A Protein That Binds to a cis-Acting Element of Wheat Histone Genes Has a Leucine Zipper Motif

TETSUYA TABATA, HISABUMI TAKASE, SHINOBU TAKAYAMA, KOJI MIKAMI, ATSUNORI NAKATSUKA, TAKEFUMI KAWATA, TAKUYA NAKAYAMA, MASAKI IWABUCHI*

The structure and function of transcription factors of higher plants was studied by isolating cDNA clones encoding a wheat sequence-specific DNA binding protein. A hexameric nucleotide motif, ACGTCA, is located upstream from the TATA box of several plant histone genes. It has been suggested that this motif is essential for efficient transcription of the wheat histone H3 gene. A wheat nuclear protein, HBP-1 (histone DNA binding protein–1), which specifically binds to the hexameric motif, has previously been identified as a putative transcription factor. A cDNA clone encoding HBP-1 has been isolated on the basis of specific binding of HBP-1 to the hexameric motif. The deduced amino acid sequence indicates that HBP-1 contains the leucine zipper motif, which represents a characteristic property of several eukaryotic transcription factors.

TRANSCRIPTIONAL REGULATION DEpends on the sequence-specific interaction of trans-acting protein factors with cis-acting DNA elements. Our understanding of the structure and function of trans-acting factors through analysis of their cloned genes (*1–4*) has been restricted to factors from animals and yeast, with little being known about those of plants.

We have shown that HBP-1 binds to the hexameric motif ACGTCA (*5*), which is conserved in several plant histone genes (*6*) and is thought to be a cis-acting element for efficient transcription of the wheat histone H3 gene (*7*); HBP-1 may therefore be a trans-acting factor. To study the structure and function of HBP-1, we attempted to isolate its cDNA clone from a wheat cDNA library in λgt11. We made use of the ability of HBP-1 to bind to the hexameric motif to screen a cDNA expression library, using a technique developed by Singh *et al.* (*8*) and Vinson *et al.* (*9*) with slight modifications (*10*). We constructed a cDNA library in λgt11 using polyadenylated [poly(A)$^+$] mRNA from wheat seedlings. This library was screened with a ^{32}P-labeled catenated fragment of the synthetic 30-bp oligonucleotide that contains the hexameric motif and surrounding sequences of the wheat H3 gene (Fig. 1A). From a primary screen of 250,000 phage plaques, two phages that bound to the probe were isolated after three rounds of screening. For one of the two phages, termed λ16, we determined the specificity of the DNA binding protein encoded by the recombinant phage in extracts of an induced phage lysogen. Lysogen extracts were subjected to SDS–polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to nitrocellulose filters. After a denaturation-renaturation procedure, the filters were treated with the ^{32}P-labeled hexamer-containing probe. A single fusion protein band of ~140 kD was detected in the lysogen extract from λ16, whereas no such band was found in a lysogen extract from nonrecombinant λgt11 (Fig. 1B).

We performed a mobility-shift DNA binding assay to confirm the sequence specificity of the binding of the fusion protein. Extracts derived from the λ16 and λgt11 lysogens were examined with an H3 probe (Fig. 1C). Two major retarded bands were detected specifically with the extract of the λ16 lysogen. The formation of the two bands may be due to β-galactosidase interactions or oligomeric contacts in HBP-1, because only one major band was detected by protein blot analysis with a DNA probe. A 20-bp oligonucleotide containing the hexameric motif efficiently competed with the probe for binding; 20-bp oligonucleotides that contained either a mutated hexameric motif or an octameric motif that is conserved in many histone H3 and H4 genes of higher plants (*11*) did not compete. These competition profiles are the same as those obtained with native wheat nuclear HBP-1 (*5*).

