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.

Recognition of TA by G was revealed by the study of the effects on triple-strand formation of 20 possible base triplets at a single common position (Table 1). The use of oligonucleotides equipped with the DNA cleaving moiety, thymidine-EDTA·Fe(II) (T*) (*12, 13*), allowed the relative stabilities of triple helix formation between 30-bp DNA duplexes containing the site $d(A_7XA_7)\cdot d(T_7YT_7)$ (XY = AT, GC, CG, or TA) and a series of 15-nt oligomers differing at one base position $d(T_7ZT_7)$ [Z = T, C, A, G, or I (I = inosine)] to be determined by the affinity cleaving method (*1*) (Fig. 2B). The 30-bp duplexes were labeled with ^{32}P at the 5′ end of the target-site strand $d(T_7YT_7)$. The DNA binding-cleaving reactions were performed under conditions that were sensitive to the stability of the variable base triplet in the middle of a thymine 15-nt fragment upon triple helix formation (*p*H 7.0, 23°C, 40% ethanol). The most intense cleavage patterns were observed for the combinations Z = T, XY = AT; Z = C, XY = GC; and Z = G, XY = TA (Fig. 2A and Table 1). The cleavage observed for two of these combinations (lanes 3 and 8) represents the known ability of T and C to form T·AT and C+GC base triplets, respectively. Remarkably, intense cleavage is also observed for a G·TA base triplet (lane 18) implying that G may form specific hydrogen bonds to TA. The lack of binding-cleavage observed for base triplets G·AT, G·GC, and G·CG (lanes 15, 16, and 17, respectively, Fig. 2) demonstrates that G is specific for a TA base pair. The lack of binding-cleavage observed for base triplets T·TA, C·TA, A·TA, and I·TA shows that the base pair TA is also specific for G in the third strand

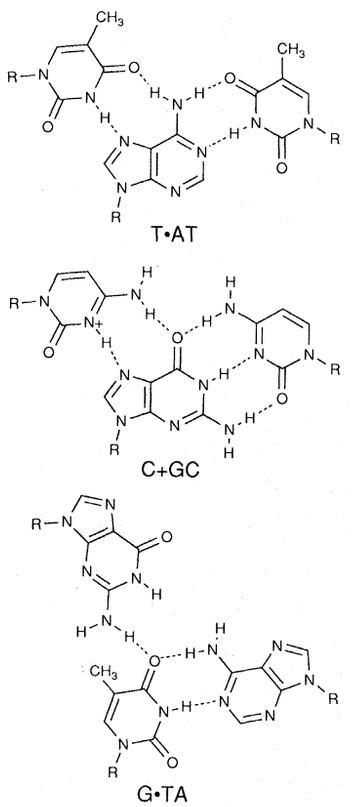

Fig. 1. (**Top**) Pyrimidine motif with isomorphous base triplets, T·AT and C+GC. Each pyrimidine base is bound in the major groove by two Hoogsteen hydrogen bonds to the purine base in the Watson-Crick duplex. (**Bottom**) Hypothetical model for G·TA base triplet within a pyrimidine triple helix motif where G (N-2) is bound in the major groove by a hydrogen bond to T (O-4) of the Watson-Crick TA base pair. With more severe perturbation of the third strand DNA backbone or disruption of the Watson-Crick hydrogen bonds between T and A, additional hydrogen bonds may conceivably be formed between G and A.

