

Intramolecular DNA triplexes in supercoiled plasmids

(homopurine-homopyrimidine/OsO₄/diethyl pyrocarbonate/Hoogsteen base pairs)

JEFFERY C. HANVEY, MITSUHIRO SHIMIZU, AND ROBERT D. WELLS

Department of Biochemistry, Schools of Medicine and Dentistry, The University of Alabama at Birmingham, Birmingham, AL 35294

Communicated by Charles C. Richardson, May 16, 1988

ABSTRACT A series of inserts with oligopurine-oligopyrimidine mirror repeat sequences was investigated at the base pair level with specific chemical probes (OsO₄ and diethylpyrocarbonate) to evaluate the *in vitro* existence of intramolecular triplexes. Two parent inserts in recombinant plasmids with (GAA)_n and (AG)₁₂ sequences and three mutant inserts (containing transitions or transversions) revealed that base pair changes at one location affected the chemical reactivity 13 base pairs away. The specificity and nature of these reactions, as well as the thermal stability of the complexes, provide direct evidence for the existence of a triplex with a portion of the pyrimidine-rich strand folded back and Hoogsteen-paired in the major groove of the Watson-Crick duplex. The biological implications of this unorthodox DNA structure are discussed.

Recent studies revealed that oligopurine-oligopyrimidine segments [oligo(R·Y)] in recombinant plasmids are not in the usual right-handed B conformations (for review, see ref. 1). The unorthodox but right-handed structures are stabilized by negative supercoiling and low pH (≈5.0) in sequences with 25–100% GC but not in similar lengths of A·T (2). The structures are different from any previously described non-B conformations (1). The oligo(R·Y) sequences are abundant in eukaryotic genomes (3) and are frequently located near regulatory regions and in recombination hotspots (refs. 4–16; for review, see ref. 17).

Several right-handed models were proposed to explain the unusual structure of oligo(R·Y) sequences: the two strands having different backbone conformations (18), a duplex in which Watson-Crick A·T pairs alternate with Hoogsteen (syn) G·C pairs (19), slipped structures (8–10), and formation of intramolecular triplex structures (also called H-DNA) (13, 20, 21). Recent two-dimensional gel electrophoretic analyses (21) as well as fine mapping studies on families of simple model repeating inserts (2, 22) gave results that were most consistent with intramolecular triplexes (Fig. 1). However, direct proof of the triple base interaction was lacking.

Three-stranded complexes (triplexes) were widely investigated in the 1960s (for reviews, see refs. 23–25) with DNA and/or RNA polymers. In general, triplexes consisted of one polypurine strand and two polypyrimidine strands with the additional Y_n strand Hoogsteen-paired in the major groove with the R_n strand (Fig. 1). Hence, the T·A·T (or U) and C·G·C arrangements provided substantial specificity for the interactions. Formation of two hydrogen bonds between (G)_n and cytosine residues in the third strand requires hemi-protonation of cytosines. However, the pH of deoxycytosine residues in polynucleotides is above pH 7 (26, 27) and hence is consistent with physiological conditions.

Herein, we prepared a family of inserts designed specifically to test the existence of the highly unorthodox triplex model in plasmids. Certain base pair changes were instituted to elicit instability in the putative triplex at specific loci some

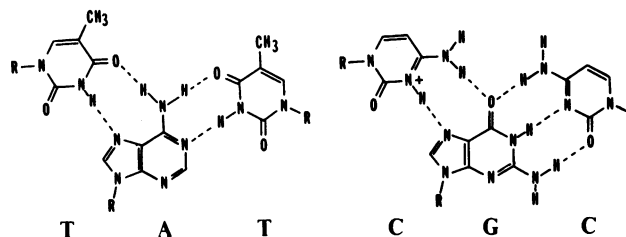
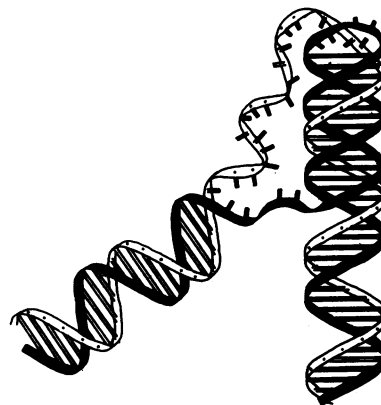


