sequences they reported, we conclude that for the sequences described here two purines can exist in a two-base loop without significant destabilization. More detailed discussion of stereochemistry is given in the following paper by Raghunathan et al. (1991).

Registry No. GGTACIAGTACC, 130641-68-8; GGTACGCGTACC, 113341-04-1; GGTAAGCGTACC, 130641-71-3; GGTAAGCITACC, 130668-56-3; GGTACGAGTACC, 130641-72-4; guanine, 73-40-5; hypoxanthine, 68-94-0; adenine, 73-24-5.

REFERENCES

Aboul-ela, F., Koh, D., & Tinoco, I., Jr. (1985) *Nucleic Acids Res. 13*, 4811-4824.

Atkinson, T., & Smith, M. (1984) in *Oligonucleotide Synthesis. A Practical Approach* (Gait, M. J., Ed.) IRL, Oxford, England.

Blommers, M. J. J., Walters, J. A. L. I., Haasnoot, C. A. G., Aelen, J. M. A., van der Marel, G. A., van Boom, J. H., & Hilbers, C. W. (1987) *Biochemistry 28*, 7491-7498.

Breslauer, K. J., Frank, R., Blocker, H., & Marky, L. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3746-3750.

Haasnoot, C. A. G., Hilbers, C. W., van der Marel, G. A., van Boom, J. H., Singh, U. C., Pattabiraman, N., & Kollman, P. A. (1986) *J. Biomol. Struct. Dyn. 3*, 843-857.

Haasnoot, C. A. G., Blommers, M. J. J., & Hilbers, C. W. (1987) in *Structure, Dynamics, and Function of Biomolecules* (Ehrenberg, A., Rigler, R., Gräslund, A., & Nilsson, L., Eds.) Springer Series in Biophysics, Vol. 1, pp 212-216, Springer-Verlag, Berlin.

Hilbers, C. W., Haasnoot, C. A. G., de Bruin, S. H., Joordens, J. J. M., van der Marel, G. A., & van Boom, J. H. (1985) *Biochemie 67*, 685-695.

Howard, F. B., Hattori, M., Frazier, & Miles, H. T. (1977) *Biochemistry 16*, 4637-4646.

McBride, L. J., & Caruthers, M. H. (1983) *Tetrahedron Lett. 24*, 245-248.

Muraoka, M., Miles, H. T., & Howard, F. B. (1980) *Biochemistry 19*, 2429-2439.

Orbons, L. P. M., van Benzekom, A. A., & Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 965-987.

Powell, J. I., Fico, R., Jennings, W. H., O'Bryan, E. R., & Schultz, A. R. (1980) *Proceedings of the IEEE Computer Society, 21st International Conference*, pp 185-190, IEEE Computer Society, Los Angeles, CA.

Privê, G. G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M. L., & Dickerson, R. E. (1987) *Science 238*, 498-504.

Raghunathan, G., Jernigan, R. L., Miles, H. T., & Sasisekharan, V. (1991) *Biochemistry* (following paper in this issue).

Scheffler, I. E., Elson, E. L., & Baldwin, R. L. (1968) *J. Mol. Biol. 36*, 291-304.

Scheffler, I. E., Elson, E. L., & Baldwin, R. L. (1970) *J. Mol. Biol. 48*, 145-171.

Sinha, N. D., Biernat, J., McManus, J., & Kaster, H. (1984) *Nucleic Acids Res. 12*, 4539-4557.

Tinoco, I., Jr., Uhlenbeck, O. C., & Levine, M. D. (1971) *Nature 230*, 362-367.

Xodo, L. E., Manzini, G., Quadrifoglio, F., van der Marel, G., & van Boom, J. H. (1988) *Biochemistry 27*, 6321-6326.

Conformational Feasibility of a Hairpin with Two Purines in the Loop. 5′-d-GGTAC*IA*GTACC-3′

G. Raghunathan,[‡] R. L. Jernigan,[‡] H. Todd Miles,*[,§] and V. Sasisekharan[§,∥]

Laboratory of Mathematical Biology, National Cancer Institute, Building 10, Room 4B56, and Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, Building 2, Room 201, National Institutes of Health, Bethesda, Maryland 20892

Received June 28, 1990; Revised Manuscript Received September 12, 1990

ABSTRACT: Structural feasibility and conformational requirements for the sequence 5′-d-GGTAC*IA*GTACC-3′ to adopt a hairpin loop with I6 and A7 in the loop are studied. It is shown that a hairpin loop containing only two nucleotides can readily be formed without any unusual torsional angles. Stacking is continued on the 5′-side of the loop, with the I6 stacked upon C5. The base A7, on the 3′-side of the loop, can either be partially stacked with I6 or stick outside without stacking. Loop closure can be achieved for both syn and anti conformations of the glycosidic torsions for G8 while maintaining the normal Watson–Crick base pairing with the opposite C5. All torsional angles in the stem fall within the standard B-family of DNA helical structures. The phosphodiesters of the loop have trans,trans conformations. Loop formation might require the torsion about the C4′–C5′ bond of G8 to be trans as opposed to the gauche$^+$ observed in B-DNA. These results are discussed in relation to melting temperature studies [Howard et al. (1991) *Biochemistry* (preceding paper in this issue)] that suggest the formation of very stable hairpin structures for this sequence.