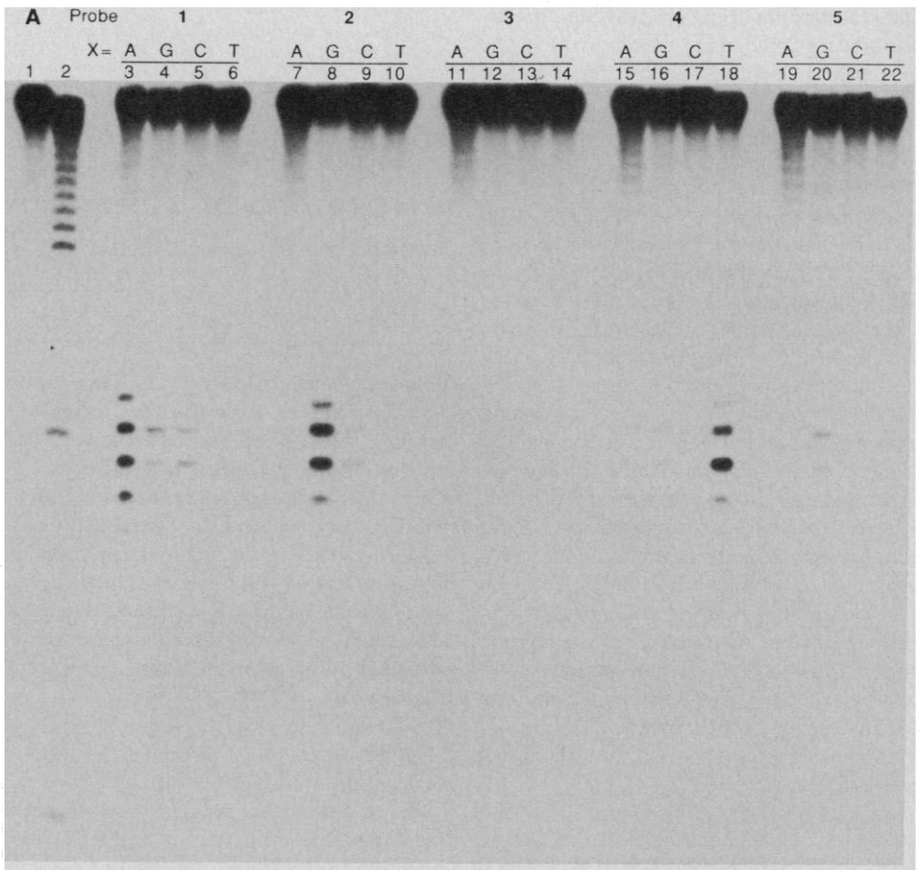

Fig. 2. (**A**) Autoradiogram of the 20% denaturing polyacrylamide gel. The cleavage reactions were carried out by combining a mixture of oligonucleotide-EDTA (1 μ*M*), spermine (1 m*M*), and Fe(II) (25 μ*M*) with the ^{32}P-labeled 30-oligomer duplex [~0.5 μ*M* in base pairs (~12,500 ± 600 cpm)] in a solution of tris-acetate, *p*H 7.0 (25 μ*M*), NaCl (100 m*M*), calf thymus DNA (100 μ*M* in base pairs), and 40% ethanol and then incubating at 0°C for 30 min and at 23°C for 30 min (*16*). Cleavage reactions were initiated by addition of dithiothreitol (DTT) (3 m*M*) and allowed to proceed for 6 hours at 23°C. The reactions were stopped by freezing and lyophilization, and the cleavage products were analyzed by gel electrophoresis. (Lanes 1 to 22) Duplexes containing 5′ end-labeled $d(A_5\text{-}T_7YT_7G_{10})$. (Lane 1) Control showing intact 5′ labeled 30-bp DNA standard obtained after treatment according to the cleavage reactions in the absence of oligonucleotide-EDTA; (lane 2) products of Maxam-Gilbert G+A sequencing reaction (*17*); (lanes 3 to 22) DNA cleavage products produced by oligonucleotide-EDTA·Fe (**1** to **5**); **1** (lanes 3 to 6), **2** (lanes 7 to 10), **3** (lanes 11 to 14), **4** (lanes 15 to 18), **5** (lanes 19 to 22). XY = AT (lanes 3, 7, 11, 15, and 19); XY = GC (lanes 4, 8, 12, 16, and 20); XY = CG (lanes 5, 9, 13, 17, and 21); XY = TA (lanes 6, 10, 14, 18, and 22). (**B**) (Above) Sequences of oligonucleotide-EDTA **1** to **5** where T* is the position of the thymidine-EDTA (*13*). The oligonucleotides differ at one base position indicated in bold type. (Below) DNA cleavage pattern derived by densitometry of the autoradiogram shown in (A) (lane 18). The heights of the arrows represent the relative cleavage intensities at the indicated bases. The box indicates the double-stranded sequence bound by oligonucleotide-EDTA·Fe(II) **1** to **5**. The Watson-Crick base pair (AT, GC, CG, or TA) opposite the variant base in the oligonucleotide is shaded (see Table 1).

(lanes 6, 10, 14, and 22, Fig. 2A and Table 1). In order to test for the importance of the amino group of G in the G·TA base triplet, G was replaced with I, a purine base that lacks NH_2 at position 2. The lack of cleavage observed with oligonucleotide **5** [Z = I] indicates that the amino group of guanine is important for recognition of TA base pairs. No binding is observed for Z = 8-bromoguanosine, XY = TA. Since 8-bromoguanosine prefers the *syn* conformation (*14*), the simplest model is a G·TA base triplet structure where G is in the *anti* conformation with respect to the sugar ring and forms at least one hydrogen bond between G and T (G, N-2 to T, O-4) (Fig. 1, bottom). Finally, minor cleavage was observed for triplets T·GC, T·CG, C·CG, and G·CG (lanes 4, 5, 9, and 17). These bases may hydrogen-bond to the Watson-Crick base pair to a lesser extent or may be sterically compatible with triple strand formation.