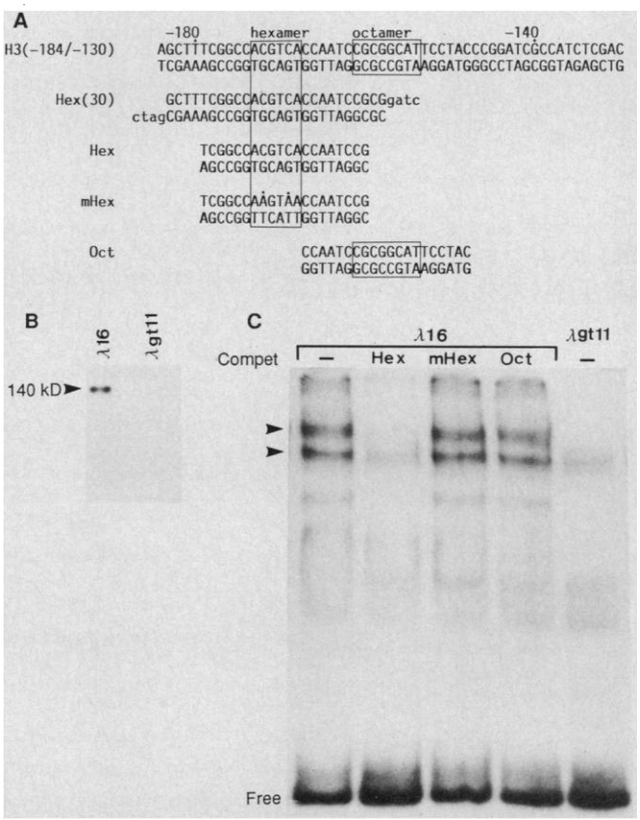

Fig. 1. Specificity of the hexamer binding activity of λ16-encoded protein. (**A**) DNA sequences of probes and competitors derived from the wheat histone H3 gene (clone pTH012) (*14*). Numbers refer to the nucleotide positions relative to the start site of transcription. The following were used: H3(−184/−130), a Hind III–Fok I fragment of the H3 gene; Hex(30), a 30-nucleotide synthetic oligomer used as a probe for cDNA screening and protein blotting; Hex, a synthetic 20-bp oligonucleotide containing the hexameric motif that was used as a competitor in the mobility-shift DNA binding assay; mHex, the same as Hex except for two point mutations indicated by dots; Oct, a synthetic 20-bp oligonucleotide containing the octameric motif. (**B**) Blot analyses of the proteins extracted from the induced lysogens of λ16 and of λgt11. Bacterial lysogens were prepared (*8*), and each lysogen extract was separated by SDS-PAGE on a 7.5% gel and transfered to nitrocellulose filters. The filters were probed with the ^{32}P-labeled oligomers under the same condition as in the cDNA screening. Molecular mass is indicated on the left of the panel. (**C**) Mobility-shift DNA binding assays of the proteins extracted from the induced lysogens of λ16 and λgt11. Each lysogen extract was incubated with the probe H3(−184/−130) in the presence or absence of a molar excess (100 times) of the specific competitor as indicated and then analyzed on a native polyacrylamide gel. Two major retarded bands are indicated with arrowheads.

T. Tabata, S. Takayama, M. Iwabuchi, Laboratory of Plant Molecular Biology, Department of Botany, Faculty of Science, Kyoto University, Sakyo-ku, Kyoto 606, Japan.
H. Takase, K. Mikami, A. Nakatsuka, T. Kawata, T. Nakayama, Department of Botany, Faculty of Science, Hokkaido University, Sapporo 060, Japan.

*To whom correspondence should be addressed.

The binding specificity of the λ16-encoded protein was more rigorously compared with that of partially purified native wheat nuclear HBP-1 by a methylation interference assay. The interference pattern obtained with the lysogen extract was centered over the hexamer motif and was similar to that obtained with native HBP-1 (Fig. 2). These results indicate that the insert of λ16 encodes wheat nuclear HBP-1.

A DNA blot of wheat genomic DNA that had been digested with restriction enzymes was probed with the phage λ16 insert (Fig. 3A). Each digest showed a simple hybridization pattern, consistent with HBP-1 being a single-copy gene or one with a small number of copies. An RNA blot analysis of poly(A)$^+$ wheat RNA showed that HBP-1 mRNA seems to be of approximately the same size as that of 17*S* ribosomal RNA (rRNA) (1.8 kb) (Fig. 3B).

The nucleotide sequence of the cDNA insert was determined by the M13 dideoxy sequencing technique (Fig. 4). The deduced amino acid sequence indicated that λ16 encodes a 278–amino acid polypeptide (including an Eco RI linker sequence that had been used for construction of the cDNA library) in frame with the β-galactosidase of λgt11. Because the cDNA insert was about 800 bp shorter than the putative HBP-1 mRNA, the λ16 insert was used to isolate additional clones from the cDNA library (*12*). Of the 13 positive clones analyzed, the insert sequence of the longest clone (λ17) was 462 and 118 bp longer than that of λ16 in the 5′ and 3′ regions, respectively, and contained a poly(A)$^+$ tract at the 3′ end (Fig. 4). The 3′ untranslated region of the λ16 insert was also followed by 22 adenosines, indicating that the HBP-1 gene has at least two polyadenylation sites. No sequence differences were noted between overlapping regions of λ16 and λ17. The most proximal AUG codon (position 1) in frame with the fusion protein of λ16 is preceded by a stop codon (position −54). We tentatively as-