FIG. 1. Model for intramolecular triplex (*Upper*). Watson-Crick paired duplexes of the vector (*Lower*). The strand containing the pyrimidine-rich sequence is shaded, whereas the strand containing the purine-rich sequences is unshaded. The 3'-half of the purine strand (termed second strand) of the insert remains Watson-Crick paired with the 5'-half of the pyrimidine strand. The 3'-half of the pyrimidine strand (termed third strand), after dissociating from its Watson-Crick complement, occupies the major groove of the duplex region described above and is specifically complexed with it through Hoogsteen base pairs (*Lower*). The mirror repeat and the oligopurine-oligopyrimidine nature of these sequences enable this specific association. The structure of the nonpaired 5'-half of the purine-rich strand is uncertain.

distance away, due to its folded-back nature. Fine mapping at the base pair level with chemical probes specific for certain nucleotides validated the presence of triplexes.

MATERIALS AND METHODS

Plasmids. pRW791 was made by filling in the *EcoRI* site of pRW790 (15, 28). Oligonucleotides were synthesized and cloned into the *BamHI* site of pRW790 (or pRW791 for the pRW1704 insert) as described (2, 15, 16, 22, 28, 29). The inserts were synthesized with GATCC on the 5' end and a guanine on the 3' end, except for the pRW1704 insert, which contained a TTCG on the 3' end of the purine strand. The insert in pRW1402 was found when screening transformants for the insert of pRW1404.

Chemical Modifications. Chemical modifications of the plasmids were as described (2). For the reactions at elevated temperatures, the mixtures were preincubated for 15 min at 25°C and then 15 min at the indicated temperatures. The reactions were stopped by placing the containers on ice, and the subsequent workup was as described (2).

When pRW1704 was treated with diethyl pyrocarbonate, the adenines within the 5' half of the (GAA)₉ stretch (positions 2–15), but not in the 3' half, were hyperreactive (Fig. 4A). These results are similar to those reported previously for a (GAA)₈-(TTC)₈ insert (pRW1406) (2).

A GAA Insert with Three Interruptions in the Strand Bias. pRW1412 contains three interruptions in the R-Y stretch, at positions 8, 14, and 21 (Fig. 2A). At each location, an A-T pair was changed to a T-A pair. The interruptions were placed such that, if the insert forms an intramolecular triplex, there would be only one unpaired A-T between strands two and three (Fig. 5).

When pRW1412 was treated with OsO₄, thymines at the center of the homopyrimidine stretch (positions 12, 15, and 17) and at the BamHI site (position -4) were modified (Fig. 3A). The pattern of reactivity was similar to that seen for the control insert (GAA)₉-(TTC)₉ (Fig. 3A) and other R-Y sequences (2). This indicates that the three pyrimidines in the purine stretch of pRW1412, including two not in the center, did not change the region over which the structure formed.

The homopurine strand of pRW1412 contains three thymines. OsO₄ reactivity was seen for the thymines at positions 8 and 14 and the thymine in the BamHI site (location -3) (Fig. 3B). However, the thymine within the 3' half of the purine strand (number 21) was not reactive to OsO₄ (Fig. 3B). If these interruptions do not significantly alter the structure, then the reactivity of the three thymines in the purine stretch to OsO₄ was exactly as expected for the triplex model (Fig. 5).

The 5'-half of the purine strand of pRW1412 was hyperreactive to diethyl pyrocarbonate (Fig. 4A). The length of the hyperreactivity was similar to that for the insert without the interruptions (pRW1704) (positions 2–15) and that seen for other R-Y sequences (Fig. 4A and ref. 2). Again, this is a good indication that the three thymines in the purine strand do not greatly interfere with the unusual structure.

The adenines in the center and in the 3'-half of the pyrimidine strand of pRW1412 (positions 8 and 14), but not

the adenine within the 5'-half (location 21), were hyperreactive to diethyl pyrocarbonate (Fig. 4B). The hyperreactivity of the central adenine was expected, based on the known reactivity of the center of both strands of R-Y sequences (Fig. 3 and ref. 2). Also, the lack of hyperreactivity of the adenine at position 21 was expected because, overall, no reactivity has been seen previously for thymines at this position (ref. 2 and Fig. 3A).

