
DNA stem-loop structures in oligopurine-oligopyrimidine triplexes

Stephen C. Harvey*, Jia Luo and Richard Lavery¹

Department of Biochemistry, University of Alabama, Birmingham, AL 35294, USA and ¹Institut de Biologie Physico-Chimique, Fondation Edmond de Rothschild, 13 rue Pierre et Marie Curie, 75005 Paris, France

Received July 1, 1988; Revised and Accepted November 13, 1988

ABSTRACT

Closed circular DNA containing polypurine-polypyrimidine sequences can adopt a triple helical stem-loop structure under supercoiling pressure. We describe an automated procedure for building model loops and its application to the investigation of the polypyrimidine loop at the end of such a triple helical stem. All possible combinations of 3'-stacked and 5'-stacked structures have been examined for loops containing three, four, five, and six nucleotides. The lowest energy conformation is a four-membered loop with all bases stacked on the strand at the 3' end of the loop. The model predicts that sequences $(GA)_n$, $(GGGA)_n$ and $(GAAA)_n$ should form the stem-loop structure more easily than $(GGA)_n$ and $(GAA)_n$. It is also predicted that when a polypurine-polypyrimidine sequence converts from a double stranded structure to a triple stranded stem-loop, the most favorable conditions are those where an even number of basepairs makes the transition. Experimental tests of these predictions are also described.

INTRODUCTION

Stretches of DNA containing long sequences of purines on one strand and pyrimidines on the other (pur·pyr sequences) occur in both prokaryotes and eucaryotes. Their biological function is not known for certain, but they are found at frequencies substantially above that which would be expected on a random basis in several human genes and eucaryotic viruses (1,2), and they are often associated with S1 nuclease hypersensitive sites upstream of eucaryotic genes (3-5). S1 sensitivity is only one of several attributes indicating that pur·pyr sequences have unusual structures; a recent review (6) discusses the various probes used to investigate these structures and lists about thirty naturally occurring pur·pyr sequences that have been examined.

Two broad classes of models for the structure of pur·pyr sequences have been put forward. First are the double stranded

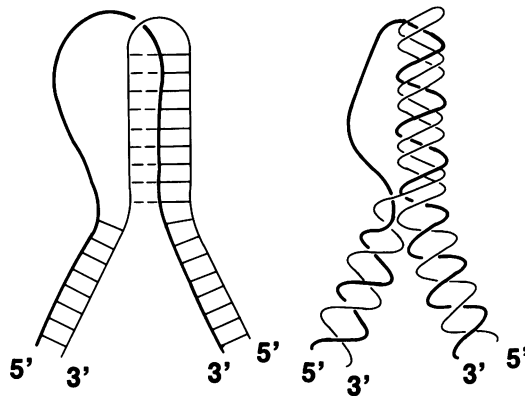


Figure 1: Schematic representations of the model structure of the pur·pyr stem and loop (6,12,16). The heavy lines represent the polypurine strand, while the light lines indicate the polypyrimidine strand. The lefthand side of the figure shows the secondary structure, with solid crossbars for Watson-Crick hydrogen bonds and broken crossbars for Hoogsteen hydrogen bonds. The righthand side of the diagram shows the three dimensional structure. The double helix coming up from the lower right is continued into the triple helix, with the extra polypyrimidine strand lying in the major groove of that double helix.

models, where it has been proposed that the two backbones have different conformations (7) or that Watson-Crick basepairs alternate with Hoogsteen basepairs (8).

Second are three stranded models (9-17). These are based on early studies of nucleic acid polymers that showed the ability of two polypyrimidine strands to associate with one polypurine strand (18,19). X-ray fiber diffraction studies revealed that the three stranded structure consists of a duplex of the A family with the third strand lying in the major groove of the duplex and oriented parallel to the polypurine strand (20). The anti-parallel polypyrimidine strand is associated with the polypurine strand through Watson-Crick hydrogen bonding, while the third strand is held to the polypurine strand by Hoogsteen hydrogen bonds. In Figure 1 the secondary and tertiary structures of the triple stranded model pur·pyr stem and loop are shown schematically.