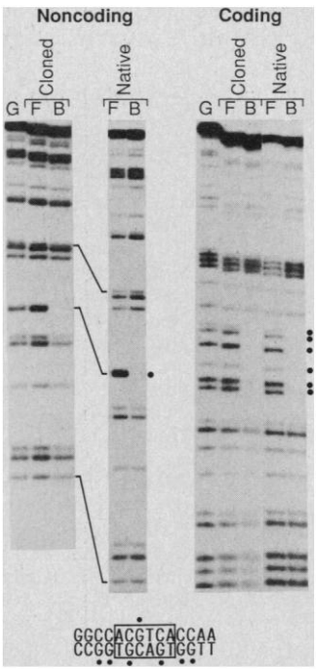

Fig. 2. Methylation interference analyses of λ16-encoded protein and native wheat nuclear HBP-1. Methylation interference analysis was performed with the probe H3(−184/−130), which had been labeled with ^{32}P on either the noncoding or coding strand with T4 polynucleotide kinase. Partially methylated probe was subjected to the binding reaction with either lysogen extract (cloned) or partially purified native wheat nuclear HBP-1 (*5*) (native), then subjected to electrophoresis on a native polyacrylamide gel. ^{32}P-labeled DNA was recovered from the free (F) and bound (B) bands, cleaved at sites of methylation, and analyzed on a sequencing gel. Lane G is a reference ladder generated by cleavage of the probe at the G residues. The positions of methylated guanine residues that interfere with binding are indicated by dots and depicted at the bottom.