The preceding paper (Howard et al., 1991) discussed the variations in the thermal stabilities of oligonucleotide duplexes due to the change in the positions of the mispairs. In comparison with the duplex d-GGTACGCGTACC, small decreases in the melting temperatures were observed for the sequences d-GGTAC*IA*GTACC and d-GGTAC*GA*GTACC and significantly larger decreases were observed for the sequences d-GGTA*A*GCGTACC and d-GGTA*A*GC*I*TACC. These and other observations reported there were interpreted

[‡] Laboratory of Mathematical Biology, NCI.

[§] Laboratory of Molecular Biology, NIDDK.

[∥] Scholar-in-Residence, FIC, National Institutes of Health, Bethesda, MD 20892. Permanent address: Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India.

in terms of a two-base loop for the former two sequences and a four-base loop for the latter two sequences. The present study was undertaken to explore the stereochemical and energetic feasibility of the two-base loop in the sequence 5′-d-GGTAC*IA*GTACC.

Hairpin loops clearly play an important structural role in the tertiary folding of RNA molecules. While they may be relatively less common in DNA, they are necessary for various genetic processes (Hobom et al., 1979; Muller & Fitch, 1982; Weaver & DePamphilis, 1984) and could play an essential role in the proper positioning and recognition of DNA by regulatory proteins. The structural and functional significance of this class of structures have in some ways provided the impetus for physicochemical characterization of several oligonucleotides containing self-complementary, partially self-complementary, and mispaired sequences. The size and stability of hairpin loops has been studied by various physical observations and theoretical calculations. Thermodynamic studies of RNA loops (Tinoco et al., 1971; Gralla & Crothers, 1973; Uhlenbeck et al., 1973; Groebe & Uhlenbeck, 1988) have indicated that at least four residues are required for the stability of hairpin loops. Again, recent physical analyses of synthetic deoxyoligonucleotide hairpins of varying loop sizes (Hilbers et al., 1985) and molecular modeling studies (Haasnoot et al., 1986) suggest that four or five nucleotides are optimal to form a DNA loop. Various techniques such as thermal stability measurements, spectroscopy, calorimetry, 2D NMR, and X-ray crystallography have been used to study hairpin helices with four or more bases in the loop (Senior et al., 1988; Chattopadhyaya et al., 1988; Garcia et al., 1988; Williams & Boxer, 1988; Hare & Reid, 1986; Pramanik et al., 1988; Benight et al., 1989), and there have also been reports of three-base loops for DNA (Summers et al., 1985) and RNA (Puglisi et al., 1990). Most of these oligomers have four thymines in the loop. A molecular mechanics study of the hairpins with an intervening loop sequence -TTTT- by Haasnoot et al. (1987) suggests a wobble base pair between the first and last thymines in such a loop. When T and A or C and G replaced the first and last positions of the four-base loop, then a two base hairpin loop can be achieved only by going to an unusual conformation with a highly buckled base pair at the loop end of the stem. In the -TTTA- loop, the two hydrogens attached to the N6 of A10 are hydrogen bonded to O2s of T8 and T9. The modeling studies of a mismatched DNA octamer containing 5-methylcytosine by Orborns et al. (1987) indicates the feasibility of a two-base loop; however, the loop conformation had an unusual conformation angle. Recently, Blommers et al. (1989) have made various changes in the intervening loop sequence -TTTT- and studied them by using NMR and UV techniques. NMR data indicate a Hoogsteen AT pair for the loop -TTTA- and adenine has the syn conformation. In the case of the loop -CTTG-, they agree with the earlier calculations of Haasnoot et al. (1987) and suggest that the GC pair would be highly buckled and hence indicate a strained conformation. They further observe that when the bases in the middle of the sequence, viz., TT, are replaced by bulky bases, the hairpin tends to become a four-base loop. There have also been reports in which the observed thermal stabilities and corresponding calculated enthalpy changes have been attributed to the formation of two-base hairpin molecules (Xodo et al., 1988; Chen, 1989). It should be informative to study the conformational features of hairpin molecules with short loops and to interpret the dependence of their thermal stabilities on base conformation and sequence.

In this paper we have made a thorough examination of the conformational feasibility of a two-base loop for the sequence d-GGTAC*IA*GTACC. It is shown that C5 and G8 can form normal Watson-Crick hydrogen bonds that are not buckled, and G8 can have either anti or syn conformation. While there are some general characteristics of the stacking pattern of the two bases in the loop, the loop formation is independent of the nature of the bases in the loop. The models are based on careful molecular modeling and molecular mechanics energy refinements. All the torsional angles are within the allowed conformational ranges and have been previously observed in oligonucleotide crystal structures.

Materials and Methods

An initial B-DNA duplex structure with the sequence

5′-d-G	G	T	A	C	T-3′
1	2	3	4	5	6
3′-d-C	C	A	T	G	A-5′
12	11	10	9	8	7

was generated by using fiber diffraction coordinates (Arnott et al., 1976). The actual coordinates used in the present study were kindly provided by Dr. Chandrasekaran of Purdue University. Then, T6 was replaced by inosine by superposition. The chain was extended beyond the TA base pair, and molecular editing was done to satisfy the chemistry of chain continuity for the formation of the hairpin loop. Preliminary structures satisfying the conditions of loop closure were generated by varying the torsional angles in the neighborhood of the loop. This was done with the program MOGLI on an Evans and Sutherland graphics system. Final pictures were made on a Silicon Graphics system with the program QUANTA. Initially, the glycosidic torsion of G8 was modeled to be in the syn conformation. To retain the Watson–Crick hydrogen bonding between G8 and C5 required some changes in the torsional angles of residues 8 and 9 in the stem of the hairpin. In the next stage, the torsional angles in the loop were varied for chain closure. Three different starting conformations were generated for the syn and three more for the anti conformations of the G8 glycosidic angle. The structures were energy minimized in the Cartesian space with CHARMM by using the parameters for nucleic acids including all hydrogens (Nilsson & Karplus, 1986). The adapted basis Newton–Raphson method was used for minimizing the energies. The torsional angles of the residues in the loop were constrained successively in an iterative manner until we arrived at structures that have all the torsional angles in the allowed ranges. A dielectric constant of 4.0 was used. Variations in the structures with changes in the dielectric constant were examined; however, the choice of dielectric constant was not found to be critical for the present study, viz., the stereochemical feasibility for the formation of a two-base hairpin loop for this sequence.