There is growing support for the three stranded model. Indirect evidence for it is found in the fact that, for sequences

containing deoxycytidine, the unusual structure is generally favored by low pH. This is significant because cytosine must be protonated to form the required Hoogsteen basepair with guanine (10-20). Stronger support comes from the observation that the ability of supercoiling to induce the unusual pur·pyr structure in a sequence with a mirror repeat is substantially reduced if that repeat is disrupted by only one nucleotide out of 32 (13). The most compelling evidence, however, is the pattern in chemical reactivity when the mirror repeat is altered. Changing one base in the stem causes a change in the reactivity of a pyrimidine that is several nucleotides away in the primary sequence, as would be predicted by the triplex model (16).

In principle, two isomers are possible for structures like those in Figure 1, depending on which half of the polypurine strand forms part of the triple helical stem. Chemical modification studies have recently shown unambiguously that the 3' end of the polypurine strand is part of the stem, while the 5' half is single stranded (14-17). Consequently, the 5' half of the polypyrimidine strand forms Watson-Crick basepairs and the 3' half forms Hoogsteen basepairs with the polypurine strand.

While a detailed model structure is available for the three stranded stem of pur·pyr structures (20), no such model is available for the polypyrimidine loop at the end of the stem. Here we describe the development of an automated procedure for building model loops and its application to the investigation of the pur·pyr stem-loop structure. The experimental implications of the model are also discussed.

METHODS

Our method for loop building begins with an idea first proposed by Haasnoot et al. (21), that an ideal stem-loop structure will have optimum base stacking interactions if the regular helical structure of one strand of the stem is propagated into the loop. For a double helical stem, the analysis of distances from a given phosphorus atom in one strand to all phosphorus atoms in the other strand can be used to identify the shortest interphosphorus distance. One or two nucleotides can be used to close that gap, giving a predicted geometry for the loop. This

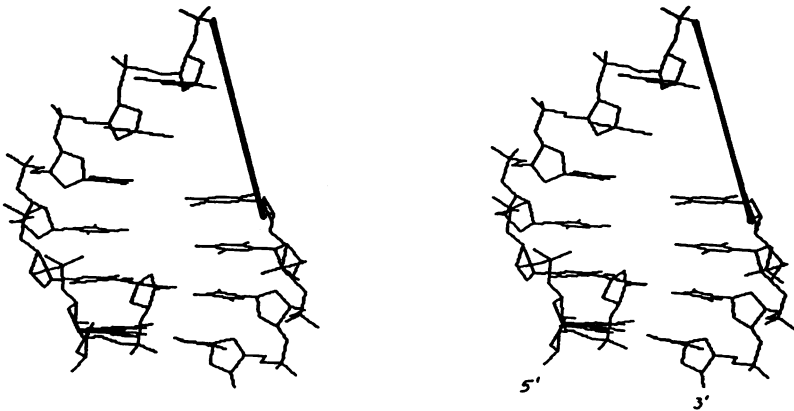


Figure 2: Initial model for pur-pyr stem and 642 loop. The polypyrimidine loop contains six nucleotides, with four of these on the 5' side of the loop (left side) and two on the 3' side (right side). The bottom two residues of each strand are part of the triple helix and are not allowed to move during refinement. The base stacking in the loop is the same as in the stem, but one O5'-C5' bond length is about 10 Å (heavy line) and must be corrected during refinement.

idea can be shown to correctly predict loop sizes and stacking patterns for A-RNA, B-DNA and transfer RNA (21).

This approach is especially useful for manually building loop structures using either physical models or computer graphics. It allows one to reduce the complexity of the problem by choosing an overall loop architecture in advance, for the Haasnoot plot tells whether the loop should be 3'-stacked or 5'-stacked and how large the stack should be.

We designate the loop size by N and use the indices L and M to represent the number of stacked bases on the 5' and 3' sides of the initial loop structure, respectively. Loop geometries are identified by the notation NLM , where $N=L+M$. The Haasnoot procedure identifies a likely value of L (or M) and an estimate of the number of bases needed to close the loop, M (or L).

We have modified the Haasnoot approach by beginning with an initial loop structure in which the ideal helix from the stem is extended into the loop on both the 5' and 3' sides. As an example, consider a six membered loop with four nucleotides on the 5' side and two on the 3' side, which we designate a 642 loop (Figure 2). Although the structure has very favorable base

stacking, it is outrageous in terms of covalent structure, because one covalent bond required for loop closure would have to be over 20 Å long (the O5'-C5' bond shown by the heavy line in Figure 2). A structure like this could be refined by molecular mechanics or molecular dynamics (22), using a package such as AMBER (23), for example. However, extremely large forces would act on the atoms at either end of the very long bond, and those forces would largely determine the path followed through conformational space as the structure is refined.

