particular flexible segment can contribute to site-specific DNA binding. The arm clearly has an important role in the λ repressor-operator interactions. The extent to which such segments are used in DNA binding may be underestimated, since some flexible segments may lack the sequence similarities that facilitate the study of well-structured DNA binding motifs.

REFERENCES AND NOTES

1. M. A. Weiss, J. L. Eliason, D. J. States, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6019 (1984).
2. C. O. Pabo *et al.*, *Nature* **298**, 441 (1982).
3. S. R. Jordan and C. O. Pabo, *Science* **242**, 893 (1988).
4. J. L. Eliason, M. A. Weiss, M. Ptashne, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2339 (1985).
5. Plasmid pWL103 expresses a synthetic gene for the NH_2-terminal 102 amino acids of repressor and contains a number of unique restriction sites (W. Lim and R. T. Sauer, personal communication). The plasmid was digested with Nco I (which overlaps the initiation codon), and the overhangs were filled in enzymatically. This procedure generates a blunt-ended ATG initiation codon. The DNA was then cut with BssH II (which overlaps codons 16 and 17), and the nearly full-length linear plasmid DNA was gel purified. Oligonucleotides corresponding to the coding strand of repressor from codon 1 through the bases complementary to the BssH II overhang were synthesized. For each of the six codons subjected to mutagenesis, a separate synthesis was done. At the codon of interest the oligonucleotide was synthesized with an equimolar mixture of all four nucleotides at the first two positions and C or G at the third position. A complementary primer was synthesized and annealed such that a BssH II overhang was generated at the 3′ end of the mutant strand. Enzymatic extension with the Klenow fragment of DNA Pol1 generated a pool of fragments that were then ligated into the vector, and the ligation mixture was used to transform *Escherichia coli* strain X90. Selection of functional repressor variants was performed by plating transformation mixes on plates containing ampicillin (100 mg/ml) and approximately 10^9 phage KH54, a mutant of λ that has a deletion covering the gene for repressor. In order to ensure that sampling errors were not responsible for the observed preferences at codons 3, 4, and 5, sequences were obtained from each of at least seven independent transformations at each codon. The construction of the specifically introduced nonfunctional mutants was done in a similar fashion, except that both strands were chemically synthesized with the desired mutation.
6. Although we find a number of functional substitutions at residue 6, substitutions to Asp, Asn, Glu, or Gln at this position are inactive and are expressed at undetectably low levels. These variants were identified by sequence screening of a restricted codon-directed mutagenesis experiment. This second round of mutagenesis was done because we were concerned that some problem in the oligonucleotide synthesis might have led to an underrepresentation of these codons. For this experiment, the mutagenic oligonucleotide was synthesized with the following mixture of nucleotides at codon 6: (CAG)-(A)-(GC). The Asn, Asp, Glu, and Gln substitutions were identified by sequence analysis. In pulse-chase labeling experiments, we did not observe any incorporation of radiolabeled methionine into these mutant proteins. Under the same conditions, the wild-type protein is the major radiolabeled species and is stable for at least 2 hours.
7. The deletion of codon 1 and the deletion of codons 1 and 2 probably arose as a result of oligonucleotides which were prematurely terminated during synthesis (by three and six nucleotides, respectively).
8. B. J. Meyer, R. Maurer, M. Ptashne, *J. Mol. Biol.* **139**, 163 (1980).
9. The ^{32}P end-labeled binding site consisted of the 17-bp O_L1 site in the center of a 50-bp double-stranded oligonucleotide. The assay mix was 100 mM tris (pH 7.8), 50 mM KCl, 2 mM $CaCl_2$, 5% glycerol, sonicated calf thymus DNA (40 mg/ml), $<10^{-11}$ M radiolabeled DNA, and varying amounts of repressor in a final volume of 20 ml. Samples were allowed to equilibrate for 30 min before loading onto the gel (5% polyacrylamide, 7 mM tris, 3 mM sodium acetate, pH 8.0). Prior to loading, the gel was pre-run with circulation of the gel buffer for ~1 hour. The samples were run for ~1 hour at 100 V.
10. Final crystallographic *R* factor of 0.189 for 25,252 reflections between 8.0 and 1.8 Å; root-mean-square deviation from ideal bond lengths of 0.020 Å and for bond angles of 2.1°. Details to be published elsewhere (L. J. Beamer and C. O. Pabo, manuscript in preparation).
11. Only the arm for the monomer bound to the consensus half-site has clear electron density. The arm appears to be disordered on the nonconsensus half-site, and we note that the contacts that the arm makes with G7 and G9 (the central base pair) are not available for the arm on the nonconsensus half-site. These differences in the consensus and nonconsensus half-sites are fully consistent with the differences in chemical protection patterns [R. T. Sauer, C. O. Pabo, B. J. Meyer, K. C. Backman, M. Ptashne, *Nature* **279**, 396 (1979)] and the differences in the contributions these regions make to the binding energy (*12*).
12. A. Sarai and Y. Takeda, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6513 (1989).
13. A. Hochschild, J. Douhan III, M. Ptashne, *Cell* **47**, 807 (1986).
14. In the partially refined 2.5 Å model, the Lys^4 side chain appeared to only contact base pair 6. The higher resolution structure described here shows unambiguously that the amine of Lys^4 makes contacts to both base pair 6 and to base pair 7.
15. C. O. Pabo and M. Lewis, *Nature* **298**, 443 (1982).
16. Gel shift experiments were done essentially as described in (*9*) except that various amounts of unlabeled competitor oligonucleotides were used in addition to the large amount of calf thymus DNA. The poor binding by the Ala^3 protein complicates the interpretation of this experiment somewhat, since much higher quantities of protein are required. For this protein, the dissociation constant for the cooperative nonspecific binding adjacent to a specific site appears to be close to that for binding to the specific site. The additional higher level bands seen in the gel shift experiment due to this binding are competed away equally well by O_L1 and by the operator site with the GC to AT change at base pair 8. The amount of operator DNA required to compete away these bands is much smaller on a mass basis than the amount of calf thymus DNA present, and addition of the equivalent mass of extra calf thymus DNA does not affect the gel shift pattern. In the case of wild-type protein, O_L1 is approximately a 15-fold better competitor than is the GC8 to AT8 operator variant.
17. Some of the features of this high-resolution structure that were not apparent at lower resolution were correctly predicted by Sarai and Takeda (*12*). By attempting to reconcile the 2.5 Å resolution structure and the results of binding studies to variant operator sequences, these authors correctly anticipated that Lys^4 contacts G7 (in addition to G6) and suggested that Lys^3 contacts G8.
18. M. Ptashne, *A Genetic Switch* (Cell/Blackwell, Cambridge, MA, 1986).
19. L. J. Beamer, thesis, Johns Hopkins University, Baltimore, MD (1991).
20. A. T. Brünger, J. Kuriyan, M. Karplus, *Science* **235**, 458 (1987).