Figure 1 gives the notation of the torsional angles used.

Results

Of the three energy-minimized hairpin models for the anti conformation of G8, two of them have minimal structural changes. Similarly, of the three structures calculated with the syn conformation for G8, two of them varied as little as 0.25 in rms deviation for all the atoms, including hydrogens. Hence only four structures will be discussed, two for the G8 having the anti conformation, designated as HA1 and HA2, and two for the G8 having the syn conformation, designated as HS1 and HS2.

(*1*) *Structures HA1 and HA2 with Anti Glycosidic Torsion for G8.* Stereoviews of the structures HA1 and HA2 are

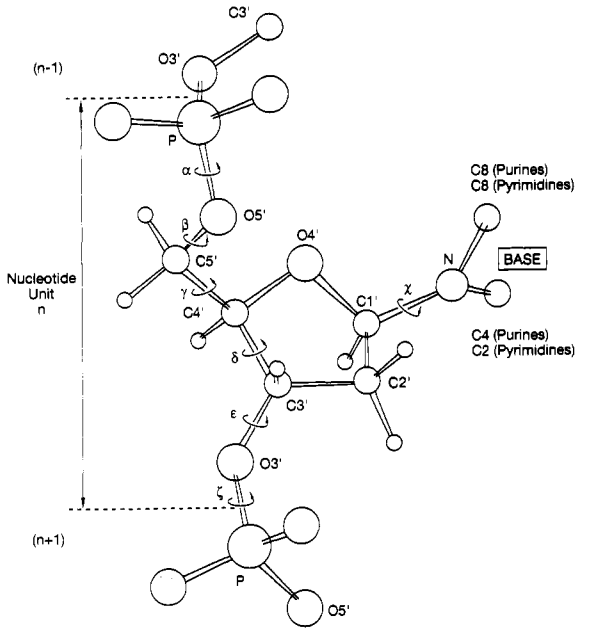

FIGURE 1: Designation of torsional angles for a polynucleotide chain.

Table I: Torsional Angles (in Degrees) for the Conformation HA1

residues	torsional angles α P–O5′	β O5′–C5′	γ C5′–C4′	δ C4′–C3′	ε C3′–O3′	ζ O3′–P	χ C1′–N9
G1	–	–	59	107	214	258	16
G2	306	138	62	118	204	217	46
T3	314	137	62	115	194	258	47
A4	317	138	62	98	212	268	29
C5	298	137	60	88	191	259	58
I6	303	137	65	126	179	268	68
A7	190	174	185	130	179	230	47
G8	221	136	177	94	180	233	41
T9	321	137	64	100	181	260	43
A10	327	137	63	100	200	263	23
C11	312	137	62	88	176	254	31
C12	332	137	64	150	–	–	51

Table II: Torsional Angles (in Degrees) for the Conformation HA2

residues	torsional angles α P–O5′	β O5′–C5′	γ C5′–C4′	δ C4′–C3′	ε C3′–O3′	ζ O3′–P	χ C1′–N9
G1	–	–	59	104	204	258	28
G2	311	137	62	98	193	256	34
T3	318	137	63	99	188	259	39
A4	323	138	62	98	206	264	19
C5	309	137	64	115	194	262	51
I6	307	138	62	129	295	183	70
A7	203	175	184	96	180	179	42
G8	207	136	183	91	180	204	41
T9	315	136	66	142	193	255	65
A10	319	138	61	94	209	264	15
C11	306	137	61	86	175	255	34
C12	332	137	64	150	–	–	52

shown in Figure 2, panels a and b, respectively. Of the two bases in the loop, viz., I6 and A7, I6 is stacked upon C5 in the stem in both HA1 and HA2. A7 is partially stacked with I6 in HA1, whereas it sticks out completely in HA2. All backbone torsional angles (Tables I and II for HA1 and HA2) in the stem for both the structures have values characteristic of B-DNA. In both structures, C5′–C4′ of A7 and G8 are trans, as opposed to gauche$^+$ for the other nucleotides. Also, the values of C5′–O5′ for A7 are 174° and 175° and are higher than those of the other residues. The phosphodiester conformation about O3′–P of A7 and P–O5′ of G8 is trans,trans. These conformational changes are needed for loop closure and

Table III: Torsional Angles (in Degrees) for the Conformation HS1

residues	torsional angles α P–O5′	β O5′–C5′	γ C5′–C4′	δ C4′–C3′	ε C3′–O3′	ζ O3′–P	χ C1′–N9
G1	–	–	59	133	240	277	351
G2	288	138	58	121	197	229	57
T3	315	137	62	104	190	258	40
A4	321	138	61	97	221	274	21
C5	292	137	59	90	191	256	65
I6	300	137	65	128	179	266	73
A7	205	174	186	130	180	206	38
G8	155	180	181	136	180	186	158
T9	307	136	66	142	190	256	67
A10	322	138	62	94	205	260	10
C11	314	137	64	133	192	200	48
C12	334	137	63	157	–	–	49