Instead of conventional molecular mechanics, we have chosen to refine the initial structures with JUMNA (Junction Minimization of Nucleic Acids), a program designed especially for building and manipulating nucleic acid structures (24). Whereas molecular mechanics will attempt to correct the very long O5'-C5' distance in Figure 2 by treating it as a badly distorted bond, JUMNA treats this linkage as an unsatisfied mathematical constraint (25), which does not lead to large energy derivatives. Further, rather than working in Cartesian coordinate space, as does molecular mechanics, JUMNA makes all structural changes in a coordinate space defined by the natural variables of the molecule (26). These include single bond torsions and, in the sugar rings, valence angles. Because of this mathematical sophistication, JUMNA is able to optimize the initial structure of Figure 2 by bringing the O5'-C5' distance back to its optimal value while maintaining reasonable base stacking and backbone conformation throughout the loop. This is shown in Figure 3, the final loop structure created by JUMNA from the starting structure of Figure 2. To maintain the structure of the triple stranded stem, we use an option in JUMNA that allows one to lock the helical parameters of the bases and the relative positions of different strands.

We have used this procedure to build and refine model loop structures for all values of N from three to six. For a given value of N , there are $N+1$ structures to be examined, corresponding to $L=0,1,\dots,N$. We have chosen the sequence poly(dA)·poly(dT) for simplicity; differences in loop conformation arising from sequence differences have not been examined.

Since we are interested in the relative stabilities of different loop sizes and, for a given size, of different loop

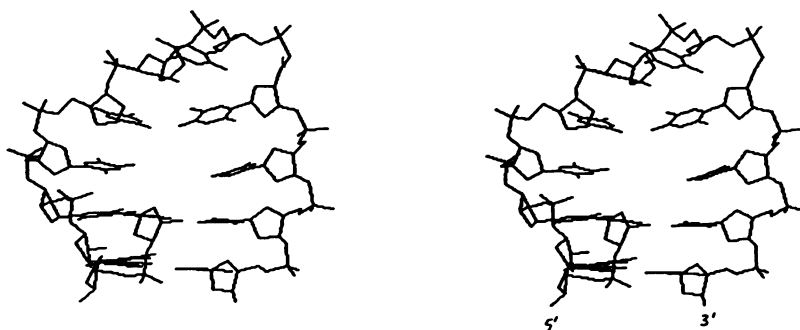


Figure 3: Energy minimized 642 loop structure produced by JUMNA when given the starting structure of Figure 2 and following the procedure described in the text.

conformations, the critical parameter that we need to estimate is the relative energies of the various model loops. The energy of the double stranded structure is not of any consequence in this study, since we are not examining the transition from the double stranded to the triple stranded state; it is the relative probabilities of forming alternative loop structures once the transition has been made that interests us.

Energies are calculated using the standard JUMNA potential function (24-27), which includes electrostatic and van der Waals interactions, an explicit hydrogen bonding term, and terms for rotations about bonds (torsions) and distortions of valence angles. Bond lengths are fixed, but as mentioned above, this is done without the introduction of energy penalties. The dielectric function $D(r)$ is roughly sigmoidal, rising from a $D=1$ at $r=0$ to a plateau value of $D=78$ for interchange distances greater than about 20 Å.

Rough estimates of solvation energy are obtained on the final optimized structures by calculating the total solvent-accessible surface area using the algorithm of Lee and Richards (28) with a probe radius of 1.4 Å and multiplying that area by 25 cal/Å².

RESULTS AND DISCUSSION

A Haasnoot plot (21) for interphosphate distances between the polypyrimidine strands of a triple helical pur-pyr structure