However, the hyperreactivity of the adenine at position 8 to diethyl pyrocarbonate revealed its accessibility, consistent with its inability to Hoogsteen pair with the thymine at position 21 in a triplex (Fig. 5). Because the thymines flanking this adenine were not reactive to OsO₄ (Fig. 3A), the distorted region in pRW1412 may be as little as one nucleotide.

A GAA Insert with One Interruption in the Strand Bias. pRW1415 contains a (GAA)₉-(TTC)₉ insert, except that one A-T pair was replaced by a T-A pair. The A-T to T-A transversion is at position 21 (Fig. 2A). This insert was constructed to specifically test the intramolecular triplex model. The interruption is positioned such that a thymine located 13 base pairs (bp) away (at position 8), which is normally not modified by OsO₄, would be unpaired and thus reactive to OsO₄ if an intramolecular triplex forms (Fig. 5).

Fig. 3A shows the pattern of reactivity on the pyrimidine strand, and Fig. 3B shows that on the purine strand, when pRW1415 was probed with OsO₄. The central thymines (positions 11–17) were reactive to OsO₄, as seen for the perfect repeat in pRW1704. In addition, two thymines at positions 8 and 9 were also reactive (Fig. 3A); the thymine at position 8 was twice as reactive. These two thymines are 12 and 13 bp away from the interruption.

If an intramolecular triplex as shown in Fig. 5 is formed for pRW1415, the thymine at position 8 would lie directly across from the thymine in the purine strand and could not Hoogsteen pair.

Also, as seen for all other R-Y sequences (2), the thymines within the BamHI site proximal to position 1 of pRW1415

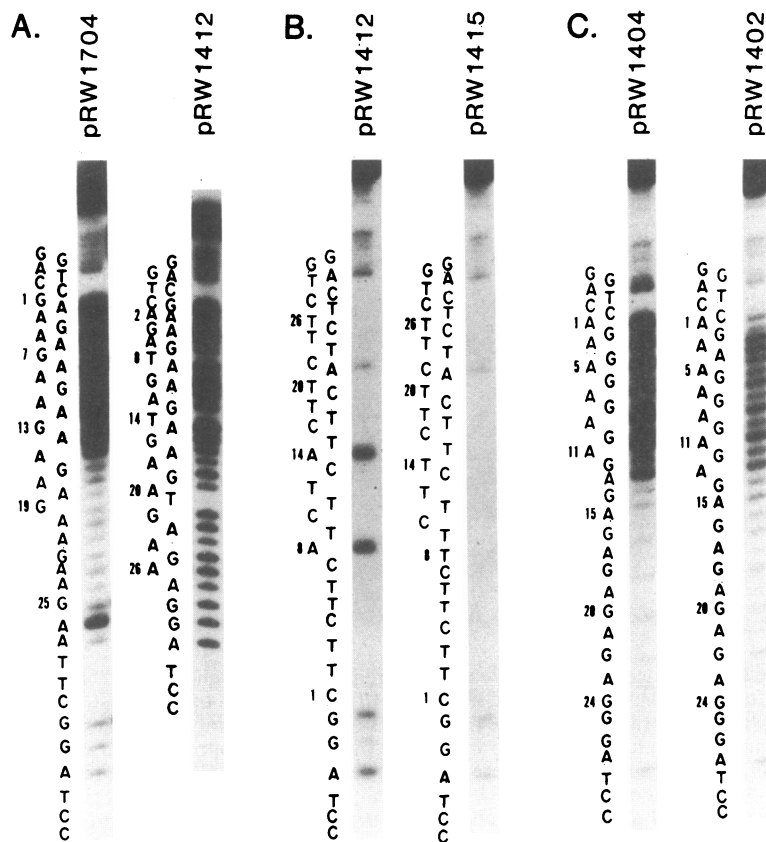


FIG. 4. Mapping of diethyl pyrocarbonate sites at the base pair level. (A) Diethyl pyrocarbonate sites in the purine strand of the insert for pRW1704 and pRW1412. (B) Diethyl pyrocarbonate sites in the pyrimidine strand of the inserts for pRW1412 and pRW1415. (C) Diethyl pyrocarbonate sites in the purine strand of pRW1404 and pRW1402.