Table IV: Torsional Angles (in Degrees) for the Conformation HS2

residues	torsional angles α P–O5′	β O5′–C5′	γ C5′–C4′	δ C4′–C3′	ε C3′–O3′	ζ O3′–P′	χ C1′–N9
G1	–	–	59	106	201	261	25
G2	313	137	63	102	192	253	33
T3	317	137	63	102	197	252	33
A4	314	138	64	112	208	269	27
C5	305	137	61	99	192	263	43
I6	313	138	63	110	254	180	48
A7	277	174	179	97	180	180	47
G8	171	137	180	91	180	185	159
T9	304	136	65	143	191	257	64
A10	319	137	62	97	197	263	21
C11	315	137	64	98	166	249	28
C12	347	137	64	158	–	–	37

account for the structural features of the bases in the loop. Sugar pucker of A7 is C2′-endo for HA1 and is C3′-endo for HA2. The observed structural changes between HA1 and HA2 also arise from the the changes in the torsional angles of the adjacent bonds, C3′–O3′ of I6 and O3′–P of I6.

(2) Structures HS1 and HS2 with Syn Glycosidic Torsion for G8. The structures HS1 and HS2 are shown in Figure 3, panels a and b, respectively, and the corresponding torsional angles are given in Tables III and IV. As in the case of HA1 and HA2, the torsional angles about C5′–C4′ of A7 and G8, O3′–P of A7, and P–O5′ of G8 are trans. The values of the torsional angle about O3′–P of G8 (185° and 186°) are relatively lower than those of the residues in the stem, and this is in part responsible for restoring the Watson–Crick base pairing for G8, which now has the syn conformation. Comparing HS1 and HS2, the torsional angle about O5′–C5′ of G8 is 180° in HS1 whereas it is 137° in HS2. The sugar puckers for the residues 6, 7, and 8 in HS1 and HS2 are different. There are also correlated changes in the torsional angles O3′–P of I6 and P–O5′ of A7.

Discussion

Molecular modeling and calculations presented in this paper demonstrate the facile formation of hairpin structures with two purines as a two-base loop on a B-DNA stem, a plausible suggestion in the previous paper (Howard et al., 1991) based on the observed differences in thermal stabilities for various mispaired oligonucleotides. This study provides a conformational rationale for loop folding in DNA.

The helical stem is characteristic of B-DNA, as is evident in the values of torsional angles for the four energy-minimized structures shown in Tables I–IV. None of the various models discussed in the present study contains any unusual torsional angles, whereas the previous modeling studies of a two-base hairpin loop (Orborns et al., 1987) contain a gauche$^+$ torsion about a C5′–O5′ bond that is neither experimentally observed

