J. Mol. Biol. (1984) **172**, 355–362

Molecular Cloning and Sequence Analysis of *trp–lac* Fusion Deletions

DNA fragments containing deletions that fuse the *trp* operon to the *lac* operon were cloned and the end-points of the fusions were determined. The results from DNA sequence analysis correlated well with those from genetic, biochemical and physiological studies previously reported. The sequence data from this study, in combination with the known properties of these fusion strains, provided information on: (1) the precise *lac* operon distal boundary of the *lac* operator; (2) the nature of the *trp* operon terminator; and (3) the messenger RNA sequences that result in inhibition of *lacZ* translation initiation in *trp–lac* fused mRNAs.

The fusion of *Escherichia coli* genes or operons has been useful in the genetic analysis of operon structures and regulation (Bassford *et al.*, 1978). Of particular interest are the *trp–lac* fusions, whose isolation was described by Mitchell *et al.* (1975). The deletions that generated these fusions all have their *lac* end-points within the *lac* regulatory region. Genetic, physiological and biochemical experiments with some of these *trp–lac* fusion deletions suggested that they have some interesting properties, which we have studied by DNA sequence analysis in this letter. For instance, two deletions (F36a and W205) fuse *lac* to *trp* but do not cut into the *trp* structural genes, thus they help define the *trp* operon terminator. Two other deletions (W211 and W225) appear to define the operon distal boundary of the *lac* operator, since they reduce the repressor–operator affinity by twofold and 11-fold, respectively (Reznikoff *et al.*, 1974*b*). Finally, the *in vivo* β-galactosidase content of these fusion strains varies within a 60-fold range, and this was found to be partially due to the differing efficiencies of translational initiation for different fusions (Reznikoff *et al.*, 1974*a*).

The *E. coli trp–lac* fusion deletions F36a, W200, W205, W211, W225, W227 and R189 are shown schematically in Figure 1 (Miller *et al.*, 1970; Mitchell *et al.*, 1975). In order to facilitate the sequence analysis of these deletions, they were crossed onto a ColE1 derivative (pRZ115-R189) that carries the trp^+–lac^- fusion R189. This was accomplished by transforming each of the appropriate *trp–lac* fusion strains with pRZ115-R189 and isolating the plasmid DNA, which was then used to transform a Lac^- strain. The resulting transformants that were Lac^+ were assumed to contain plasmids that had incorporated the relevant Lac^+ fusion *via* homologous recombination events between the chromosome and plasmid pRZ115-R189. These plasmids were checked by restriction endonuclease analyses. The *Hin*dII fragments containing some of the *trp–lac* fusions were subcloned into pVH51 prior to sequence analysis.

The DNA sequence analysis by primed synthesis was performed as described by Smith (1980). The DNA sequence analysis by chain chemical cleavage was carried out as described by Maxam & Gilbert (1980).

0022-2836/84/030355-08 $03.00/0 © 1984 Academic Press Inc. (London) Ltd.

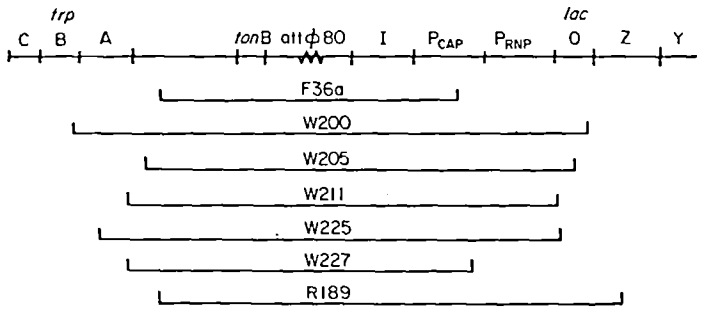

FIG. 1. *lac* and *trp* ends of fusion deletions. The symbols *trpC*, *B* and *A* and *lacZ* and *Y* refer to structural genes in the *trp* and *lac* operons. The other genetic designations include *tonB* (encodes a function required for ϕ80 infection and colicin V and B killing), attϕ80 (an attachment site for the ϕ80*lac* prophage), *I* (encodes the *lac* repressor), P_{CAP} (the site required for CAP binding in the *lac* controlling elements), P_{RNP} (the RNA polymerase binding site for *lac*) and *O* (the *lac* operator).

(a) *Specificity of deletion end-points*

The sequenced end-points of the *trp–lac* fusions F36a, W200, W205, W211, W225 and W227 are shown in Figure 2.

It has been observed that in many cases spontaneous deletions occur between repeated sequences (Miller, 1978). None of the six sequenced fusion deletions described in this letter occurred between extensively repeated sequences (the W200 deletion has end-points within the repeated sequence G-A-A-A, the W225 deletion was between the related sequences A-A-T-G-T (*trp*) and A-A-T-T-G-T (*lac*), the W211 deletion within repeated sequence T-G) (see Fig. 2). The lack of obvious extensive sequence homology between deletion end-points may reflect the fact that the fusion strains were obtained through specific selection procedures constraining their end-points.