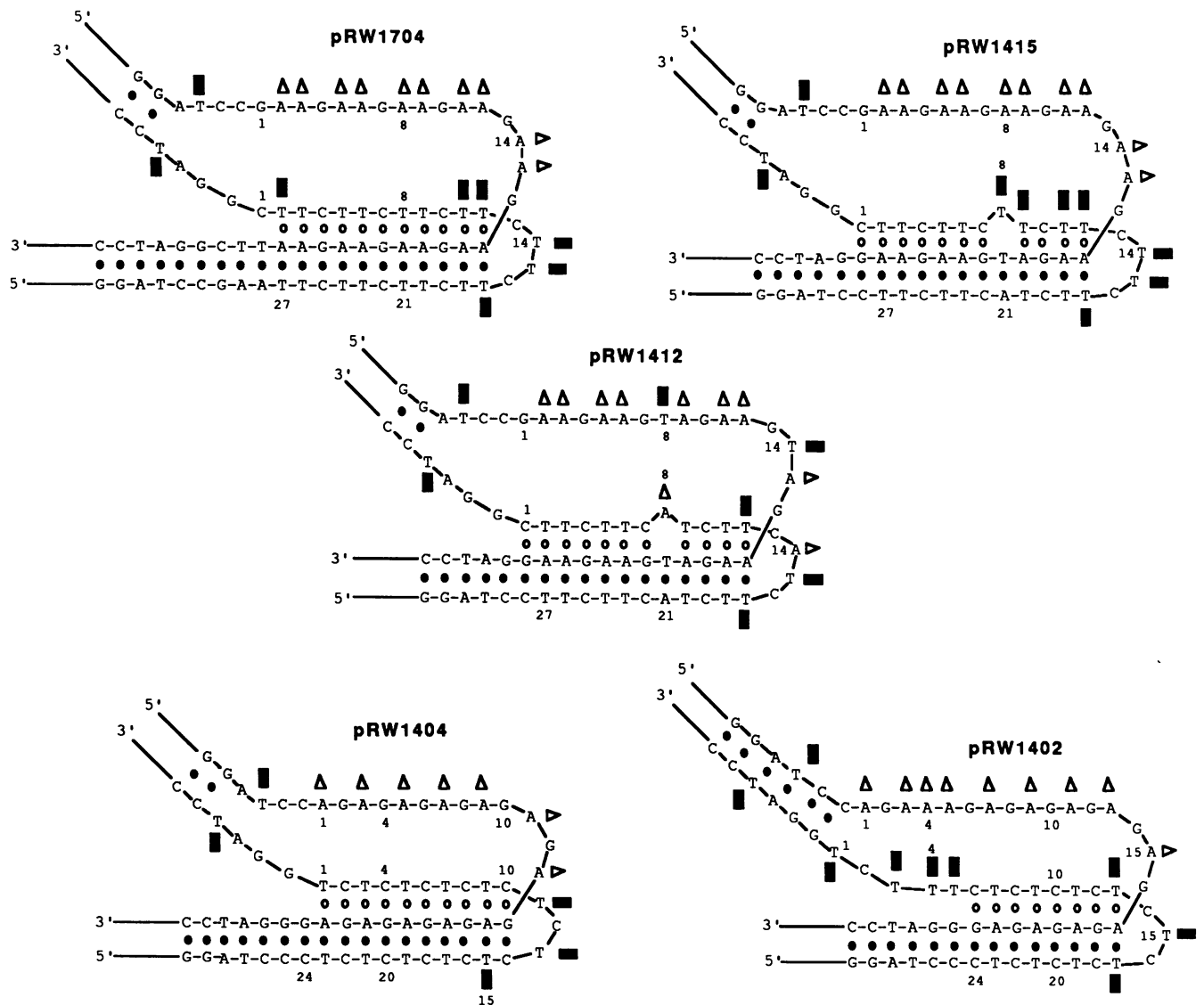


FIG. 5. Possible intramolecular triplexes formed by the inserts in the plasmids. The triplexes were drawn in accord with the experimental findings. Therefore, the interruptions in the R·Y bias in pRW1412 and pRW1415 do not change the length of the triplex. However, the G → A change in pRW1402 has shortened the triplex, compared to pRW1404. ■, OsO₄-reactive sites; △, diethyl pyrocarbonate-reactive sites. The relative extents of the reactions are not shown. ●, Watson-Crick bp; ○, Hoogsteen bp.