21. We thank T. Steitz for use of the data collection facilities at Yale University and members of his laboratory, J. Friedman and M. Rould for help with data collection and processing, the National Cancer Institute for use of the CRAY supercomputer, and C. Wendling and A. Collector for oligonucleotide synthesis. Supported by NIH grant no. GM31471 and the Howard Hughes Medical Institute.

8 May 1991; accepted 23 July 1991

Structure and Stability of X·G·C Mismatches in the Third Strand of Intramolecular Triplexes

ROMÁN F. MACAYA, DARA E. GILBERT, SHIVA MALEK, JANET S. SINSHEIMER, JULI FEIGON

Intramolecular DNA triplexes that contain eight base triplets formed from the folding of a single DNA strand tolerate a single X·G·C mismatch in the third strand at acidic pH. The structure and relative stability of all four triplets that are possible involving a G·C Watson-Crick base pair were determined with one- and two-dimensional proton nuclear magnetic resonance techniques. Triplexes containing A·G·C, G·G·C, or T·G·C triplets were less stable than the corresponding parent molecule containing a C·G·C triplet. However, all mismatched bases formed specific hydrogen bonds in the major groove of the double helix. The relative effect of these mismatches on the stability of the triplex differs from the effect assayed (under different conditions) by two-dimensional gel electrophoresis and DNA cleavage with oligonucleotide EDTA·Fe(II).

TRIPLE HELICAL NUCLEIC ACID structures formed from one homopurine and two homopyrimidine RNA sequences were proposed more than 30 years ago (*1*). Recent evidence indicates that such DNA sequences may also fold back on themselves to form intramolecular triplexes, termed H-DNA (*2*), when contained in supercoiled plasmids; moreover, these structures may be relevant in vivo (*3*). Triplex formation is currently being widely investigated because of potential therapeutic applications in the specific inhibition of transcription (*4*) and for use in chromosome mapping (*5*). Proposed models for pyrimidine·purine·pyrimidine DNA triplexes have the second pyrimidine strand Hoogsteen base-paired to purines in the

Department of Chemistry and Biochemistry and Molecular Biology Institute, University of California, Los Angeles, CA 90024.

major groove of Watson-Crick base-paired duplex DNA (or RNA) (*6, 7*). The base pairing schemes for T·A·T and C^+·G·C triplets in pyrimidine-purine-pyrimidine DNA triplexes and the parallel orientation of the third strand relative to the purine strand have been unambiguously confirmed by our nuclear magnetic resonance (NMR) studies (*8, 9*). However, there is some evidence that DNA triplexes containing other base triplets can be formed (*10–15*). The extent to which other base triplets can form or be accommodated within this pyrimidine-purine-pyrimidine triple helical motif is not well understood. We report our investigations of the base pairing schemes for third strand X·G·C mismatches and their effect on the stability of intramolecular triplexes formed from the folding of a single DNA strand.

We synthesized oligonucleotides with the sequences shown in Fig. 1A, where X denotes C, A, G, or T. These 32-base oligonucleotides are designed to fold into intramolecular triplexes at acidic pH as shown in Fig. 1A, with the 5′ purines (1–8) forming Watson-Crick base pairs with the central pyrimidines (13–20) and Hoogsteen base pairs with the 3′ pyrimidines (25–32). A related 28-base sequence does form a stable intramolecular triplex containing T·A·T and C·G·C triplets under appropriate conditions (*9, 16*). For this study, we lengthened the triplex by one triplet and changed one of the loops from TTT to TATA in order to increase the thermal stability of the triplex. One-dimensional spectra of the exchangeable resonances of the four oligonucleotides at pH 5.2 are shown in Fig. 1. The parent oligonucleotide, with X = C, is shown in Fig. 1B. Resonance intensity for 16 H-bonded imino resonances and the non-H–bonded imino resonances in the TATA loop and C^+ amino resonances are observed. All of the imino and amino resonances have been assigned by analysis of nuclear Overhauser effect spectroscopy (NOESY) (*17*) spectra to Watson-Crick A·T and G·C and Hoogsteen T·A or C^+·G base pairs in base triplets as described (*8, 9*). The exchangeable proton spectra obtained for the oligonucleotides with X = A, G, or T are shown in Fig. 1, C, D, and E, respectively.