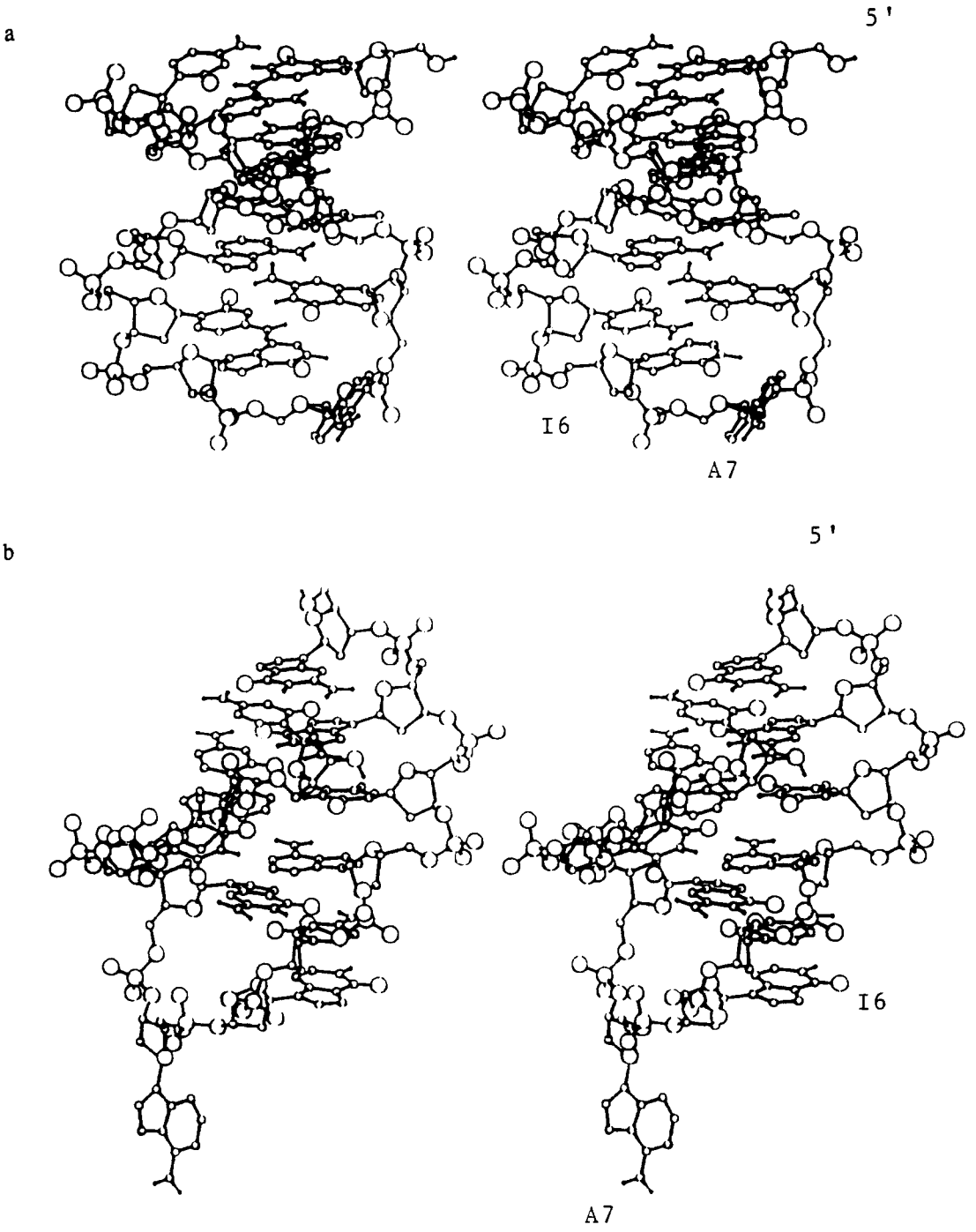

FIGURE 2: (a) Stereoview of the conformation HA1 as viewed into the major groove. The sequence is 5'-d-GGTAC*IA*GTACC. G8 is in the anti conformation; I6 is stacked with C5, whereas A7 is partially stacked with I6; G8 and C5 are Watson–Crick hydrogen bonded. (b) Stereoview of the conformation HA2 as viewed into the minor groove. G8 is in the anti conformation. While I6 is stacked upon C5, A7 sticks out.

nor theoretically allowed [Lakshminarayanan & Sasisekharan, 1969; Pullman et al., 1972; Saran & Govil, 1971; see also Saenger (1984)]. Contrary to the previous studies (Haasnoot et al., 1986; Blommers et al., 1989), the present calculations suggest that the last base pair in the helical stem, viz., C5-G8, can be Watson–Crick hydrogen bonded without significant buckle.

The sequence examined in the present study has two purines, viz., I6 and A7, in the loop. Thus, our results contradict the earlier proposal by Blommers et al. (1989), who suggested that replacement of the central thymines in the -TTTT- loop by bulkier purine residues would result in a hairpin with a four-base loop. The present calculations indicate that it need not be so. The accommodation of the bulky purines in the loop suggests that hairpin formation with a two-residue loop could be independent of the nature of the residues in the loop.

In all four energy-minimized structures, I6 is stacked upon C5. This pattern is independent of the glycosidic torsion of G8, which could be either syn or anti. This feature is seen in Figures 2 and 3 and is also evident from the interaction energy values between the bases given in Table V. The stacking of the base at the 5'-side of the loop seems to be a natural consequence of the stereochemical requirements of loop folding. The base on the 3'-side of the loop, viz., A7, either could be partially stacked with I6, as in HA1 and HS1 (Figures 2a and 3a), or could be looped out completely, as in HA2 and HS2 (Figures 2b and 3b). The partial stacking of the bases in the loop in HA1 and HS1 is further stabilized by a hydrogen bond between N1H of I6 and N3 of A7. The differences in the stacking pattern of the loop bases, viz., A6