were reactive to OsO₄ (Fig. 3). As predicted by the intramolecular triplex model, the thymine (which is the actual interruption within the homopurine strand of pRW1415) (at position 21), was not reactive to OsO₄. This result was expected since the thymine was Watson-Crick-paired to the adenine at 21 in the stem of the triplex.

The adenine residues within the 5'-half of the purine strand of pRW1415 (positions 2-15) were hyperreactive to DEPC (data not shown), similar to the results with all other R·Y sequences. Also, the adenine within the homopyrimidine stretch of pRW1415 (position 21) was not hyperreactive to diethyl pyrocarbonate (Fig. 4B), as expected and similar to the adenine in the same location of pRW1412.

In summary, the composite results with pRW1412 and pRW1415 compared to the control pRW1704 strongly support the triplex model because a change of an A·T to a T·A pair at position 21 alters the reactivity at the precise expected location (position 8), which is 13 bp away. These data are not consistent with the other models.

A Perfect AG Repeat. pRW1404 contains an (AG)₁₂ insert in the BamHI site of pRW790. Previously, OsO₄ reactivity was seen at the center and 3' end of the pyrimidine strand of

pRW1404 and at the thymines of the BamHI site proximal to position 1 (ref. 2 and Fig. 3C). Also, adenine residues within the 5' half of the purine stretch of (AG)₁₂ (positions 1-13) were hyperreactive to diethyl pyrocarbonate (Fig. 4C and ref. 2). These data supported the triplex model (2).

An AG Repeat Containing One G → A Mutation. pRW1402 is identical to pRW1404 except that it has an adenine instead of a guanine at position 4 (Fig. 2B). The position of the mutation is such that, if the sequence forms an intramolecular triplex identical to that for pRW1404, a G·T Hoogsteen mismatch would occur three nucleotides from the end of the triplex.

The G·C to A·T transition in pRW1402 does not prevent the formation of an unusual DNA structure by the R·Y insert, as demonstrated by the OsO₄ reactivity (Fig. 3C). However, the location of the thymine residues within the TC repeat that were modified by OsO₄ was changed significantly. Reactivity within the pyrimidine stretch was shifted toward position 24 [away from the center and away from the mutation (upward on the gel in Fig. 3C)] by one CT repeat. In addition, the thymines in positions 3-5 were strongly reactive to OsO₄ (Fig. 3C), whereas only a weak reaction was seen previously

for pRW1404 (Fig. 3C and ref. 2). Furthermore, the strong reactivity at the *Bam*HI site proximal to location 1 of pRW1404 was reduced in pRW1402 (Fig. 3C).

If an identical triplex is formed by pRW1404 and pRW1402, stronger reactivity of the thymine at position 4 in pRW1402 would be predicted because of a disallowed T-G Hoogsteen pair. However, the reduced reactivity of the *Bam*HI site near position 1 and the shifting of reactivity within the central region of the sequence would not be expected. On the other hand, if the R-Y sequence of pRW1402 forms a shorter triplex than pRW1404, to "avoid" the T-G Hoogsteen mismatch (Fig. 5), the predicted reactivity would be precisely as observed. The loop of the triplex would shift toward the 5' end of the pyrimidine strand, and the TT repeat (positions 4 and 5) would become the end of the triplex.

The mutation at location 4 in pRW1402 did not significantly affect the diethyl pyrocarbonate reactivity of the insert (Fig. 4C). As predicted by the triplex model (Fig. 5), the diethyl pyrocarbonate hyperreactive sites in the insert were at positions 1–15 for pRW1402. Much less modification by diethyl pyrocarbonate was found at positions 17–24.