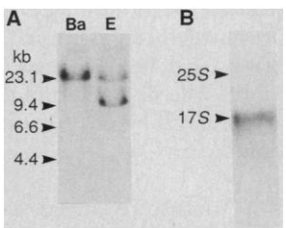

Fig. 3. DNA (**A**) and RNA (**B**) blot hybridization analyses with the λ16 cDNA. (A) Wheat genomic DNA (10 μg in each case) was digested with Bam HI (Ba) and Eco RI (E), and the fragments were separated by electrophoresis and transferred to nitrocellulose filters. The filters were probed with the ^{32}P-labeled λ16 insert in a hybridization solution containing 6× SSC (standard saline citrate), 50% formamide, 0.2% polyvinylpyrrolidone, 0.2% Ficoll, 0.2% bovine serum albumin, 0.5% SDS, sonicated *Escherichia coli* DNA (10 μg/ml), and poly(A)$^+$ RNA (30 μg/ml) at 37°C for 36 hours. The filters were then washed in 0.1× SSC at 65°C and autoradiographed. The positions of Hind III–digested λ standards are indicated. (B) Wheat poly(A)$^+$ RNA was resolved in a formaldehyde-containing 1.5% agarose gel and transferred to a nitrocellulose filter. The hybridization and washing conditions were as in (A). The positions corresponding to 17*S* (1.8 kb) and 25*S* (3.5 kb) rRNAs are indicated.

```
                                                    ⌐λ17
-242 GGAGTTCCCGCATTTGCTGGTCCCGGCGAGGTCCGCTCCTCTGCACGCGTGGGTGGCGGCGCCCTCGCCGGCGTCGCGTCGCGCCGGCCC

-152 CCGAATCTCCGGCGAGCCGCTCGCGCCCCAAGCCCGAGCCCGGTGAGGCCTACTGCTCTAGGTCTTCCTGCCACCCAGAGTTTTCGATAT

                                                        1
 -62 ACTGGTTTTGATATGGAGTGATTGGTGGAGCTGGTCGTCCGGAAGCTCTGCTGGTGAATCCGATGGGGAGCAACGATCCTAGCACGCCGT
                                                          M  G  S  N  D  P  S  T  P  S

  29 CTAAGGCTTCGAAGCCACCAGAACAGGAGCAACCTCCGGCTACTACCTCTGGCACCACAGCTCCAGTTTACCCTGAATGGCCCGGCTTTC
  11  K  A  S  K  P  P  E  Q  E  Q  P  P  A  T  T  S  G  T  T  A  P  V  Y  P  E  W  P  G  F  Q

 119 AGGGCTACCCAGCGATGCCACCGCATGGGTTTTTCCCCCCTCCTGTTGCTGCAGGCCAGGCTCATCCATACATGTGGGGACCTCAGCACA
  41  G  Y  P  A  M  P  P  H  G  F  F  P  P  P  V  A  A  G  Q  A  H  P  Y  M  W  G  P  Q  H  M

                                ⌐λ16
 209 TGGTGCCACCTTACGGGACACCACCACCCCCCTATATGATGTATCCACCAGGAACAGTATATGCCCATCCGACTGCTCCTGGTGTGCATC
  71  V  P  P  Y  G  T  P  P  P  P  Y  M  M  Y  P  P  G  T  V  Y  A  H  P  T  A  P  G  V  H  P

 299 CATTTCATTATCCTATGCAAACAAATGGAAATCTTGAACCTGCTGGAGCTCAGGGAGCTGCACCAGGTGCTGCAGAAACGAATGGGAAAA
 101  F  H  Y  P  M  Q  T  N  G  N  L  E  P  A  G  A  Q  G  A  A  P  G  A  A  E  T  N  G  K  N

 389 ATGAGCCCGGCAAAACATCTGGTCCATCTGCCAATGGGGTTACGTCCAACAGTGAAAGCGGAAGTGATAGTGAAAGTGAAGGAAGTGATG
 131  E  P  G  K  T  S  G  P  S  A  N  G  V  T  S  N  S  E  S  G  S  D  S  E  S  E  G  S  D  A

 479 CCAATTCTCAAAATGATTCACATTCAAAGGAGAATGATGTAAATGAAAATGGCAGTGCTCAGAATGGCGTATCTCATTCATCATCGCATG
 161  N  S  Q  N  D  S  H  S  K  E  N  D  V  N  E  N  G  S  A  Q  N  G  V  S  H  S  S  S  H  G

 569 GAACATTTAATAAACCCATGCCGTTGGTTCCAGTACAATCAGGTGCAGTGATAGGAGTTGCTGGTCCTGCGACAAACTTGAACATAGGGA
 191  T  F  N  K  P  M  P  L  V  P  V  Q  S  G  A  V  I  G  V  A  G  P  A  T  N  L  N  I  G  M

 659 TGGACTACTGGGGTGCAACTGGCTCTTCACCTGTTCCTGCAATGCGCGGCAAAGTACCGTCTGGTTCAGCTCGAGGAGAGCAATGGGATG
 221  D  Y  W  G  A  T  G  S  S  P  V  P  A  M  R  G  K  V  P  S  G  S  A  R  G  E  Q  W  D  E

 749 AAAGGGAACTGAAGAAGCAGAAGAGGAAGCTGTCCAACCGGGAATCAGCGCGCAGGTCCCGGCTGCGCAAGCAGGCTGAGTGTGAAGAGC
 251  R  E  L  K  K  Q  K  R  K  L  S  N  R  E  S  A  R  R  S  R  L  R  K  Q  A  E  C  E  E  L

                                  GCN4  N  T  E  A  A  R  R  S  R  A  R  K  L  Q  R  M  K  Q  L
                                       269
 839 TCGGGCAGCGCGCTGAGGCTTTGAAGTCAGAGAACTCGTCCCTCAGGATCGAGCTCGACCGGATCAAAAAGGAGTACGAGGAGCTCCTTT
 281  G  Q  R  A  E  A  L  K  S  E  N  S  S  L  R  I  E  L  D  R  L  K  K  E  Y  E  E  L  L  S

      E  D  K  V  E  E  L  L  S  K  N  Y  H  L  E  N  E  V  A  R  L  K  K
                                                                        310
 929 CGAAGAACACCTCTCTGAAGGCGAAGCTAGGGGAGTCCGGCGGCGGCGGCGGCAGCGATGCGGTCCCCGACATGAACGAACGGGGCGACA
 311  K  N  T  S  L  K  A  K  L  G  E  S  G  G  G  G  G  S  D  A  V  P  D  M  N  E  R  G  D  T

1019 CCAACGGCGGCAGCCACCAGAAGGAGCCCTGACTGACCGATCGACGACGCATGCGGCCTCTGCTACTCTACTTACATAGCTCGTGATGAT
 341  N  G  G  S  H  Q  K  E  P  End

                                                                              ⌐λ16→
1109 TGATGATGATGATTCTTGTAACAAACAAACTTGGCGACCCCGATCCGCCACACACTCACACAATGAGATGGTTCTTCTTTTTTGTGCCTC

1199 TTATCGACCTGATGACAATGGCGACGCCGTCGCCGCCACCTGCAGACTCCTGTGCTTGTTTCGTCCCTACTGTGGAAAGAGAGATCATTT

                                                                    λ17→
1289 GTGGGCGCGTGCACTTCTCAGTTCTCACCGCCAAAACTCAAAAAAAAA
```