To determine whether this specificity can be observed at mixed sequences within larger DNA fragments, cleavage of a 799-bp simian virus 40 (SV40) restriction fragment by an oligonucleotide-EDTA·Fe(II) series (**6** to **10**) (5′-T*$T_3C_2TC_2T_2ZTCT_2$C-3′) was examined. This restriction fragment contains the 17-bp sequence d(AAAAGGAGGAATAGAAG), which represents a purine-rich site containing one pyrimidine (T). The cleavage efficiencies of oligonucleotides **6** to **10**, differing at one base position opposite the TA Watson-Crick base pair, were examined under conditions sensitive to the stability of the base triplet at the TA site (*p*H 6.6, 37°C). Oligonucleotide-EDTA **9**, but not **6**, **7**, **8**, or **10**, produced significant site-specific cleavage on the 799-bp fragment (Fig. 3A). Oligonucleotide **9** contains G opposite the TA base pair. Site-specific

Table 1. Twenty base triplets were examined for binding specificity compatible with the pyrimidine-Hoogsteen triple helix motif by the experiment described in Fig. 2. Relative cleavage efficiencies are (+++) 25 to 30 (±3)% cleavage, (+) 10 to 15 (±1.5)% cleavage, (−) <5 (±0.5)% cleavage. The data are obtained from scintillation counting and densitometric analysis of the autoradiogram shown in Fig. 2A.

W-C duplex (XY)	Pyrimidine third strand (Z)				
	T	C	A	G	I
AT	+++	−	−	−	−
GC	+	+++	−	−	−
CG	+	+	−	+	−
TA	−	−	−	+++	−