Stability of Intramolecular Triplexes. The triplexes described herein are stabilized by low pH (≈ 5.0) and by negative supercoiling. The triplexes of pRW1412 and pRW1415 contain one nonpair between the second and third strands in the stem. It could be expected, therefore, that these triplexes would be less stable than the triplex formed by the perfect R-Y repeat (pRW1704). To test the thermal stability of these triplexes, the plasmids were probed with OsO₄ at various temperatures, and the sites of modification were determined. The pattern of OsO₄ reactivity of the control (TTC)₉ insert (pRW1704) was not changed from 25°C to 55°C, although there was reduced reactivity at 55°C. By 60°C, all thymines within the insert were modified to the same extent, indicating that the triplex structure was lost (data not shown). The inserts in pRW1412 and pRW1415 maintained their patterns of OsO₄ reactivity only to 50°C, and by 55°C the triplex structures were lost. Therefore, the perfect R-Y repeat appears to be more stable by about 5°C than those containing interruptions.

When a plasmid containing a 3-bp interruption in the center of the insert [pRW1702, (GAA)₄TTC(GAA)₄] was examined (data not shown), the thermal stability of its triplex was similar to that of the perfect repeat, (GAA)₉ (pRW1704). This is further support for the triplex model because a single interruption located in the putative stem of the triplex causes more instability than three interruptions located in its loop.

DISCUSSION

These results with R-Y inserts containing interruptions in the strand bias or in the mirror repeat provide direct evidence for triple base interactions in plasmids. Considered along with our previous studies on perfect R-Y inserts (2, 22) and the results of Frank-Kamenetskii and coworkers (20, 21), these data demonstrate the formation of intramolecular triplexes.

Several features of this conformation might be involved in protein recognition, including the three-stranded segment in which the Watson-Crick segment is A-form (38), as well as the nonpaired pyrimidine-rich loop, the junctions, and the homopurine segment. These unpaired regions would provide ample sites for specific recognition. Also, these unpaired regions are possible sites for strand invasion in recombination. Alternatively, if the duplex R-Y tract were recognized by a specific ligand, formation of a triplex would likely prevent such an interaction. Interestingly, *E. coli* RNA polymerase-mediated transcription of DNA polymers was substantially inhibited by formation of specific triplexes with RNA polymers (24).

Although two isomers of a triplex are possible (20, 21, 35), our chemical and enzymatic mapping studies (refs. 2 and 22 and references therein) show that only the isomer with the

5'-end of the purine strand unpaired (Figs. 1 and 5) does form. It is conceivable that the two isomers, if formed under appropriate conditions, would have different biological properties (i.e., relative to inhibition of transcription).

The recent development of methods for determining the *in vivo* existence of non-B DNA (39) enables an evaluation of the properties and potential functions of triplexes in living cells.

Note Added in Proof. Recent studies (40, 41) agree with the conclusions reported herein.

This work was supported by grants from the National Institutes of Health (GM 30822) and the National Science Foundation (86-07785). J.C.H. was supported, in part, by a National Institutes of Health Postdoctoral Fellowship (T32 AI 07150) and M.S. by a grant from the Yoshida Foundation for Science and Technology.