The observation of more than the eight Watson-Crick imino protons expected if only the partially folded duplex formed, plus resonance intensity at the characteristic chemical shift of C^+ aminos, is strongly indicative of triplex formation (*8, 9, 18*). Triplex formation for all of the oligonucleotides containing the alternative base triplets A·G·C, G·G·C, and T·G·C was confirmed by analysis of two-dimensional NOESY spectra in H_2O and D_2O. The assignments for the imino and amino resonances are given in Table 1. In addition to the imino resonances observed for the standard T·A·T and C^+·G·C triplets, an additional imino resonance is observed for each oligonucleotide, which we have identified as arising from the base X = A, G, or T. These are labeled in Fig. 1 and are discussed below. Thus, all of the oligonucleotides form intramolecular triplexes at pH 5.2 and 1°C, indicating that a single mismatch in the Hoogsteen base-paired strand can be accommodated in the formation of a nonstandard X·G·C triplet.

The H-bonding scheme for the X·G·C triplets was determined by analysis of NOESY spectra in H_2O. A portion of the NOESY spectrum of the oligonucleotide in H_2O in which X = A is shown in Fig. 2. This spectrum shows several features that are characteristic of triplex DNA containing T·A·T and C^+·G·C triplets (*8, 9, 18, 19*), and these characteristics are used in making assignments (see below). Another characteristic feature of triplex DNA is nuclear Overhauser effect (NOE) cross peaks between the Hoogsteen base-paired iminos and the 5′ neighboring purine deoxyribose H-2′, H-2″ resonances; NOE cross peaks between iminos and H-2′, H-2″ are never observed in duplex DNA. The imino resonances are assigned from the two sets of sequential imino-imino connectivities along the Watson-Crick and Hoogsteen base-paired strands (Fig. 2A, above and below the diagonal, respectively). Depending on the X·G·C base pairing scheme and the relative strength of any H bonds formed, these sequential connectivities might be disrupted at the X·G·C triplet. For X = A, no sequential connectivity between T29 and the imino assigned to A28 is observed, although there is one between T27 and A28. Although adenine is not normally protonated at N-1, observation of this very low field–shifted imino resonance is one indication that there is a protonated H-bonded adenine imino in the A·G·C triplet. Similar low field shifts are observed for the C^+ iminos in DNA triplexes. The C^+ iminos are identified by their characteristic NOEs to the C^+ aminos (Fig. 2C). Although only two sets of C^+ amino resonances are expected, three sets of cross peaks appear in this region. The additional pair of amino resonances arises from the amino group of A28. Their unusual low field chemical shift can be attributed to protonation of A28 at N-1.

The observation of the A^+ imino and two resolved amino resonances indicates that A28 forms specific H bonds from the imino and one amino proton to the Watson-Crick G4 · C17 base pair. On the basis of model building, we conclude that the sterically most favorable base-pairing scheme is that shown in Fig. 3B. This base-pairing scheme is confirmed by observed NOE cross peaks between the A28 imino and the C17 amino resonances and between the A28 imino and G4H-8 and A28H-2. The