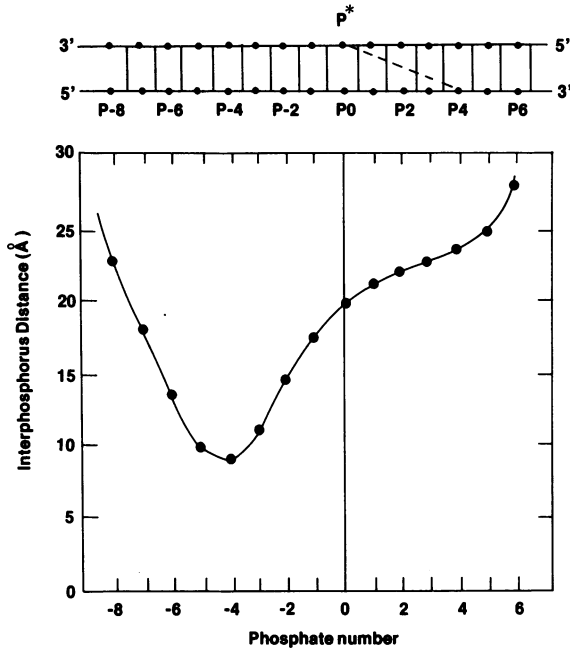


Figure 4: Haasnoot plot (20) of distances between phosphorus atoms of pyrimidine strands in a regular pur-pyr triple helix. Distances are measured from a given phosphorus atom on one strand (P^*) and the phosphorus atoms on the other strand. The numbering in the second strand proceeds in the 5' to 3' direction, as indicated in the schematic representation at the top of the figure. As an example, the distance between P^* and P_4 (dashed line) is 23.1 Å.

is shown in Figure 4; the atomic coordinates for the triple helix are those provided by x-ray fiber diffraction (20). Figure 4 would predict that the optimum size for the polypyrimidine loop would be about five nucleotides. With four bases on the 3' side of the loop stacked in an extension of the stem structure, the interphosphate gap of 8.9 Å might be closed by one nucleotide, especially if the backbone of the helical extension were allowed to move slightly to shorten the gap; this would give a 514 loop. Alternative structures would have three bases stacked on the 3' side, with one or two bases to close the 11.0 Å gap, giving a 413 or a 523 loop.

The results of the automated loop building procedure are given in Tables 1-4, which present the calculated energies for

TABLE 1
Energetics of Three-Membered Loops
(all energies in kcal/mol)

Structure	E _C	E _{SA}	E _C + E _{SA}
303	-120.4	48.0	-72.4
312	-108.4	49.0	-59.4
321	-123.0*	46.7	-76.3*
330	-120.9	46.4	-74.5

E_C = Conformational energy

E_{SA} = Solvation energy, based on solvent-accessible surface areas multiplied by 25 cal/Å².

* Denotes lowest energy structure.

loops containing from three to six nucleotides, and in Figures 5-8, which show the lowest energy structures for each size of loop. With the exception of the three membered loop, the lowest energy structures are all found to be 3' stacked (404, 514, and 615), in agreement with the prediction of Figure 4. The lowest energy three membered loop (321) is essentially 5' stacked, although the base at the top of the loop partly bridges the 5' and 3' stacks (Figure 5).

TABLE 2
Energetics of Four-Membered Loops
(all energies in kcal/mol)

Structure	E _C	E _{SA}	E _C + E _{SA}
404	-148.5*	49.9	-98.6*
413	-141.0	51.5	-89.5
422	-135.1	52.0	-83.1
431	-129.5	53.5	-76.0
440	-132.5	51.3	-81.2

See Table I for nomenclature

* Denotes lowest energy structure.

TABLE 3
Energetics of Five-Membered Loops
(all energies in kcal/mol)

Structure	E_C	E_{SA}	$E_C + E_{SA}$
505	-156.4	58.7	-97.7
514	-157.7*	54.7	-103.0*
523	-142.0	53.4	-88.6
532	-148.7	55.7	-93.0
541	-146.2	56.6	-89.6
550	-138.0	60.0	-78.0

See Table I for nomenclature

* Denotes lowest energy structure.

Although the range of solvation energies can be substantial for a given loop size (it is over 6 kcal/mol for five and six membered loops), the corrections for solvation energy do not affect the rank ordering of any of the structures within a given loop size. This is because loops with good base stacking and relatively unstrained backbone conformations are generally more compact and consequently have low solvent accessible surface areas.

TABLE 4
Energetics of Six-Membered Loops
(all energies in kcal/mol)

Structure	E_C	E_{SA}	$E_C + E_{SA}$
606	-156.2	65.8	-90.4
615	-170.5*	62.0	-108.5*
624	-161.2	62.3	-98.9
633	-145.1	64.6	-80.5
642	-156.7	63.4	-93.3
651	-162.1	61.4	-100.7
660	-134.7	67.6	-67.1

See Table I for nomenclature.

* Denotes lowest energy structure.