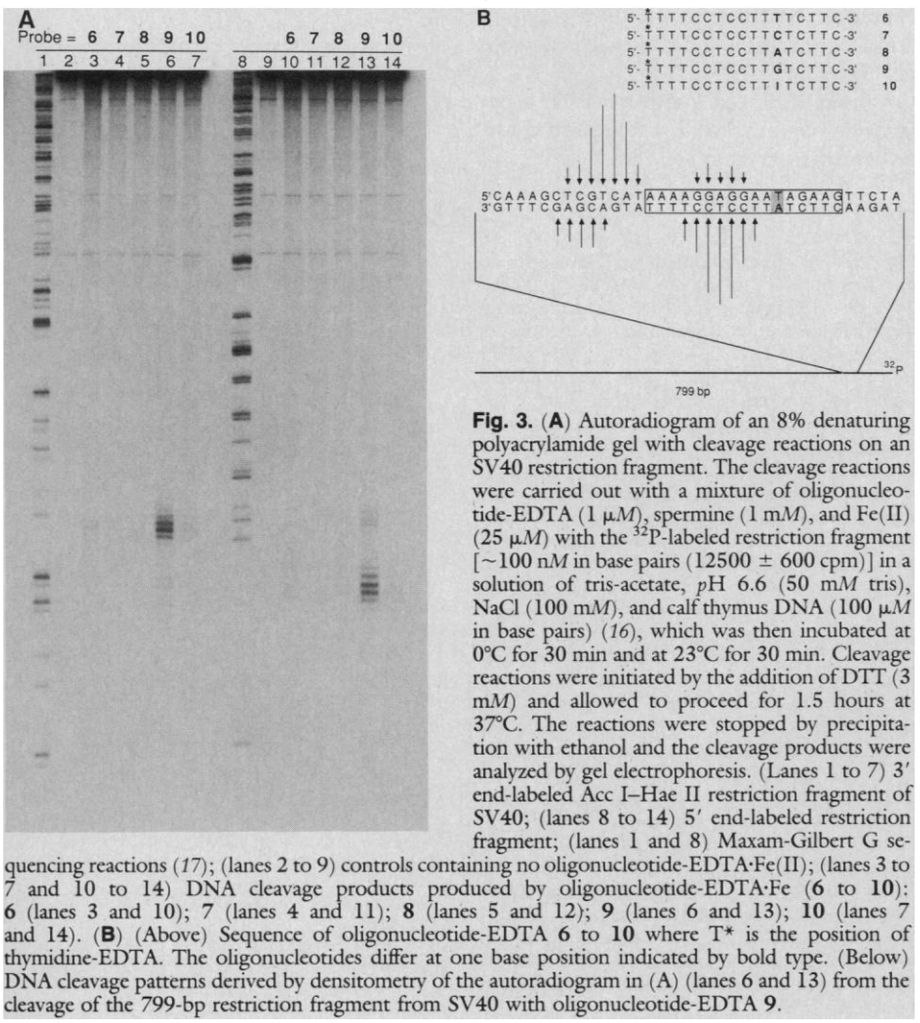

Fig. 3. (**A**) Autoradiogram of an 8% denaturing polyacrylamide gel with cleavage reactions on an SV40 restriction fragment. The cleavage reactions were carried out with a mixture of oligonucleotide-EDTA (1 μ*M*), spermine (1 m*M*), and Fe(II) (25 μ*M*) with the ^{32}P-labeled restriction fragment [~100 n*M* in base pairs (12500 ± 600 cpm)] in a solution of tris-acetate, *p*H 6.6 (50 m*M* tris), NaCl (100 m*M*), and calf thymus DNA (100 μ*M* in base pairs) (*16*), which was then incubated at 0°C for 30 min and at 23°C for 30 min. Cleavage reactions were initiated by the addition of DTT (3 m*M*) and allowed to proceed for 1.5 hours at 37°C. The reactions were stopped by precipitation with ethanol and the cleavage products were analyzed by gel electrophoresis. (Lanes 1 to 7) 3′ end-labeled Acc I–Hae II restriction fragment of SV40; (lanes 8 to 14) 5′ end-labeled restriction fragment; (lanes 1 and 8) Maxam-Gilbert G sequencing reactions (*17*); (lanes 2 to 9) controls containing no oligonucleotide-EDTA·Fe(II); (lanes 3 to 7 and 10 to 14) DNA cleavage products produced by oligonucleotide-EDTA·Fe (**6** to **10**): **6** (lanes 3 and 10); **7** (lanes 4 and 11); **8** (lanes 5 and 12); **9** (lanes 6 and 13); **10** (lanes 7 and 14). (**B**) (Above) Sequence of oligonucleotide-EDTA **6** to **10** where T* is the position of thymidine-EDTA. The oligonucleotides differ at one base position indicated by bold type. (Below) DNA cleavage patterns derived by densitometry of the autoradiogram in (A) (lanes 6 and 13) from the cleavage of the 799-bp restriction fragment from SV40 with oligonucleotide-EDTA **9**.

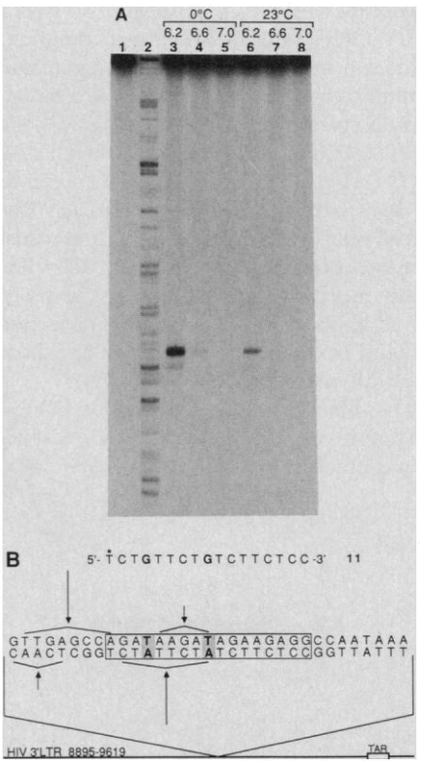

Fig. 4. (**A**) Autoradiogram of a 4% denaturing polyacrylamide gel with cleavage reactions on an HIV restriction fragment. The cleavage reactions were carried out with a mixture of oligonucleotide-EDTA probe **11** (1.67 μ*M*), spermine (1 m*M*), and Fe(II) (25 μ*M*) with the ^{32}P-labeled restriction fragment [~100 n*M* in base pairs (8000 ± 300 cpm)] in a solution of tris-acetate, *p*H 6.6 (50 m*M* tris), NaCl (100 m*M*) calf thymus DNA (100 μ*M* in base pairs), and 10% ethanol; the mixture was incubated 1 hour at the reaction temperature (*16*). Cleavage reactions were initiated by the addition of DTT (3 m*M*) and allowed to proceed for 16.5 hours at 0° or 23°C. The reactions were stopped by precipitation with ethanol and the resulting cleavage products were analyzed by gel electrophoresis (Fig. 4A). (Lanes 1 to 8) The 3′ end-labeled Hind III–Xho I restriction fragment of pHIV-CAT; (lane 1) control containing no oligonucleotide-EDTA·Fe(II); (lane 2) Maxam-Gilbert G sequencing reaction (*32, 33*); (lanes 3 to 8) the DNA cleavage products in the presence of oligonucleotide **11**; (lanes 3 to 5) at 0°C, (lanes 6 to 8) at 23°C, (lanes 3 and 6) at *p*H 6.2, (lanes 4 and 7) at *p*H 6.6, (lanes 5 and 8) at *p*H 7.0. (**B**) (Above) Sequence of oligonucleotide-EDTA **11** where T* is the position of thymidine-EDTA. (Below) The arrows represent the sites of cleavage over the indicated bases and the box indicates the double-stranded sequence in the HIV LTR bound by oligonucleotide **11**.