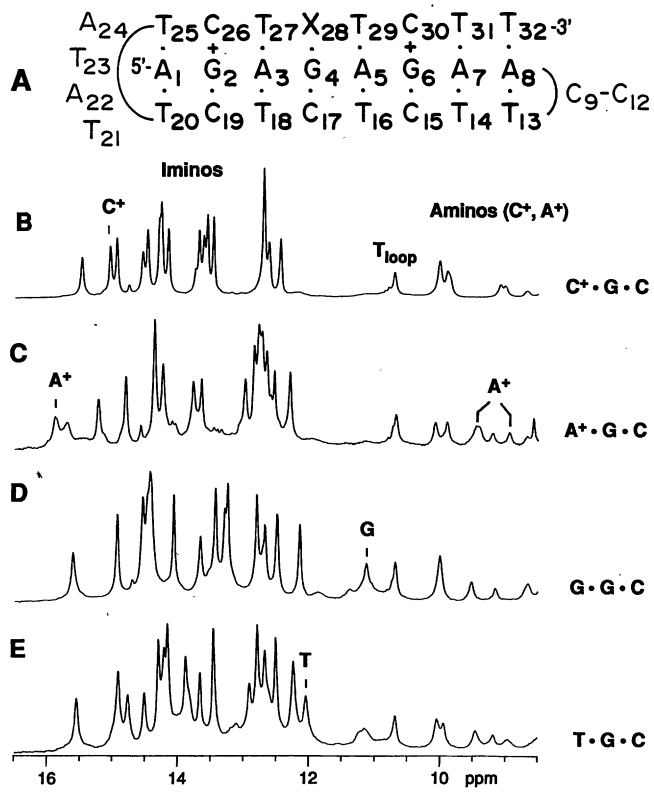

Fig. 1. (**A**) Sequence and folded conformation of the oligonucleotides investigated. The X indicates the position at which the bases A, G, T, or C were substituted. (**B** through **E**) One-dimensional 500-MHz ^1H NMR spectra of the exchangeable proton resonances at 1°C are shown for (B) X = C, (C) X = A, (D) X = G, and (E) X = T. All NMR samples were 2 mM DNA strand, 100 mM NaCl, 5 mM $MgCl_2$, pH 5.2, 400 μl in 90% H_2O–10% D_2O. DNA oligonucleotides were synthesized and purified and NMR samples were prepared as described (*9*). Assignments of the imino protons of X are as indicated. The assignments of all of the resonances shown are given in Table 1. Complete ^1H resonance assignments of (B) have been obtained (*27*). Spectra were obtained with a $1\bar{1}$ spin echo pulse sequence (*28*) (τ = 90 to 100 μs). Spectra were acquired with 4,096 complex points, 256 acquisitions, and a spectral width of 12,048 to 12,396 Hz. Spectra were processed with GE NMR software (GEM16) and line-broadened by 3 Hz before Fourier transformation.

G4H-8 and A28H-2 resonances also show a strong NOE cross peak in NOESY spectra in D_2O. Although the pK_a of adenosine N-1 is ~3.9 (*20*), which is even lower than that for cytosine, protonated adenines in G·A mismatches contained in duplex DNA have been reported both in crystal structures (*21*) and solution (*22*) at low pH.

The base pairing schemes for the G·G·C and T·G·C triplets contained in the oligonucleotides with X = G and T, respectively, were determined by the same methods as those for X = A discussed above. The imino proton of G28 in the X = G oligonucleotide resonates in the same region as the iminos from the thymines in the loop (Fig. 1D), indicating that the imino proton is not H-bonded, although it is somewhat protected from solvent exchange. The NOESY spectrum for this oligonucleotide in H_2O shows a strong NOE cross peak between the G28 imino and amino resonances. This cross peak, although broad because of the usual fast rotation around the C–N bond for the guanine aminos, indicates that one of the two amino protons must be H-bonded. The proposed base pairing scheme for the G·G·C triplet is given in Fig. 3C. Although we cannot determine unambiguously which amino proton is H-bonded to the C-6 carbonyl, model building indicates that the base pairing scheme shown is sterically more favorable in terms of the phosphodiester backbone. This H-bonding scheme is similar to the scheme proposed by Griffin and Dervan (*10*) for the G·T·A triplet in that both schemes involve a single H bond between an amino proton from the guanine in the third strand and a carbonyl group in the Watson-Crick base pair. However, our investigations of the G·T·A triplet contained within the intramolecular triplex indicate that in this case the H bond is to the other amino proton (*23*). The G·G·C H-bonding scheme differs from that suggested by Kohwi and Kohwi-Shigematsu (*13*) for their proposed $dG_n{\cdot}dG_n{\cdot}dC_n$ triplex. This is not surprising, however, given the proposed antiparallel orientation for their second purine strand.

The imino proton for T28 in the X = T oligonucleotide resonates at about 12 ppm (Fig. 1E), indicating that it is H-bonded. This shift is upfield of the usual chemical shift range for Watson-Crick A·T iminos but is in the range where Hoogsteen base-paired thymine iminos are observed (*8, 9, 18, 19, 24*). This T28 imino shows a strong NOE cross peak to G4H-8, which confirms the H bond to G4N-7. The base pairing scheme deduced for the T·G·C triplet is shown in Fig. 3D. The triplex containing the $A^+{\cdot}G{\cdot}C$ triplet shows the largest difference in chemical shifts compared with the parent X = C and the other oligonucleotides. This may indicate some distortion in the helix to accommodate this mismatched triplet.

The data given above show that intramolecular triplexes containing a single X·G·C mismatch do form. However, these mismatches do have a significant effect on the thermal stability of the triplexes formed. One can qualitatively monitor the relative stabilities of the triplexes by obtaining NMR spectra of the exchangeable resonances as a function of temperature. On the basis of such data, the order of decreasing stability for the four triplexes and therefore for the

Table 1. Chemical shifts (ppm) of H-bonded imino and C^+, A^+ amino resonances of intramolecular triplexes at 1°C.*

Base pair	Triplex X =			
	C	A	G	T
	Iminos			
A1 · T20	14.54	14.37	14.42	14.50
G2 · C19	12.63	12.70	12.65	12.65
A3 · T18	14.28	14.37	14.42	14.21
G4 · C17	12.71	12.77	12.12	12.22
A5 · T16	14.28	13.78	14.39	14.28
G6 · C15	12.71	12.85	12.77	12.77
A7 · T14	14.15	14.23	14.04	14.15
A8 · T13	14.94	14.79	14.90	14.89
T25 · A1	13.62	13.64	13.62	13.64
C26 · G2	15.47	15.22	15.58	15.53
T27 · A3	13.69	12.52	13.26	13.85
X28 · G4	15.04	15.88	11.10†	12.03
T29 · A5	13.57	12.62	13.22	12.89
C30 · G6	14.46	15.69	14.51	14.74
T31 · A7	13.47	12.97	13.40	13.45
T32 · A8	12.46	12.22	12.46	12.90
	C^+, A^+ *aminos*			
C26‡	9.09	9.20	9.18	9.14
	10.00	10.05	9.93	10.00
X28‡	9.01	8.95		
	9.90	9.43		
C30‡	9.85	9.40	9.45	9.50
	10.02	9.97	10.04	9.97

*Chemical shifts were determined by reference to the chemical shift of H_2O, which had been previously calibrated relative to DSS. †G28 imino is not H bonded. ‡Lower field resonance in each pair is H bonded.