Fig. 4. DNA sequence of the HBP-1 cDNA and the deduced amino acid sequence of the HBP-1 protein. Numbering of nucleotides and amino acids begins with the first in-frame AUG codon. Amino acid sequence is shown using single letters (*15*). A partial amino acid sequence of GCN4 (*16*) is also aligned with the corresponding region of HBP-1. Asterisks denote sequence identities. The basic motif is underlined. Periodically appearing leucines and a single isoleucine in the leucine zipper motifs are circled. The extents of the inserts λ16 and λ17 are indicated by arrows.

signed this as the initiation codon. Beginning at this putative initiation codon, the open reading frame apparently encodes a protein of 349 amino acids in length, with a predicted molecular mass of 36,700 daltons.

A computer search for related protein sequences revealed definite similarity between HBP-1 and GCN4 (*4*), a yeast transcription factor. The two proteins shared 19 identical residues within a sequence of 42 amino acids localized near their respective carboxyl termini (Fig. 4). This region contains two characteristic motifs, namely the basic motif and the leucine zipper motif, which are both found in other eukaryotic transcription factors such as C/EBP (*2*), GCN4 (*2*), CREB (*3*), and the proto-oncogene products Jun and Fos (*2*). The basic motif constitutes the DNA binding domain in GCN4 (*4*) and the Fos-Jun complex (*13*), suggesting that this motif functions as the DNA binding domain of HBP-1. In the typical leucine zipper motif, when the amino acid sequence is arranged to form an idealized α helix, a periodic repetition of leucine residues present at every seventh position over a distance of eight helical turns aligns along one side. A similar array of five leucines and one isoleucine was noted in the sequence of HBP-1. This motif is thought to represent a part of the scaffold that molds a protein to interact with its target site on DNA and allows dimerization of polypeptides that have this motif. Thus our finding that HBP-1, a putative trans-acting factor, appears to have the basic motif and the leucine zipper motif extends the range of evolutionary conservation of the structure of transcription factors.

REFERENCES AND NOTES

1. J. T. Kadonaga, K. R. Carner, F. R. Masiarz, R. Tjian, *Cell* **51**, 1079 (1987); T. Williams, A. Admon, B. Luscher, R. Tjian, *Genes Dev.* **2**, 1557 (1988); C. Santoro, N. Mermod, P. C. Andrews, R. Tjian, *Nature* **334**, 218 (1988); M. Ptashne, *ibid.* **335**, 683 (1988); G. Wiederrecht, D. Seto, C. S. Parker, *Cell* **54**, 841 (1988); L. M. Staudt *et al.*, *Science* **241**, 577 (1988); H.-C. Liou, M. R. Boothby, L. H. Glimcher, *ibid.* **242**, 69 (1988); D. K. Didier *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7322 (1988); M. Miyamoto *et al.*, *Cell* **54**, 903 (1988); M. Bodner *et al.*, *ibid.* **55**, 505 (1988); H. A. Ingraham *et al.*, *ibid.*, p. 519.
2. W. H. Landschulz, P. F. Johnson, S. L. McKnight, *Science* **240**, 1759 (1988).
3. J. P. Hoeffler, T. E. Meyer, Y. Yun, J. L. Jameson, J. F. Habener, *ibid.* **242**, 1430 (1988).
4. I. A. Hope and K. Struhl, *Cell* **46**, 885 (1986).
5. K. Mikami *et al.*, *FEBS Lett.* **223**, 273 (1987); K. Mikami *et al.*, *Plant Cell Physiol.* **30**, 107 (1989).
6. S.-C. Wu *et al.*, *Plant Mol. Biol.* **11**, 641 (1988).
7. T. Nakayama *et al.*, *Plant Cell Physiol.*, in press.
8. H. Singh, J. H. LeBowitz, A. S. Baldwin, Jr., P. A. Sharp, *Cell* **52**, 415 (1988).
9. C. R. Vinson *et al.*, *Genes Dev.* **2**, 801 (1988).
10. A wheat cDNA library was constructed from wheat seedling poly(A)$^+$ RNA with cDNA synthesis and λgt11 cloning kits (Amersham). Screening of cDNA expression was performed as described (*9*) except that the binding buffer contained 25 m*M* Hepes (*p*H 7.9), 100 m*M* KCl, 0.5 m*M* dithiothreitol, 0.25% powdered milk, poly(dI-dC) (10 μg/ml), and nick-translated–catenated probe (0.01 μg/ml) (made by ligating 30-bp oligonucleotides).
11. N. Chaubet *et al.*, *Plant Mol. Biol.* **6**, 253 (1986).
12. The cDNA library (2×10^6 clones) was screened by the plaque hybridization technique with the λ16 insert used as a probe, and 30 positive signals were identified. The hybridization and subsequent washing conditions were the same as those for the DNA and RNA blot hybridization analyses.
13. T. Kouzarides and E. Ziff, *Nature* **336**, 646 (1988); P. Sassone-Corsi, L. J. Ransone, W. W. Lamph, I. M. Verma, *ibid.*, p. 692.
14. T. Tabata, M. Fukasawa, M. Iwabuchi, *Mol. Gen. Genet.* **196**, 397 (1984).
15. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
16. A. G. Hinnebusch, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6442 (1984).
17. The SWISS-PROT database was searched for amino acid sequence homology by the use of a computer system belonging to Research Center for Molecular Genetics, Hokkaido University. This research was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, and Culture of Japan, and by grants from the Mitsubishi Foundation and the Research Council of the Ministry of Agriculture, Forestry, and Fisheries of Japan for original and creative research in biotechnology.