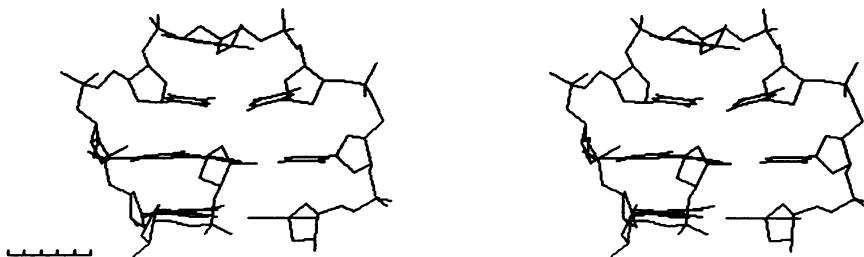


Figure 5: The 321 loop structure, the lowest energy three membered loop conformation. In this and the following figures, the 5' end of the molecule is at the lower left.

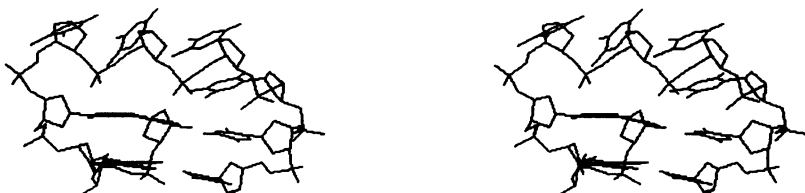


Figure 6: The 404 loop structure, the lowest energy four membered loop conformation.

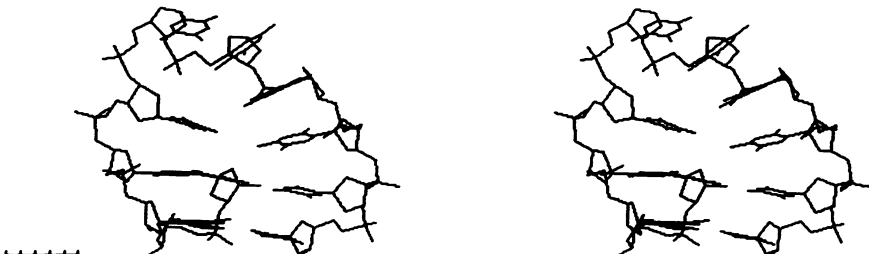


Figure 7: The 514 loop structure, the lowest energy five membered loop conformation.

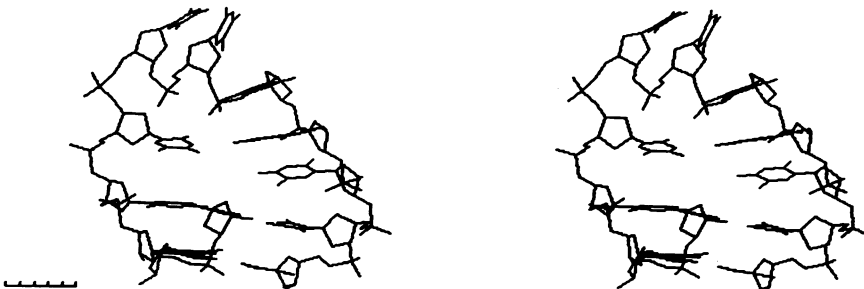


Figure 8: The 615 loop structure, the lowest energy six membered loop conformation.

TABLE 5
Energetics of Loops of Various Sizes

Loop Size N	Lowest Energy Structure	E _{SS}	E _C +E _{SA} -E _{SS}
3	321	-22.9	-53.4
4	404	-34.5	-64.1*
5	514	-46.1	-56.9
6	615	-57.7	-50.8

E_{SS} = Correction term for difference between single strand in loop vs. single strand incorporated into triplex.
See Table 1 for further nomenclature.

Again, it should be emphasized that it is the relative energies that are of interest here, because we want to know the relative probabilities of the various loop conformations once the transition from the double stranded structure to the triple stranded stem-loop has taken place. We have not attempted to compare these energies with that of the double stranded structure, because we are not examining the transition itself, only the final state.