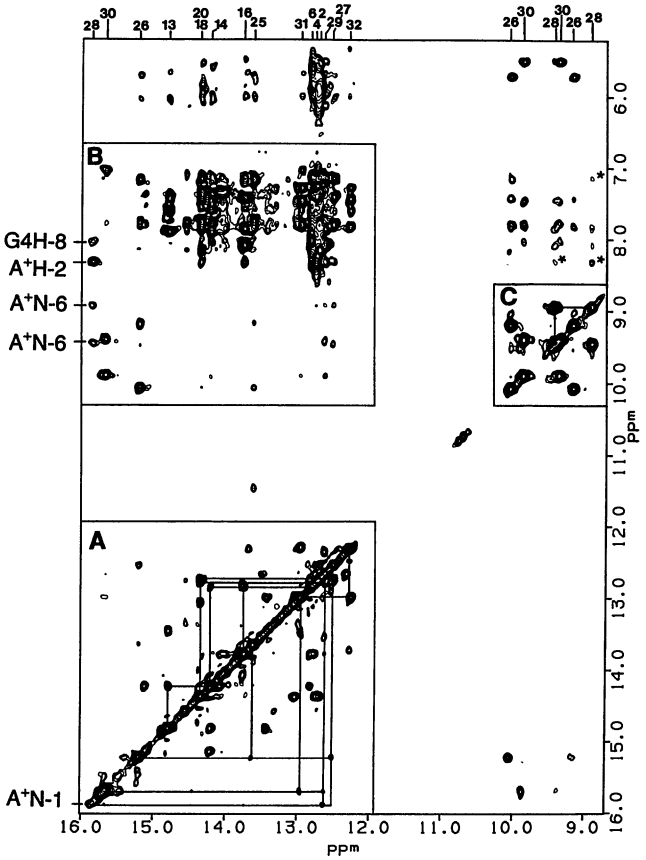

Fig. 2. Portion of NOESY spectrum in 90% H_2O–10% D_2O of the oligonucleotide with X = A at 1°C. The boxed regions contain cross peaks between (**A**) imino-imino, (**B**) imino-aromatic, amino, and (**C**) C^+ aminos and A^+ aminos. The sequential imino-imino connectivities are indicated by the solid lines in (A) for the Watson-Crick base pairs (above the diagonal) and the Hoogsteen base pairs (below the diagonal). Connectivities between the two A^+ amino resonances are indicated by solid lines in (C). Cross peaks between the A^+ imino and the aromatic and amino proton resonances that define the base pairing scheme are labeled in (B). The cross peaks from G4H-8 and A28H-2 could be distinguished by comparison to spectra obtained after deuteration of the purine H-8s (*8, 9*). Cross peaks between A28 amino and C17 amino resonances are starred (*). Assignments of the imino and amino resonances are given in Table 1. Assignments were made as described (*9, 27*). Some additional cross peak intensity that arises from a minor, unidentified form of the oligonucleotide can be seen in (A) and (B). The NOESY spectrum was acquired with a $1\bar{1}$ spin echo pulse sequence (τ = 90 μs) replacing the third pulse of a standard NOESY sequence as described (*9*). Acquisition parameters were 2,048 complex points in t_2, 256 t_1 values, 96 acquisitions per t_1 value, sweep width of 12,346 Hz, recycle delay of 2 s, and mixing time of 100 ms. The spectrum was processed with the program FTNMR (Hare Research). Before Fourier transformation the free-induction decays (FIDs) were corrected by a Gaussian window function with K = 32 and extrapolation with M = 16 (*29*) to remove the residual water signal. The 256 points were processed in both dimensions, and the spectrum was zero-filled in t_1 to give a final 2048 by 2048 real data matrix and was apodized in both dimensions by a skewed, sine-bell squared (skew = 1.5) function with a 60° phase shift.

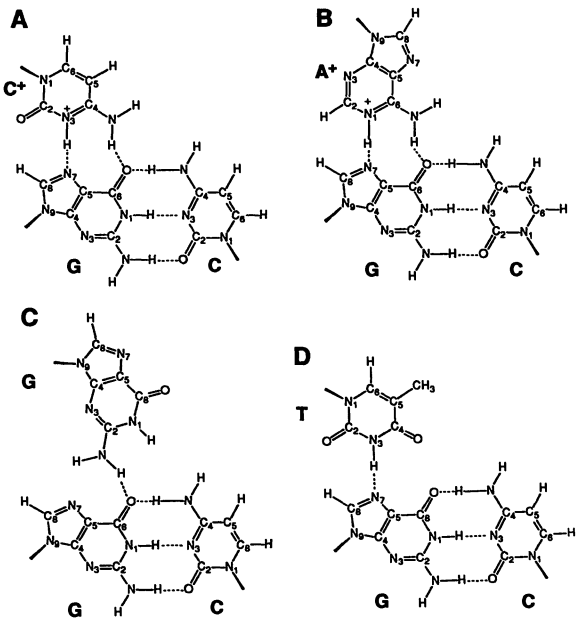

Fig. 3. H-bonding schemes for the (**A**) C^+·G·C, (**B**) A^+·G·C, (**C**) G·G·C, and (**D**) T·G·C triplets within the intramolecular triplex, as determined from the NMR studies, where the third strand is parallel to the homopurine strand. NOESY spectra in D_2O (not shown) showed that all of the bases were in the anti conformation.