- Wells, R. D. & Harvey, S. C. (1987) *Unusual DNA Structures* (Springer, New York).
- Hanvey, J. C., Klysik, J. & Wells, R. D. (1988) *J. Biol. Chem.* **263**, 7386–7396.
- Behe, M. J. (1987) *Biochemistry* **26**, 7870–7875.
- Hentschel, C. C. (1982) *Nature (London)* **295**, 714–716.
- Htun, H., Lund, E. & Dahlberg, J. E. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7288–7292.
- Evans, T., Schon, E., Gora-Maslak, G., Patterson, J. & Efstratiadis, A. (1984) *Nucleic Acids Res.* **12**, 8043–8058.
- Dybvig, K., Clark, C. D., Aliperti, G. & Schlesinger, M. J. (1983) *Nucleic Acids Res.* **11**, 8495–8508.
- Mace, H. A. F., Pelham, H. R. B. & Travers, A. A. (1983) *Nature (London)* **304**, 555–557.
- McKeon, C., Schmidt, A. & deCrombrugge, B. (1984) *J. Biol. Chem.* **259**, 6636–6640.
- Shen, C.-K. J. (1983) *Nucleic Acids Res.* **11**, 7899–7910.
- Nickol, J. M. & Felsenfeld, G. (1983) *Cell* **35**, 467–477.
- Schon, E., Evans, T., Welsh, J. & Efstratiadis, A. (1983) *Cell* **35**, 838–848.
- Christophe, D., Cabrer, B., Bacolla, A., Targovnik, H., Pohl, V. & Vassart, G. (1985) *Nucleic Acids Res.* **13**, 5127–5144.
- Fowler, R. F. & Skinner, D. M. (1986) *J. Biol. Chem.* **261**, 8994–9001.
- Wohlrab, F., McLean, M. J. & Wells, R. D. (1987) *J. Biol. Chem.* **262**, 6407–6416.
- Kilpatrick, M. W., Torri, A., Kang, D. S., Engler, J. A. & Wells, R. D. (1986) *J. Biol. Chem.* **261**, 11350–11354.
- Wells, R. D. (1986) in *DNA Ligand Interactions: From Drugs to Proteins*, ed. Guschlbauer, W. (Plenum, New York), pp. 63–84.
- Evans, T. & Efstratiadis, A. (1986) *J. Biol. Chem.* **261**, 14771–14780.
- Pulleyblank, D. E., Haniford, D. B. & Morgan, A. R. (1985) *Cell* **42**, 271–280.
- Lyamichev, V. I., Mirkin, S. M. & Frank-Kamenetskii, M. D. (1986) *J. Biomol. Struct. Dyn.* **3**, 667–669.
- Mirkin, S. M., Lyamichev, V. I., Drushlyak, K. N., Dobrynin, V. N., Filippov, S. A. & Frank-Kamenetskii, M. D. (1987) *Nature (London)* **330**, 495–497.
- Collier, D. A., Griffin, J. D. & Wells, R. D. (1988) *J. Biol. Chem.* **263**, 7397–7405.
- Felsenfeld, G. & Miles, H. T. (1967) *Ann. Rev. Biochem.* **36**, 407–448.
- Morgan, A. R. & Wells, R. D. (1968) *J. Mol. Biol.* **37**, 63–80.
- Chamberlin, M. J. (1965) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **24**, 1446–1457.
- Inman, R. B. (1964) *J. Mol. Biol.* **9**, 624–637.
- Wells, R. D. & Larson, J. E. (1972) *J. Biol. Chem.* **247**, 3405–3409.
- McLean, M. J., Blahó, J. A., Kilpatrick, M. W. & Wells, R. D. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5884–5888.
- Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
- Lilley, D. M. J. & Palecek, E. (1984) *EMBO J.* **3**, 1187–1192.
- Lukasova, E., Jelen, F. & Palecek, E. (1982) *Gen. Physiol. Biophys.* **1**, 53–70.
- Galazka, G., Palecek, E., Wells, R. D. & Klysik, J. (1986) *J. Biol. Chem.* **261**, 7093–7098.
- Johnston, B. H. & Rich, A. (1985) *Cell* **42**, 713–724.
- Herr, W. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 8009–8013.
- Wells, R. D., Collier, D. A., Hanvey, J. C., Shimizu, M. & Wohlrab, F. (1988) *FASEB J.*, in press.
- McLean, M. J. & Wells, R. D. (1988) *J. Biol. Chem.* **263**, 7370–7377.
- McLean, M. J., Lee, J. W. & Wells, R. D. (1988) *J. Biol. Chem.* **263**, 7378–7385.
- Arnott, S. & Selsing, E. (1974) *J. Mol. Biol.* **88**, 509–521.
- Jaworski, A., Hsieh, W.-T., Blahó, J. A., Larson, J. E. & Wells, R. D. (1987) *Science* **238**, 773–777.
- Voloshin, O. N., Mirkin, S. M., Lyamichev, V. I., Belotserkovskii, B. P. & Frank-Kamenetskii, M. D. (1988) *Nature (London)* **333**, 475–476.
- Kohwi, Y. & Kohwi-Shigematsu, T. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 3781–3785.