In comparing the energies of loops of different sizes to determine the global minimum energy structure, it is necessary to correct for loop size. The lower energies associated with larger loops in Tables 1-4 arise from increased base stacking, so that larger loops always appear to be favored. But the critical issue is the energy difference between that of a given N-membered loop and that of the same number of nucleotides stacked in the triplex. The latter is calculated by subtracting the single strand conformational energy (calculated by JUMNA) and the solvation energy difference, along with the energy of the Hoogsteen hydrogen bonds. When this correction is made (Table 5), we find that the lowest energy structure is the 404 loop, followed by the 514 and 321 loops.

It is seen that the Haasnoot plot does correctly identify the basic loop size and stacking pattern (3'-stacked or 5'-stacked). But it does not provide a method of examining the steric problems that must be resolved in building loops, so

TABLE 6
 Structural Parameters of Polypyrimidine Strand of the 404 Loop
 (All angles in degrees.)

Residue	α	β	γ	δ	ϵ	ζ	χ	Pucker
T1				150	-152	-70	39	C3'-exo
T2	-94	170	49	91	-130	-50	53	04'-endo
T3	148	179	49	95	-158	-82	10	C2'-exo
T4	-82	183	58	90	-171	-82	52	C3'-endo
T5	-77	185	57	84	-151	-61	26	C3'-endo
T6	-80	178	62	92	-154	-67	33	C3'-endo
T7	-71	176	70	147	-170	-101	70	C2'-endo
T8	-74	181	53	126			67	C1'-exo
A-DNA	-55	172	45	83	-160	-69	28	C3'-endo

detailed molecular models are necessary. We believe that our procedure provides a relatively unbiased method of building, refining, and comparing models of alternative structures.

The structural details of the 404 loop are given in Table 6. The changes in backbone conformation from the standard A-DNA conformation are surprisingly small. With only four exceptions, none of the backbone torsion angles has changed dramatically, and the changes are at the boundaries between the stem and loop, rather than in the loop itself. Three of the large changes are in the endocyclic torsion angle δ (C5'-C4'-C3'-O3'), which has gone from the gauche+ configuration to trans in thymidines 1 and 7 and from gauche+ to an eclipsed configuration in thymidine 8. The fourth change occurs in α (O3'-P-O5'-C5') for T3, which has changed from gauche- to trans.

Four sugars have repuckered. The changes in δ mentioned above represent repuckering of three deoxyribosees from the northern (N) quadrant (C3'-endo and C2'-exo) to the southern (S) quadrant (C1'-exo, C2'-endo, and C3'-exo). In addition, deoxyribose 2 has adopted O4'-endo puckering. While this conformation is of higher energy (29,30) than either the N or S conformation, it is observed in crystal structures of both riboses and deoxyribosees (30-32) and occurs spontaneously in molecular dynamics simulations on both RNA and DNA (29,33).

Finally, let us consider two predictions arising from this study. First, because a four-membered loop is indicated as being much more stable than a three-membered loop (Table 5), our model would predict that it will be more difficult to cause the formation of the unusual pur·pyr structure if the sequence is a repeat of the form (XYZ)_n. Thus, the sequences (GA)_n, (GGGA)_n and (GAAA)_n should form the triple stranded stem-loop structure more easily than (GGA)_n and (GAA)_n. These differences should manifest themselves in the levels of supercoiling pressure that are necessary to induce the transition in different sequences. Both *in vivo* and *in vitro* studies might shed light on this issue.

Second, we note that the stem of the triplex always contains an even number of pyrimidines. Since a loop with an even number or pyrimidines (four) is favored, our model would also predict that the length of the polypurine-polypyrimidine sequence that is

actually converted into the unusual pur·pyr structure should contain an even number of residues. This feature should be detectable in patterns of chemical reactivity and enzyme susceptibility, using standard probes for pur·pyr sequences (6).

ACKNOWLEDGEMENTS

We are indebted to Dr. Krystyna Zakrzewska for numerous discussions and to Dr. M. Prabhakaran for assistance in preparing the figures showing molecular structures. R.L. would like to thank the Association for International Cancer Research (Brunel and St. Andrews Universities, U.K.) for their support of this research. This research was also supported by a grant to S.C.H from the National Institutes of Health (GM-34015).