four different triplets at pH 5.2 is C^+·G·C ≫ A^+·G·C > G·G·C > T·G·C. The stabilities of the four triplexes were also monitored optically as a function of temperature (Fig. 4). The melting profile for oligonucleotide X = C at pH 5.2 shows a single cooperative transition with a melting point T_m of 68°C rather than the biphasic transitions that have been reported elsewhere for DNA triplexes (*24–26*). The melting profile of the oligonucleotide at pH 8.7, where it is predominantly the partially folded duplex (*9*), also shows a single but broader transition with a T_m of 62°C (Fig. 4A). This result is consistent with our previous NMR work that showed that third strand binding stabilized the Watson-Crick duplex in triplexes formed from $d(GA)_4$ + 2 $d(TC)_4$ (*8*).

Apparently the formation of the intramolecular triplex increases the melting temperature of this oligonucleotide above that of the partially folded duplex, with the result that as soon as the third strand begins to dissociate the rest of the molecule also melts. This melting behavior will clearly depend on both salt and pH. In contrast to the parent oligonucleotide, the oligonucleotide with the T·G·C mismatch does show a biphasic melting profile, which is best fit with a three-state model. This analysis gives two melting transitions, T_m = 32°C and 63°C, corresponding to the melting of the Hoogsteen and the Watson-Crick base-paired strands, respectively. The melting profile of the G·G·C oligonucleotide is also best fit by a three-state model with T_m = 34°C and 64°C. For the A·G·C oligonucleotide, the two transitions are T_m = 44°C and 64°C. Thus, all of the mismatches decrease the stability of third strand binding in the order given above at pH 5.2.

The relative effect of X·G·C triplets on the formation and stability of DNA triplexes has been probed by different methods in two other laboratories, with differing results from each other and this study. Belotserkovskii *et al.* (*12*) assayed for H-DNA formation by two-dimensional gel electrophoresis at pH 4.2 in plasmids containing potential H-DNA–forming sequences with a single X·Y·Z mismatch. Plasmids containing T·G·C and A·G·C mismatches required two and G·G·C required three more superhelical turns to convert to H-DNA than did plasmids with the canonical T·A·T or C·G·C triplets at the same position. Griffin and Dervan (*10*) assayed for relative stabilities of triplex formation at pH 7.0 in 15-base-long T·A·T triplexes with a single X·Y·Z mismatch by putting the DNA-cleaving moiety thymidine EDTA·Fe(II) on the third strand. Triplexes containing X·Y·Z = T·A·T, C·G·C, and G·T·A showed a relative cleavage efficiency of 30 to 35%, those containing T·G·C were cleaved at 10 to 15%, and G·G·C and A·G·C showed less than 5% cleavage.

The difference in stability of the triplexes with an A·G·C mismatch in these three assays can easily be explained by the difference in pH at which the assays were done. However, both the Frank-Kamenetskii and the Dervan assays show a higher relative stability for triplexes containing a single T·G·C mismatch than those in our study, which indicates that T·G·C is the least stable of the triplexes. Although we are unable to explain this difference at this time, our results on the relative effect of X·G·C triplets on the formation and stability of DNA triplexes point up the importance of pH, solvent conditions, and base composition in determining the potential effect of alternative base triplets on triplex formation and stability.

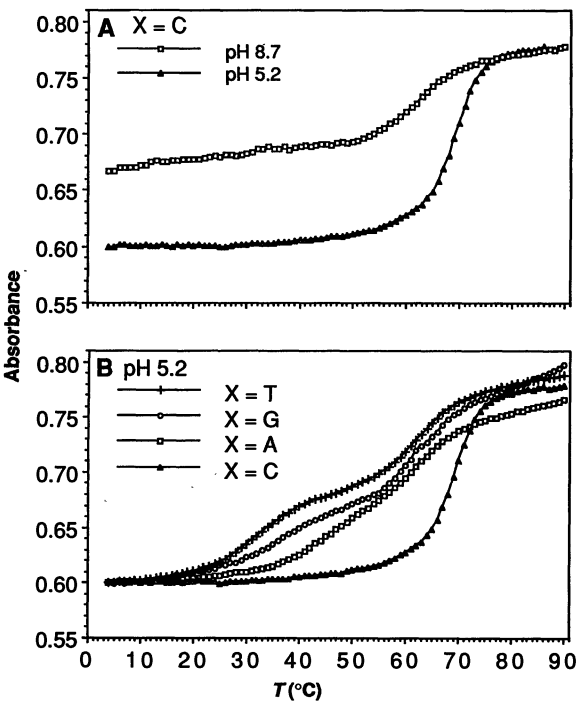

Fig. 4. Graphs of absorbance at 264 nm versus temperature for the oligonucleotides. (**A**) Melting profiles of oligonucleotide X = C at pH 5.2 and pH 8.7. Oligonucleotide is triplex at pH 5.2 and predominantly the partially folded duplex at pH 8.7. (**B**) Melting profiles of oligonucleotides X = C, A, G, and T at pH 5.2. The T_m for the triplexes are 68°C (X = C), 44° and 64°C (X = A), 34° and 64°C (X = G), and 32° and 63°C (X = T). At pH 8.7, T_m for X = C is 62°C. Absorbances were measured on an HP 8452A diode array spectrophotometer. A blank was run for each melting study and was subtracted from the baseline. The spectra in (B) were normalized to the same starting intensity. We determined the melting temperature of the oligonucleotide X = C by fitting the absorbance as a function of temperature with a two-state model (*30*) with the nonlinear least squares regression program BMDP3R (BMDP Statistical Software, Los Angeles, California). A three-state model did not give reasonable results for oligonucleotide X = C. In contrast, the X = A, G, and T oligonucleotide melting profiles fit well with a three-state model (*26*) and did not give reasonable results when fit with a two-state model.