cleavage by oligonucleotides **6**, **7**, **8**, and **10** can be increased under less stringent binding conditions for triple helix formation such as lower *p*H or temperature or added ethanol (*3*).

The formation of G·TA triplets at DNA sites containing more than one T within a local purine tract was tested by examining a site in the 3′ long terminal repeat (LTR) of HIV DNA (Fig. 4B). The 17-bp sequence d(AGATAAGATAGAAGAGG) in a pHIV-CAT Hind III–Xho I restriction fragment is a mixed-purine duplex target containing two thymines (*14*). DNA cleavage was observed by oligonucleotide-EDTA **11** (5′-T*CTGT$_2$CTGTCT$_2$CTC$_2$-3′) at 0° and 23°C (*p*H range 6.2 to 7.0) in the presence of 10% ethanol. Strand scission was observed only at the target site, with maximal cleavage occurring at *p*H 6.2, 0°C. For comparison, comparable cleavage was observed at *p*H 6.6, 37°C, for oligonucleotide **9**, which recognizes a purine site of similar size with one pyrimidine (Fig. 3A).

The ability of oligonucleotide-EDTA·Fe(II) **11** to cause site-specific double-strand breaks in pHIV-CAT DNA is documented in Fig. 5. The plasmid pHIV-CAT was digested with Bam HI to produce a 4.95-kbp fragment, which contained the 3′ LTR of HIV with the site d(AGATAAGATAGAAGAGG) located 1.54 and 3.41 kbp from the ends. The ^{32}P end-labeled DNA was allowed to react with oligonucleotide-EDTA·Fe(II) in the presence of ascorbate at 0° or 23°C (*p*H 6.2 to 7.0). Separation of the cleavage products by agarose gel electrophoresis revealed one major cleavage site producing two DNA fragments, 1.54 and 3.41 kbp in size (Fig. 5). The HIV-DNA data indicate that an oligonucleotide binding a purine site with two thymines requires lower temperature and *p*H and the presence of ethanol. Nevertheless, under these less stringent binding conditions, additional cleavage sites do not appear. This result suggests a sequence composition limitation on the number of G residues within a pyrimidine oligonucleotide that is tolerated for TA recognition (such as T > C >> G). If one assumes that an A′-like conformation forms (*15*), model-building studies suggest that a G·TA triplet within a pyrimidine triple helix motif displaces the position of the deoxyribosyl group and hence the phosphodiester backbone in the third strand, a displacement that is not likely to be energetically favorable.

Although the G·TA triplet within a pyrimidine oligonucleotide extends, in a formal sense, triple helix specificity to three of the four possible base pairs of double-helical DNA, some limitations on sequence composition are likely. It is not clear how far one can deviate from homopurine-homopyrimidine target sequences and still obtain triple helix formation. Rather this result may provide structural leads for the design of deoxyribonucleosides (or their analogs) with nonnatural heterocycles directed toward a more general solution. With regard to the putative purine·purine·pyrimidine motif (A·AT and G·GC triplets) (*7, 10, 11*), these data suggest that there may be distinct conformational families of intermolecular triple helices (such as homopyrimidine versus homopurine donor third strand), which are not mutually compatible. The specificity of base triplets may differ for each structural motif because of different alignment of the deoxyribosyl-phosphodiester backbone and hence the heterocyclic bases along the major groove, which would create different optimal hydrogen bonding patterns. For example, from the data presented here, within a pyrimidine-rich oligonucleotide, G prefers to bind TA base pairs, whereas within a purine-rich oligonucleotide, G apparently binds GC base pairs.