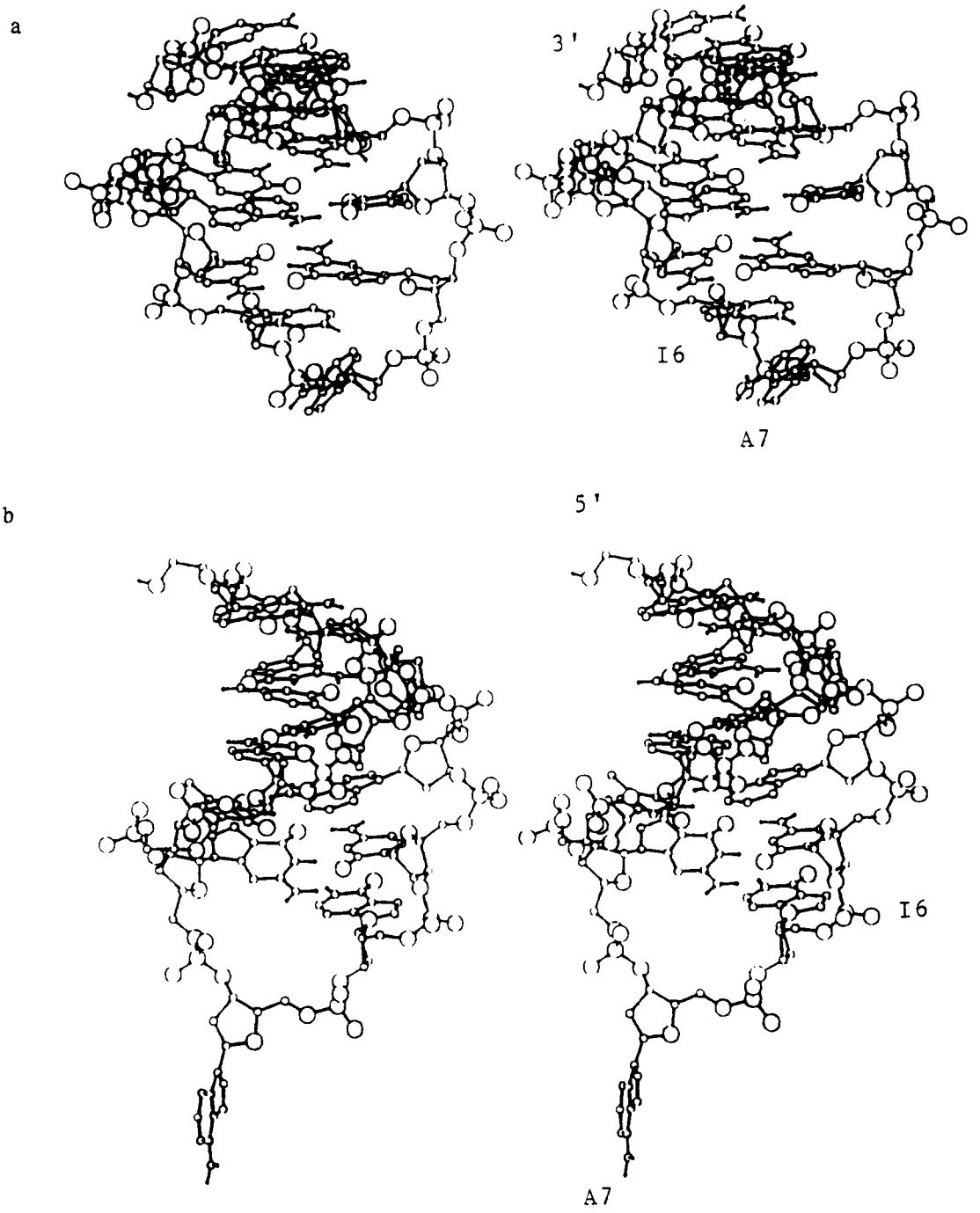

FIGURE 3: (a) Stereoview of the conformation HS1, looking into the major groove. The glysosidic torsion of G8 is syn. Normal Watson–Crick hydrogen bonding is maintained between G8 and C5. Stacking is continued on the 5′-side of the loop, with I6 stacked upon C5. (b) Stereoview of the conformation HS2 looking into the minor groove. G8 has a syn glycosidic torsional angle. The other structural features are similar to the conformation HA2 in Figure 2b.

and I7, between the pairs of structures HA1 and HS1 and HA2 and HS2 are within the general characteristics of loop folding, and these variations are evidence of some possible flexibility in short two-base loops. The stacking continuity at the 5′-side of the loop in DNA hairpins is reminiscent of, but opposite in direction to, the stacking in tRNA molecules, where stacking continuity has been observed at the 3′-side of the loop. This type of base stacking pattern in DNA hairpin loops has also been observed by Haasnoot et al. (1987).

The usual distance between O3′ at the 3′-end of one strand and O3′ attached to the phosphate at the 5′-end of the complementary strand in B-DNA is 15.9 Å. This distance is too large to be linked by two nucleotides, even with extended phosphodiester conformations and variations in sugar puckerings. However, the corresponding distance between O3′ of C5 and O3′ attached to the phosphate of G8 in the four energy-minimized models HA1, HA2, HS1, and HS2 is 12.1, 12.3, 11.6, and 10.8 Å, respectively. This shorter distance is achieved in part due to the torsional angle about C4′–C5′ bond of G8, which is trans in all four models (see Tables I–IV). It is a significant departure from the corresponding value of gauche$^+$ for this torsional angle observed in B-DNA and the residues in the hairpin stem. Thus, this bond might act as a hinge about which the rest of the loop could swing for loop closure. Taken together, the rotation about C4′–C5′ bond at the 3′-side of the loop and the base stacking at the 5′-side could be essential factors for the folding of very short loops. The stereochemistry of loop formation also requires the torsional

Table V: Base Stacking and Base Pairing Energies for the Conformations HA1 and HA2[a]

	HA1	HA2
base pairing		
G1-C12	−19.9	−19.9
G2-C11	−19.8	−19.8
T3-A10	−12.4	−12.3
A4-T9	−12.4	−12.3
C5-G8	−19.9	−19.6
base stacking		
G1/G2	−5.7	−5.7
G2/T3	−6.4	−6.5
T3/A4	−5.8	−5.4
A4/C5	−5.3	−5.4
C5/I6	−5.1	−5.5
I6/A7	−8.0	0.0
A7/G8	−1.8	0.0
G8/T9	−7.2	−6.8
T9/A10	−5.4	−5.6
A10/C11	−5.6	−5.6
C11/C12	−3.1	−3.2

[a] Base pairing and stacking energies have been computed with CHARMM by using the parameters of Nilson and Karplus (1986) and are given in kilocalories per mole. 0.0 corresponds to no stacking between bases. The values emphasize the relative stabilities of the various base pairing and base stacking arrangements.

angles O5′–C5′ of A7 to be trans and the phosphodiester conformations about O3′–P of I6, P–O5′ of A7, O3′–P of A7, and P–O5′ of G8 to be trans,trans conformations.

Both syn and anti conformations for G8 are found to be possible. It may be emphasized that syn glycosidic torsional angles for guanosine have been observed in several crystals of oligonucleotides and polymers (Drew et al., 1980; Wang et al., 1979; Arnott et al., 1980). The preference of the syn conformation of guanosine has also been found in physicochemical measurements (Guschlbauer et al., 1972) and energy calculations (Yathindra & Sundaralingam, 1973; Olson, 1973; Lespinasse et al., 1974; Berthod & Pullman, 1973). As discussed earlier, the distances between the O3′s on the opposite strands in the two G8-syn structures are 10.8 and 11.6 Å, while the distances in the two G8-anti structures are somewhat larger, at 12.1 and 12.3 Å. Thus, though it might be possible to accommodate conformations for both families of structures through local rearrangements, we suggest that it might be easier to close the loop with G8-syn structures due to the shorter distance between the O3′s on the opposite strands in these cases.