REFERENCES AND NOTES

1. G. Felsenfeld, D. R. Davies, A. Rich, *J. Am. Chem. Soc.* **79**, 2023 (1957).
2. S. M. Mirkin *et al.*, *Nature* **330**, 495 (1987); H. Htun and J. E. Dahlberg, *Science* **241**, 1791 (1988).
3. R. D. Wells *et al.*, *FASEB J.* **2**, 2939 (1988); J. S. Lee *et al.*, *Nucleic Acids Res.* **15**, 1047 (1987); M. Kato *et al.*, *Chem. J.* **268**, 175 (1990); A. J. Kinniburgh, *Nucleic Acids Res.* **17**, 7771 (1989); J.

G. McCarthy and S. M. Heywood, *ibid.* **15**, 8069 (1987); D. A. Collier *et al.*, *J. Biol. Chem.* **263**, 7397 (1988); J. S. Lee *et al.*, *Gene* **82**, 191 (1989).
4. L. J. Maher III, B. Wold, P. B. Derman, *Science* **245**, 725 (1989); M. Cooney, G. Czernuszewicz, E. H. Postel, S. J. Flint, M. E. Hogan, *ibid.* **241**, 456 (1988); D. Praseuth *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 1349 (1988); T. L. Doan *et al.*, *Nucleic Acids Res.* **15**, 7749 (1987); J.-C. François *et al.*, *J. Biol. Chem.* **264**, 5891 (1989).
5. S. A. Strobel *et al.*, *J. Am. Chem. Soc.* **110**, 7927 (1988); H. E. Moser and P. B. Dervan, *Science* **238**, 645 (1987); J. C. Hanvey *et al.*, *Nucleic Acids Res.* **18**, 157 (1990); J.-C. François *et al.*, *Biochemistry* **28**, 9617 (1989).
6. S. Arnott and E. Selsing, *J. Mol. Biol.* **88**, 509 (1974).
7. J. S. Lee *et al.*, *Nucleic Acids Res.* **6**, 3073 (1979).
8. P. Rajagopal and J. Feigon, *Nature* **339**, 637 (1989); *Biochemistry* **28**, 7859 (1989).
9. V. Sklenář and J. Feigon, *Nature* **345**, 836 (1990).
10. L. C. Griffin and P. B. Dervan, *Science* **245**, 967 (1989).
11. R. Häner and P. B. Dervan, *Biochemistry* **29**, 9761 (1990).
12. B. P. Belotserkovskii *et al.*, *Nucleic Acids Res.* **18**, 6621 (1990).
13. Y. Kohwi and T. Kohwi-Shigematsu, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3781 (1988).
14. C. Marck and D. Thiele, *Nucleic Acids Res.* **5**, 1017 (1978); J. Bernués *et al.*, *ibid.* **18**, 4067 (1990).
15. A. G. Letai *et al.*, *Biochemistry* **27**, 9108 (1988); J. C. Hanvey *et al.*, *J. Biol. Chem.* **264**, 5950 (1989).
16. From fiber diffraction data of Arnott *et al.* (*6*) for poly dT·dA·dT it was concluded that the triplex forms an A′-DNA type helix with C3′-endo sugar puckers. Our analysis of the sugar puckers of 31- and 32-base intramolecular triplexes from phase-sensitive correlated spectroscopy (COSY) coupling patterns indicates that almost all of the sugars except the cytosines in the third strand are predominantly S-type (C2′-endo). Thus, the intramolecular triplex may form a structure closer to an underwound B-DNA type helix than an A′-DNA helix (*27*).
17. A. Kumar *et al.*, *Biochem. Biophys. Res. Commun.* **95**, 1 (1980).
18. C. de los Santos *et al.*, *Biochemistry* **28**, 7282 (1989).
19. M. M. W. Mooren *et al.*, *Nucleic Acids Res.* **18**, 6523 (1990).
20. W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984).
21. T. Brown *et al.*, *J. Mol. Biol.* **207**, 455 (1989).
22. X. Gao and D. J. Patel, *J. Am. Chem. Soc.* **110**, 5178 (1988).
23. E. Wang and J. Feigon, in preparation.
24. D. S. Pilch *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1942 (1990).
25. G. E. Plum *et al.*, *ibid.*, p. 9436.
26. G. Manzini *et al.*, *J. Mol. Biol.* **213**, 833 (1990).
27. R. F. Macaya, E. Wang, P. Schultze, V. Sklenář, J. Feigon, in preparation.
28. V. Sklenár and A. Bax, *J. Magn. Reson.* **74**, 469 (1987).
29. D. Marion *et al.*, *ibid.* **84**, 425 (1989).
30. C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry*, part 3, *The Behavior of Biological Macromolecules* (Freeman, New York, 1980).
31. Supported by grants from the NIH (R01 GM 37254-01) and the Office of Naval Research (contract N0001̈4-88-K-0180), an NSF Presidential Young Investigator Award with matching funds from AmGen, Inc., Sterling Drug, Inc., E. I. DuPont De Nemours & Company, and Monsanto Co. to J.F.; a Dissertation Year Fellowship to D.E.G.; and predoctoral training grants from NIH (GM-08185) to J.S.S. and (GM-08375) to R.F.M. We thank F. Smith for setting up and programming the optical melting apparatus and K. Koshlap for synthesizing and purifying the DNA oligonucleotides.