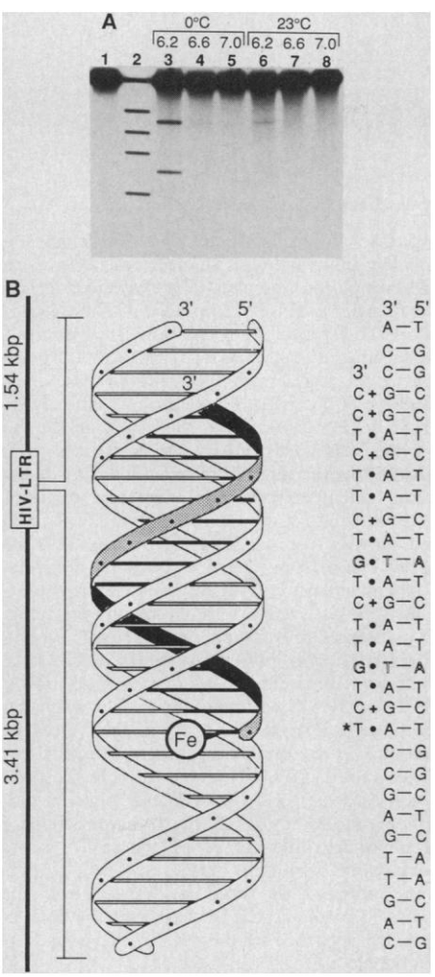

Fig. 5. (**A**) Autoradiogram of double-strand cleavage of pHIV-CAT DNA (4.95 kbp) analyzed on a 1% agarose gel. The cleavage reactions were carried out by combining a mixture of oligonucleotide-EDTA (1 μ*M*), spermine (1 m*M*), and Fe(II) (1.25 μ*M*) with the ^{32}P-labeled linearized plasmid [~0.1 μ*M* in base pairs (~27,250 ± 250 cpm)], in a solution of tris-acetate, *p*H 7.0 (50 μ*M*), NaCl (100 m*M*), calf thymus DNA (100 μ*M* in base pairs), and 10% ethanol; the mixture was incubated 1 hour at the reaction temperature (*16*). Cleavage reactions were initiated by addition of ascorbate (1 m*M*) and allowed to proceed for 18 hours at 0° and 23°C. The reactions were stopped by precipitation with ethanol, and the cleavage products were analyzed by gel electrophoresis. (Lanes 1 to 8) pHIV-CAT linearized with Bam HI and 3′ end-labeled at both ends; (lane 1) control containing no oligonucleotide-EDTA·Fe(II); (lane 2) DNA size markers obtained by digestion of Bam HI linearized pHIV-CAT with Hind II and Xho I: 4950 (undigested DNA), 3725, 3003, 1947, 1225; (lanes 3 to 8) DNA cleavage products produced by oligonucleotide-EDTA·Fe(II) **11**; (lanes 3 to 5) at 0°C, (lanes 6 to 8) at 23°C, (lanes 3 and 6) at *p*H 6.2, (lanes 4 and 7) at *p*H 6.6, (lanes 5 and 8) at *p*H 7.0. (**B**) (Left) The coarse resolution cleavage pattern from (A). (Right) Simplified model of the triple helix complex between the bound oligonucleotide-EDTA·Fe(II) **11** and a single site within the 4.95-kb plasmid HIV-CAT DNA.