Added in Proof

A recent report contains further relevant information on two-base hairpin loops (Cheong et al., 1990). An RNA hairpin has a reverse wobble GU pair at the internal terminus of the stem, thus leaving an effective two-base loop. The central two nucleotides U6 and C7 have extended B-form DNA-like C2′-endo sugar puckers. G8 has a syn conformation, and the G8·U5 base pair is buckled. The C4′–C5′ bond of G8 has a trans conformation, consistent with our models. The phosphodiester in the loop between residues C7 and G8 has a gauche$^+$,gauche$^+$ conformation, whereas the corresponding phosphodiester in all our models has a trans,trans conformation. The hairpins with two-base loops, the one on the A-RNA stem cited above as well as the ones on the B-DNA stem presented in this paper, appear to be energetically stable.

Acknowledgments

We thank Dr. James V. Silverton of the Laboratory of Biophysical Chemistry, National Heart, Lung, and Blood Institute, NIH, for help in making Figure 1 by using the PLUTO program (Motherwell, 1976).

Registry No. 5′-d-GGTACIAGTACC-3′, 130641-68-8; 5′-d-GGTACT-3′, 130641-69-9; 3′-d-CCATGA-5′, 130641-70-2; guanine, 73-40-5.

References

Arnott, S., Campbell-Smith, P., & Chandrasekaran, R. (1976) in *CRC Handbook of Biochemistry and Molecular Biology* (Fasman, G. D., Ed.) Vol. 2, pp 411–422, CRC Press, Cleveland, OH.

Arnott, S., Chandrasekaran, R., Birdsall, D. L., Leslie, A. G. W., & Ratcliff, R. L. (1980) *Nature 283*, 743–745.

Benight, A. S., Wang, Y., Amaratunga, M., Chattopadhyaya, R., Henderson, J., Hanlon, S., & Itakura, S. (1989) *Biochemistry 8*, 3323–3332.

Berthod, H., & Pullman, B. (1973) *FEBS Lett. 3*, 146–150.

Blommers, M. J. J., Walters, J. A. L. I., Haasnoot, C. A. G., Aelen, J. M. A., van der Marel, G. A., van Boom, J. H., & Hilbers, C. W. (1989) *Biochemistry 28*, 7491–7498.

Broyde, S. B., Wartell, R. M., Stellman, S. D., Hingerty, B., & Langridge, R. (1975) *Biopolymers 14*, 1597–1613.

Chattopadhyaya, R., Ikuta, S., Greskowiak, K., & Dickerson, R. E. (1988) *Nature 334*, 175–179.

Chen, F. M. (1989) *J. Biomol. Struct. Dyn. 6*, 1239–1257.

Cheong, C., Varani, G., & Tinoco, I., Jr. (1990) *Nature 346*, 680–682.

Drew, H., Takano, T., Tanaka, S., Itakura, K., & Dickerson, R. E. (1980) *Nature 286*, 567–573.

Garcia, A. E., Gupta, G., Sarma, M. H., & Sarma, R. H. (1988) *J. Biomol. Struct. Dyn. 6*, 525–542.

Groebe, D. R., & Uhlenbeck, O. C. (1988) *Nucleic Acids Res. 16*, 11725–11735.

Guschlbauer, W., Fric, I., & Holy, A. (1972) *Eur. J. Biochem. 31*, 1–13.

Haasnoot, C. A. G., Blommers, M. J. J., & Hilbers, C. W. (1987) in *Structure, Dynamics and Function of Biomolecules* (Ehrenberg, A., Rigler, R., Gräslund, A., & Nilsson, L., Eds.) Springer Series in Biophysics, Vol. 1, pp 212–216, Springer-Verlag, Berlin.

Haasnoot, C. A. G., Hilbers, C. W., van der Marel, G. A., van Boom, J. H., Singh, U. C., Pattabiraman, N., & Kollman, P. A. (1986) *J. Biomol. Struct. Dyn. 3*, 843–857.

Hare, D. R., & Reid, B. R. (1986) *Biochemistry 25*, 5341–5350.

Hilbers, C. W., Haasnoot, C. A. G., de Bruin, S. H., Joordens, J. J. M., van der Marel, G. A., & van Boom, J. H. (1985) *Biochimie 67*, 685–695.

Hobom, G., Grosschedl, R., Lusky, M., Scherer, G., Schwartz, E., & Kossel, H. (1979) *Cold Spring Harbor Symp. Quant. Biol. 43*, 165–178.

Howard, F. B., Chen, C., Ross, P. D., & Miles, H. T. (1991) *Biochemistry* (preceding paper in this issue).

Lakshminarayanan, A. V., & Sasisekharan, V. (1969) *Biopolymers 8*, 489–503.

Lespinasse, J.-N., Broch, H., Cornillon, R., & Vasilescu, D. (1976) *J. Theor. Biol. 57*, 225–230.

Motherwell, W. D. S. (1976) PLUTO: Program for plotting molecular and cyrstal structures. University of Cambridge, Cambridge, U.K.

Muller, U. R., & Fitch, W. M. (1982) *Nature 298*, 582–585.

Olson, W. K. (1973) *Biopolymers 12*, 1787–1814.

Nilsson, L., & Karplus, M. (1986) *J. Comput. Chem. 7*, 591–616.