21 March 1991; accepted 7 June 1991

A Retroviral Oncogene, *akt*, Encoding a Serine-Threonine Kinase Containing an SH2-Like Region

ALFONSO BELLACOSA, JOSEPH R. TESTA, STEPHEN P. STAAL,* PHILIP N. TSICHLIS†

The v-*akt* oncogene codes for a 105-kilodalton fusion phosphoprotein containing Gag sequences at its amino terminus. Sequence analysis of v-*akt* and biochemical characterization of its product revealed that it codes for a protein kinase C–related serine-threonine kinase whose cellular homolog is expressed in most tissues, with the highest amount found in thymus. Although Akt is a serine-threonine kinase, part of its regulatory region is similar to the Src homology–2 domain, a structural motif characteristic of cytoplasmic tyrosine kinases that functions in protein-protein interactions. This suggests that Akt may form a functional link between tyrosine and serine-threonine phosphorylation pathways.

THE AKT8 VIRUS (*1*), THE ONLY acute transforming retrovirus isolated from a rodent T cell lymphoma (AKR) to date, transforms mink lung cells in culture. Virus rescued from nonproducer mink cells by two poorly leukemogenic amphotropic murine leukemia viruses was inoculated into newborn mice and shown to be tumorigenic (*2*). A defective clone of the AKT8 virus clone contained v-*akt*, a gene of cellular origin (*3*). The presumed human homolog of v-*akt* was cloned by screening a human genomic DNA library with a virus-derived probe under conditions of reduced stringency (*3*), and it was mapped to chromosome 14q32, proximal to the immunoglobulin heavy chain locus (*4*), a region frequently affected by translocations and inversions in human T cell leukemia or lymphoma, mixed lineage childhood leukemia, and clonal T cell proliferations in ataxia telangiectasia (*5*). The putative *AKT* gene was amplified in a human gastric carcinoma (*3*). The molecular characterization of a nondefective AKT8 virus clone presented here reveals that Akt is a protein kinase C (PKC)–related serine-threonine kinase whose noncatalytic domain contains a Src homology–2 (SH2)–like region. This structural feature suggests that Akt may form a functional link between tyrosine and serine-threonine phosphorylation pathways.

We cloned the AKT8 proviral DNA from AKT8-transformed mink lung cells using a Moloney murine leukemia virus (M-MuLV) long terminal repeat (LTR) probe. A restriction map of the integrated AKT8 provirus is shown in Fig. 1. Hybridization of restriction endonuclease–digested AKT8 DNA to a M-MuLV probe representing the entire viral genome identified a nonhybridizing region of possible cellular origin included in a 3.5-kb Bgl II–Sma I DNA fragment. Sequence analysis of this fragment, which we expected to contain the transduced cellular oncogene (Fig. 2), revealed that AKT8 has sequences from a mink cell focus-forming virus and the gene encoding Akt. The 5′ recombination breakpoint maps at nucleotide 785 from the *gag* ATG codon in the region coding for the capsid protein p30. The 3′ recombination breakpoint maps at nucleotide 298 from the *env* ATG codon in the region coding for the *env* gene product gp70. The v-*akt* gene codes for a 763–amino acid protein (86 kD) generated by the fu-

Department of Medical Oncology, Fox Chase Cancer Center, Philadelphia, PA 19111.

*Present address: Kaiser Permanente, Landover, MD 20785.
†To whom correspondence should be addressed.

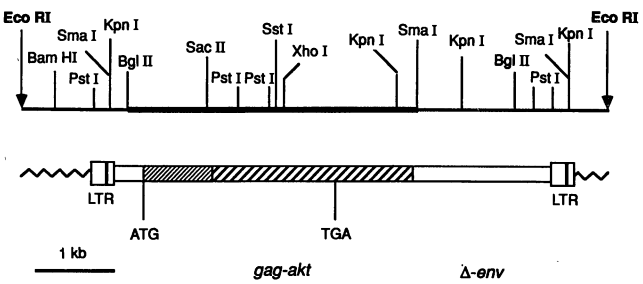

Fig. 1. Restriction endonuclease map and structural organization of the AKT8 provirus. (**Top**) Restriction map of the clone λAKT8-AB obtained by screening a partial genomic library of Eco RI–digested DNA from AKT8-transformed mink lung cells. The Bgl II–Sma I fragment shown in bold was sequenced. (**Bottom**) Structure of the AKT8 provirus. The wavy lines at both ends represent the cellular DNA sequences flanking the provirus. The hatched bar represents v-*akt*, which is subdivided into 5′ *gag*-derived and 3′ c-*akt*–derived segments. ATG and TGA define the beginning and the end of the v-*akt* open reading frame.