3 March 1989; accepted 10 July 1989

Recognition of Thymine·Adenine Base Pairs by Guanine in a Pyrimidine Triple Helix Motif

LINDA C. GRIFFIN AND PETER B. DERVAN*

Oligonucleotide recognition offers a powerful chemical approach for the sequence-specific binding of double-helical DNA. In the pyrimidine-Hoogsteen model, a binding size of >15 homopurine base pairs affords >30 discrete sequence-specific hydrogen bonds to duplex DNA. Because pyrimidine oligonucleotides limit triple helix formation to homopurine tracts, it is desirable to determine whether oligonucleotides can be used to bind all four base pairs of DNA. A general solution would allow targeting of oligonucleotides (or their analogs) to any given sequence in the human genome. A study of 20 base triplets reveals that the triple helix can be extended from homopurine to mixed sequences. Guanine contained within a pyrimidine oligonucleotide specifically recognizes thymine·adenine base pairs in duplex DNA. Such specificity allows binding at mixed sites in DNA from simian virus 40 and human immunodeficiency virus.

THE SEQUENCE-SPECIFIC RECOGNITION of double-helical DNA is essential for the regulation of cellular functions including transcription, replication, and cell division. The ability to design synthetic molecules that bind sequence-specifically to unique sites on human DNA could have major implications for the treatment of genetic, neoplastic, and viral diseases (*1, 2*). Pyrimidine oligodeoxyribonucleotides [15- to 18-nucleotide (nt) oligomers] bind homopurine sites within large double-stranded DNA by triple helix formation (*3–5*). Pyrimidine oligonucleotides bind in the major groove, parallel to the purine strand of Watson-Crick double-helical DNA (*3*). Specificity is due to Hoogsteen hydrogen bonding, wherein thymine (T) recognizes adenine·thymine (AT) base pairs (T·AT triplet) and protonated cytosine (C) recognizes guanine-cytosine (GC) base pairs (C+GC triplet) (*6–8*) (Fig. 1). In addition to length and sequence composition, the binding affinity and specificity of the pyrimidine oligonucleotide for duplex DNA is sensitive to *p*H, organic cosolvent, added cations, and temperature (*3, 4*). Less well understood is purine recognition of double-helical DNA (purine·purine·pyrimidine triplets) (*7, 9–11*). Recently, purine oligonucleotides have been postulated to bind parallel to purines in duplex DNA by triple helix formation (A·AT and G·GC base triplets) (*11*).

We examined the relative affinities of common bases for all four base pairs within a pyrimidine triple helix motif. We report that G in a pyrimidine oligonucleotide specifically recognizes TA base pairs within mixed purine-pyrimidine sites. We believe that this G·TA triplet represents a new specific interaction stabilizing triplex formation in mixed purine-pyrimidine sequences. Although there will undoubtedly be sequence-composition effects, this finding extends specific recognition within the pyrimidine triple-helix motif to three of the four possible base pairs in double-helical DNA.

Arnold and Mabel Beckman Laboratories of Chemical Synthesis, Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125.

*To whom correspondence should be addressed.