Orborns, L. P. M., van Beuzekom, A. A., & Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 965–987.

Pramanik, P., Kanhouwa, N., & Kan, L. S. (1988) *Biochemistry 27*, 3024-3031.
Puglisi, J. D., Wyatt, J. R., & Tinoco, I., Jr. (1990) *Biochemistry 29*, 4215-4226.
Pullman, B., Perahia, D., & Saran, A. (1972) *Biochim. Biophys. Acta 269*, 1-14.
Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, Berlin.
Saran, A., & Govil, G. (1971) *J. Theor. Biol. 33*, 407-418.
Senior, M. M., Jones, R. A., & Breslauer, K. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6242-6246.
Tinoco, I., Jr., Uhlenbeck, O. C., & Levine, M. D. (1971) *Nature 230*, 362-367.
Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G., & Rich, A. (1979) *Nature 282*, 680-686.
Weaver, D. T., & DePamphillis, M. L. (1984) *J. Mol. Biol. 180*, 961-986.
Williamson, J. R., & Boxer, S. G. (1989) *Biochemistry 28*, 2819-2831.
Xodo, L. M., Manzini, G., Quadrifoglio, F., van der Marel, G. A., & van Boom, J. H. (1988) *Biochemistry 27*, 6321-6326.
Yathindra, N., & Sundaralingam, M. (1973) *Biopolymers 12*, 2075-2082.

A Novel Combined Chemical–Enzymatic Synthesis of Cross-Linked DNA Using a Nucleoside Triphosphate Analogue†

Marlon Cowart and Stephen J. Benkovic*
Department of Chemistry, The Pennsylvania State University, 152 Davey Laboratory, University Park, Pennsylvania 16802
Received May 2, 1990; Revised Manuscript Received September 18, 1990

ABSTRACT: A novel method using combined chemical and enzymatic reactions to allow the preparation of covalently cross-linked DNA duplexes has been described. The method can be used to specifically link two complementary bases of a DNA duplex containing all four natural bases. The modified nucleotide 9-(2-deoxy-5-*O*-triphospho-β-D-ribofuranosyl)-N^6,N^6-ethano-2,6-diaminopurine (6edDTP) was prepared by total chemical synthesis and was found to be incorporated into DNA duplexes in the place of 2′-deoxyguanosine 5′-*O*-triphosphate by the Klenow fragment of *Escherichia coli* DNA polymerase I, T4 and T7 DNA polymerases, avian myeloma virus reverse transcriptase, and rat DNA polymerase β. Once incorporated, the aziridine of the nucleotide is rapidly opened by the N4 of the cytosine on the complementary strand to give cross-linked DNA, where the modified nucleotide is covalently joined to the complementary base by an ethano linkage. The duplexes produced were found to be recognized as substrates by various DNA polymerases. The K_m for the incorporation of the 6edDTP into DNA catalyzed by the Klenow fragment of *E. coli* DNA polymerase I was found to be 29 μM, and the k_{cat} was found to be 0.014 s^{-1}. The modified nucleoside also served as a substrate for terminal deoxynucleotidyltransferase, where it was added to single-stranded DNA and then hybridized to a complementary strand, after which cross-linking of the two strands occurred within 1 min.

The ability to covalently cross-link polynucleotide duplexes is of considerable biochemical and clinical importance. The mechanisms of action of various enzymes (such as integrases, recombinases, restriction enzymes, and exonucleases) on DNA duplexes require local melting or distortion of the duplexes for activity. The study of the interaction of these enzymes with DNA would be aided by the availability of covalently cross-linked DNA duplexes, if such duplexes could be prepared to resemble closely normal DNA, except for the cross-linked bases. The interaction of the exonuclease activity of the Klenow fragment of *Escherichia coli* DNA polymerase I has been studied, for example, by the action of the enzyme on duplexes containing a cross-linked cytosine–cytosine base pair (Cowart et al., 1989). It was found that at least four base pairs of the primer terminus must melt out in order for the enzyme to be able to remove a base pair through exonuclease activity. Other polymerases had differing requirements.

The cytotoxicity of many antitumor compounds is thought to be due to their ability to induce interstand cross-linking of cellular DNA (Mitchell et al., 1989). DNA cross-linked by naturally occurring compounds like psoralens or the clinically important synthetic bifunctional alkylating agents becomes resistant to in vivo repair, and these lesions impair the ability of the target to undergo transcription. This has been implicated as the mechanism of efficacy for some compounds active against malignant cells, and as a result, various cross-linked DNA duplexes have been synthesized for study (Kohn, 1979).

The ability of single-stranded polynucleotides to recognize and bind to their complementary strands forms the basis for various exciting clinical and biological applications (Knorre & Vlassov, 1985). A DNA oligomer containing an alkylating agent could bind to its complementary messenger RNA and, once covalently cross-linked, shut down the synthesis of the corresponding protein. The inhibition of protein biosynthesis with antisense oligonucleotides has been reported (Zamecnik & Stevenson, 1978; Jayaraman et al., 1981), but the ability to form a covalent bond with their complementary oligonucleotides might make them more efficient in this process. This principle has been demonstrated by Vlassov et al. (1984), who showed that antibody production in a mouse myeloma

† This research was supported by grants from the National Institutes of Health, General Medical Sciences, 2R01 GM 13306 (S.J.B.), and M.C. was supported in part by National Service Research Award 1 F 32 GM 12649 BI-2 from the National Institute of General Medical Sciences.
* Correspondence should be addressed to this author.

0006-2960/91/0430-788$02.50/0