*To whom correspondence should be addressed

REFERENCES

1. Behe, M.J. (1987) *Biochemistry* 26, 7870-7875.
2. Beasty, A.M. and Behe, M.J. (1988) *Nucleic Acids Res.* 16, 1517-1528.
3. Hentschel, C.C. (1982) *Nature* 295, 714-716.
4. Mace, H.A.F., Pelham, H.R.B. and Travers, A.A. (1983) *Nature* 304, 555-557.
5. Larsen, A. and Weintraub, H. (1982) *Cell* 29, 609-622.
6. Wells, R.D., Collier, D.A., Hanvey, J.C., Shimizu, H. and Wohlrab, F. (1988) *FASEB J.*, in press.
7. Evans, T. and Efstratiadis, A. (1986) *J. Biol. Chem.* 261, 14771-14780.
8. Pulleyblank, D.A., Haniford, D.B. and Morgan, A.R. (1985) *Cell* 42, 271-280.
9. Lee, J.S., Johnson, D.A. and Morgan, A.R. (1979) *Nucleic Acids Res.* 6, 3073-3091.
10. Lee, J.S., Woodsworth, M.L., Latimer, L.J.P. and Morgan, A.R. (1984) *Nucleic Acids Res.* 12, 6603-6614.
11. Christophe, D., Cabrer, B., Bacolla, A., Targovnik, H., Pohl, V. and Vassart, G. (1985) *Nucleic Acids Res.* 13, 5127-5144.
12. Lyamichev, V.I., Mirkin, S.M. and Frank-Kamenetskii, M.D. (1986) *J. Biomolec. Struct. Dyns.* 3, 667-669.
13. Mirkin, S.M., Lyamichev, V.I., Drushlyak, K.N., Dobrynin, V.N., Filippov, S.A. and Frank-Kamenetskii, M.D. (1987) *Nature* 330, 495-497.
14. Hanvey, J.C., Klysiak, J. and Wells, R.D. (1988) *J. Biol. Chem.*, 263, 7386-7396.
15. Collier, D.A., Griffin, J.A. and Wells, R.D. (1988) *J. Biol. Chem.*, 263, 7397-7405.
16. Hanvey, J.C., Shimizu, M. and Wells, R.D. (1988) *Proc. Natl. Acad. Sci. USA*, 85, 6292-6296.
17. Voloshin, O.N., Mirkin, S.M., Lyamichev, V.I., Belotserkovskii, B.P. and Frank-Kamenetskii, M.D. (1988) *Nature* 333, 475-476.

-
18. Riley, M., Maling, B. and Chamberlin, M.J. (1966) *J. Mol. Biol.* 20, 359-389.
 19. Morgan, A.R. and Wells, R.D. (1968) *J. Mol. Biol.* 37, 63-80.
 20. Arnott, S. and Selsing, E. (1974) *J. Mol. Biol.* 88, 509-521.
 21. Haasnoot, C.A.G., Hilbers, C.W., van der Marel, G.A., van Boom, J.H., Singh, U.C., Pattabiraman, N. and Kollman, P.A. (1986) *J. Biomolec. Struct. Dyns.* 3, 843-857.
 22. McCammon, J.A. and Harvey, S.C. (1987) *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, London.
 23. Weiner, P.K. and Kollman, P.A. (1981) *J. Comput. Chem.* 2, 287-303.
 24. Lavery, R. (1987) in *Unusual DNA Structures*, R.D. Wells and S.C. Harvey, Eds., pp 189-206, Springer-Verlag, New York.
 25. Lavery, R., Parker, I. and Kendrick, J. (1986) *J. Biomolec. Struct. Dyns.* 4, 443-462.
 26. Lavery, R., Sklenar, H., Zakrzewska, K. and Pullman, B. (1986) *J. Biomolec. Struct. Dyns.* 3, 989-1014.
 27. Lavery, R., Zakrzewska, K. and Pullman, A. (1984) *J. Comp. Chem.* 5, 363-373.
 28. Lee, B. and Richards, F.M. (1971) *J. Mol. Biol.* 55, 379-400.
 29. Harvey, S.C. and Prabhakaran, M. (1986) *J. Amer. Chem. Soc.* 108, 6128-6136.
 30. Olson, W.K. (1982) *J. Amer. Chem. Soc.* 104, 278-286.
 31. Hingerty, B., Brown, R.S. and Jack, A. (1978) *J. Mol. Biol.* 124, 523-534.
 32. Fratini, A.V., Kopka, M.L., Drew, H.R. and Dickerson, R.E. (1982) *J. Biol. Chem.* 257, 14686-14707.
 33. Singh, U.C., Weiner, S.J. and Kollman, P. (1985) *Proc. Natl. Acad. Sci. USA* 82, 755-759.