REFERENCES AND NOTES

1. P. B. Dervan, *Science* **232**, 464 (1986).
2. ———, in *Nucleic Acids and Molecular Biology*, F. Eckstein and D. M. J. Lilley, Eds. (Springer-Verlag, Heidelberg, 1988), vol. 2, pp. 49–64; C. Helene and N. T. Thuong, *ibid.*, pp. 105–123; M. B. Fleisher, H.-Y. Mei, J. K. Barton, *ibid.*, pp. 65–84.
3. H. E. Moser and P. B. Dervan, *Science* **238**, 645 (1987).
4. V. I. Lyamichev *et al.*, *Nucleic Acids Res.* **16**, 2165 (1988); S. A. Strobel *et al.*, *J. Am. Chem. Soc.* **110**, 7927 (1988); T. J. Povsic and P. B. Dervan, *ibid.* **111**, 3059 (1989); L. J. Maher *et al.*, *Science*, in press.
5. For α-oligodeoxyribonucleotides see: T. L. Doan *et al.*, *Nucleic Acids Res.* **15**, 7749 (1987); D. Praseuth *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 1349 (1988).
6. G. Felsenfeld, D. R. Davies, A. Rich, *J. Am. Chem. Soc.* **79**, 2023 (1957); K. Hoogsteen, *Acta Crystallogr.* **12**, 822 (1959); A. M. Michelson, J. Massouilie, W. Guschlbauer, *Prog. Nucleic Acid Res. Mol. Biol.* **6**, 83, (1967); G. Felsenfeld and H. T. Miles, *Annu. Rev. Biochem.* **36**, 407 (1967); M. N. Lipsett, *Biochem. Biophys. Res. Commun.* **11**, 224 (1963); *J. Biol. Chem.* **239**, 1256 (1964); F. B. Howard, J. Frazier, M. N. Lipsett, H. T. Miles, *Biochem. Biophys. Res. Commun.* **17**, 93 (1964); J. H. Miller and H. M. Sobell, *Proc. Natl. Acad. Sci. U.S.A.* **55**, 1201 (1966); A. R. Morgan and R. D. Wells, *J. Mol. Biol.* **37**, 63 (1968); J. S. Lee, D. A. Johnson, A. R. Morgan, *Nucleic Acids Res.* **6**, 3073 (1979); S. M. Mirkin *et al.*, *Nature* **330**, 495 (1987); H. Htun and J. E. Dahlberg, *Science* **241**, 1791 (1988); J. C. Hanvey *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6292 (1988); J. C. Hanvey *et al.*, *J. Biol. Chem.* **263**, 7386 (1988).
7. Y. Kohwi and T. Kohwi-Shigematsu, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3781 (1988).
8. P. Rajagopal and J. Feigon, *Nature* **239**, 637 (1989).
9. C. Marck and D. Thiele, *Nucleic Acids Res.* **5**, 1017 (1978).
10. S. L. Broitman, D. D. Im, J. R. Fresco, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 5120 (1987); A. G. Letai *et al.*, *Biochemistry* **27**, 9108 (1988).
11. M. Cooney *et al.*, *Science* **241**, 456 (1988).

12. R. P. Hertzberg and P. B. Dervan, *Biochemistry* **23**, 3934 (1984).
13. G. B. Dreyer and P. B. Dervan, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 963 (1985).
14. S. S. Tavale and H. M. Sobell, *J. Mol. Biol.* **48**, 109 (1970); F. Jordan and H. Niv, *Biochim. Biophys. Acta* **476**, 265 (1977).
15. S. Arnott, P. J. Bond, E. Selsing, P. J. C. Smith, *Nucleic Acids Res.* **3**, 2459 (1976).
16. The *p*H values are not corrected for temperature or presence of ethanol and are given for the tenfold-concentrated buffer solutions at 25°C.
17. T. Maniatis *et al.*, *Molecular Cloning, A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982); A. M. Maxam and W. Gilbert, *Methods Enzymol.* **65**, 499 (1980).
18. We thank the National Institutes of Health (GM-35724), the Office of Naval Research, and the Parsons Foundation for generous support. We thank S. Singleton for valuable assistance in model building, A. Frankel for a gift of pHIV-CAT, and T. Povsic for large-scale plasmid preparation.

26 April 1989; accepted 7 July 1989

Synthesis of Functional Human Hemoglobin in Transgenic Mice

RICHARD R. BEHRINGER, THOMAS M. RYAN, MICHAEL P. REILLY, TOSHIO ASAKURA, RICHARD D. PALMITER, RALPH L. BRINSTER, TIM M. TOWNES

Human α- and β-globin genes were separately fused downstream of two erythroid-specific deoxyribonuclease (DNase) I super-hypersensitive sites that are normally located 50 kilobases upstream of the human β-globin gene. These two constructs were coinjected into fertilized mouse eggs, and expression was analyzed in transgenic animals that developed. Mice that had intact copies of the transgenes expressed high levels of correctly initiated human α- and β-globin messenger RNA specifically in erythroid tissue. An authentic human hemoglobin was formed in adult erythrocytes that when purified had an oxygen equilibrium curve identical to the curve of native human hemoglobin A (Hb A). Thus, functional human hemoglobin can be synthesized in transgenic mice. This provides a foundation for production of mouse models of human hemoglobinopathies such as sickle cell disease.

CORRECTLY REGULATED EXPRESSION of human β-globin genes in transgenic mice is well documented (*1, 2*). The human gene is expressed only in adult erythroid tissue and, in some animals with relatively high transgene copy numbers, the level of human β-globin mRNA is equivalent to endogenous mouse β-globin mRNA. Analysis of constructs with β-globin gene fragments inserted upstream of a reporter gene demonstrate that sequences located immediately upstream, within and downstream of the gene contribute to the correct temporal and tissue specific expression (*3*). Sequences located 50 kb upstream of the β-globin gene also have an effect on globin gene expression (*4–8*). When these sequences that contain erythroid-specific, DNase I super-hypersensitive (HS) sites are fused upstream of the human β-globin gene and injected into fertilized mouse eggs, large amounts of human β-globin mRNA are synthesized in virtually all transgenic mice that develop (*5, 7*). These experiments suggest that the super-hypersensitive sites define locus activation regions that "open" a large chromosomal domain for expression specifically in erythroid cells and dramatically enhance globin gene expression.

