Trans. 2* **1987**, 1015; Nakai, Y.; Inoue, K.; Yamamoto, G.; Oki, M. *Bull. Chem. Soc. Jpn.* **1989**, *62*, 2923; Hirota, M.; Sakakibara, K.; Suezawa, H.; Yuzuri, T.; Ankai, E.; Nishio, M. *J. Phys. Org. Chem.* **2000**, in press.
25. Hirota, M.; Sekiya, T.; Abe, K.; Tashiro, H.; Karatsu, M.; Nishio, M.; Osawa, E. *Tetrahedron* **1983**, *39*, 3091; Nishio, M.; Hirota, M. *Tetrahedron* **1989**, *45*, 7201.
26. Umezawa, Y.; Tsuboyama, S.; Honda, K.; Uzawa, J.; Nishio, M. *Bull. Chem. Soc. Jpn.* **1998**, *71*, 1207.
27. The concept of hydrophobic effect is a subject of controversy among physical chemists. See, e.g., Shinoda, K. *Kagaku to Kogyo* **1967**, *21*, 1400; Hildebrand, J. H. *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 194; Privalov, P. L.; Gill, S. J. *Adv. Protein Chem.* **1988**, *39*, 191; Murphy, K. P.; Privalov, P. L.; Gill, S. J. *Science* **1990**, *247*, 559; Dill, K. *Science* **1990**, *250*, 297; Privalov, P. L.; Gill, S. J.; Murphy, K. P. *Science* **1990**, *250*, 297.
28. Nishio, M. *Introduction to Intermolecular Forces in Organic Chemistry*; Kodansha: Tokyo, 2000 (in Japanese).