The human α1-globin gene is also expressed at high levels in erythroid tissue of transgenic mice when the injected gene is flanked by super-hypersensitive sites from the human β-globin locus (*8*). Thus a complete human hemoglobin could be synthesized in mice if human α- and β-globin gene constructs were coinjected into fertilized eggs. Previous studies demonstrated that two of the five HS sites in the β-globin locus were sufficient for high-level expression (*7, 8*). Therefore, we inserted HS I and II (a 12.9-kb Mlu I–Cla I fragment) upstream of the human α1- and β-globin genes (Fig. 1) and coinjected equimolar amounts of these constructs into fertilized mouse eggs (*9*). The eggs were transferred into the oviducts of pseudopregnant foster mothers, and seven transgenic mouse lines were established from founder animals that contained intact copies of the injected fragments. Total RNA from ten tissues of adult progeny were then analyzed for correctly initiated human α-, human β-, mouse α-, and mouse β-globin mRNA by primer extension (*10*) (Fig. 2A). Human α- and β-globin transgenes were expressed only in blood and spleen, which are both erythroid tissues in mice; detection in the lung is the result of blood contamination (*11*) because both human and mouse α- and β-globin mRNA are observed in this nonerythroid tissue. Human α- and β-globin mRNA levels in blood, as measured by solution hybridization, were 100% and 120% of endogenous mouse β-globin mRNA, respectively. Therefore, erythroid-specific, human α- and β-globin gene expression can be achieved in adult transgenic mice after coinjection of α- and β-globin constructs that contain HS I and II.

To determine whether complete human hemoglobins were formed, we separated hemolysates (*12*) of the blood of animals from two different transgenic lines by nondenaturing isoelectic focusing (IEF) (Fig. 2B). The first lane is a mouse control and the last lane is a normal human sample. The predominant band in each of the controls is the major adult hemoglobin; mouse $\alpha_2\beta_2$ or human $\alpha_2\beta_2$, respectively. In both transgenic mouse samples 5394 and 5393, bands that run at the same pI as human Hb A

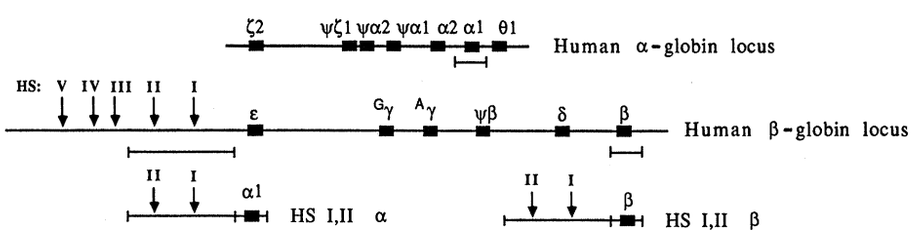

Fig. 1. HS I,II α-globin and HS I,II β-globin gene constructs. Eighty-five kilobases of the human β-globin locus and 35 kb of the human α-globin locus are drawn to scale. The brackets beneath the HS sites, α1-globin gene, and β-globin gene indicate fragments used for construction. A 12.9-kb Mlu I–Cla I fragment that contained erythroid-specific, DNase I super-hypersensitive (HS, arrow) sites I and II from the human β-globin locus was inserted into a modified pUC19 plasmid upstream of a 3.8-kb Bgl II–Eco RI fragment carrying the human α1-globin gene or a 4.1-kb Hpa I–Xba I fragment with the human β-globin gene. The 16.7- and 17.0-kb fragments with HS I,II α-globin and HS I,II β-globin were separated from plasmid sequences and coinjected into fertilized mouse eggs (*9*).

R. R. Behringer and R. L. Brinster, Laboratory of Reproductive Physiology, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA 19104.
T. M. Ryan and T. M. Townes, Department of Biochemistry, Schools of Medicine and Dentistry, University of Alabama at Birmingham, Birmingham, AL 35294.
M. P. Reilly and T. Asakura, Department of Pediatrics and Department of Biochemistry and Biophysics, The Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, PA 19104.
R. D. Palmiter, Department of Biochemistry, Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195.