(b) *The* trp *end-point of the deletions*

Sequence data presented here indicate that fusion strains F36a and W205 have intact *trp*A genes, while the W211 and W227 deletions have removed 23 and 25 base-pairs from the C-terminal end of *trp*A gene. Because the stop codon at the end of the *trp*A gene was deleted, translation from the *trp*A gene could read through into the *lac* regulatory region, terminating at an UGA codon about 30 nucleotides before the initiation codon of *lac*Z (see Fig. 2). As a result, the *trp*A product made in W211 is six amino acids shorter than normal. In the case of W227, seven C-terminal amino acid residues were removed, and 18 amino acid residues were added. There is no sequence homology between the amino acid residues removed and added (see Fig. 3). The sequence results are in good agreement with the results from genetic mapping and physiological and protein structure studies, which indicate that F36a and W205 made a normal *trp*A product, whereas W211 and W227 made a mutant *trp*A protein that retained part of *trp*A enzyme activity at 37°C (Mitchell *et al.*, 1975,1976). Apparently, the removal of six to seven amino acid residues from the C-terminal end of the *trp*A

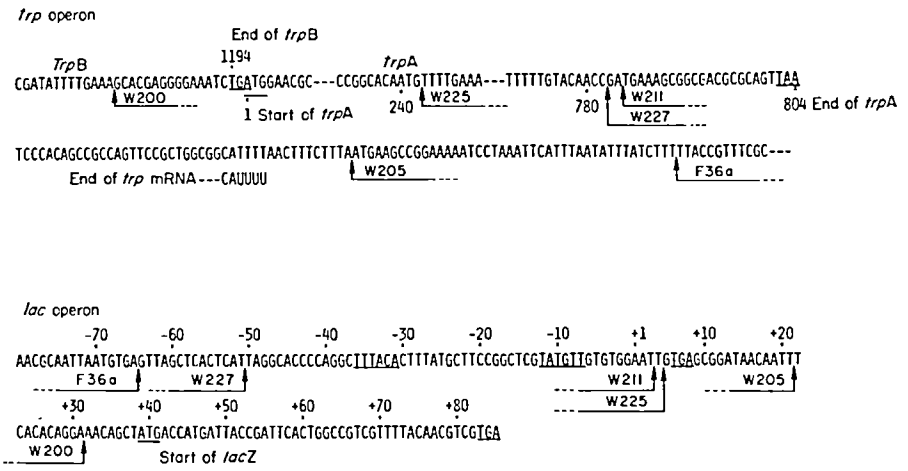

FIG. 2. End-points of fusion deletions F36a, W200, W205, W211, W225 and W227 in *trp* and *lac* operons. Base-pairs in the *trp* operon are numbered relative to the start sites of the *trp*B or *trp*A genes. Base-pairs in the *lac* operon are numbered relative to the site of *lac*Z transcription initiation. The translation initiation codon of the *trp*A gene, the translation termination codons of *trp*B and *trp*A, the −35 and −10 consensus sequences of the *lac* promoter and the translation initiation codon of the *lacZ* gene are underlined. The 2 potential translation termination codons in the *lac* regulatory region are also underlined. Translation readthrough from *trp* in fusion strains W211, W225 and W227 stops at a UGA codon positioned at base-pair +6. Translation readthrough from *trp*B in fusion strain W200 stops at a U-G-A codon positioned at base-pair +83. Arrows in *trp* operon regions indicate that the nucleotides downstream are deleted. Arrows in the *lac* operon region indicate that the nucleotides upstream are deleted. The *trp* and *lac* operon sequence is according to Nichols & Yanofsky (1979) and Dickson *et al.* (1975). Hyphens have been omitted from the sequences for clarity.

product does not totally inactivate the enzyme, although cells with these deletion mutations cannot grow under very stringent conditions (at 42°C with 5-methyltryptophan).

In fusion W225, the deletion removed two thirds of the *trp*A gene, and in W200 the entire *trp*A gene was deleted. These strains had been previously shown to be phenotypically TrpA$^-$ (Mitchell *et al.*, 1975). The W200 deletion also removed the last five amino acids from the *trp*B gene and allowed the synthesis of a "readthrough" protein that contained 18 additional amino acids. Since previous studies have shown that the *trp*B protein in W200 is active (Mitchell *et al.*, 1975), it is evident that the five C-terminal amino acids are not required for, and the additional amino acids do not interfere with, the *trp*B protein activity.

(c) *The* trp *terminator*

Fusion strains F36a and W205 are characterized as phenotypically TrpA$^+$, that is, they make perfectly normal *trp*A protein (tryptophan synthetase A) (Mitchell *et al.*, 1976). Since the fusion of two operons requires the elimination of the transcriptional terminators between them, it was proposed that the *trp* end-points of deletions F36a and W205 were located in between the *trp*A gene and the transcription terminator of the *trp* operon. The deletion end-points of F36a and

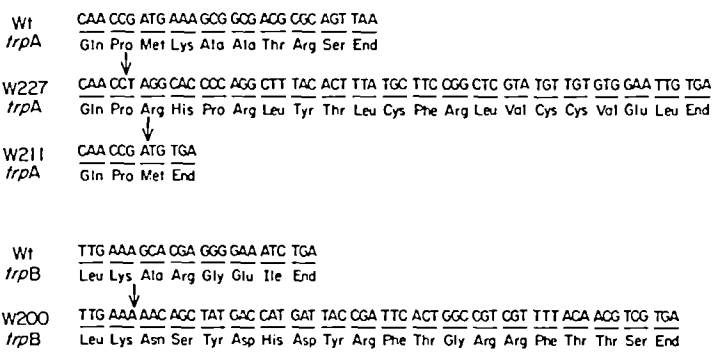

FIG. 3. C-Terminal nucleotide and amino acid sequence of the *trp*A and *trp*B proteins in fusion strains W200, W227 and W211. The arrows indicate the deletion junction points. Wt, wild type.

W205 are 11 and 53 base-pairs past *trp* t (terminator), respectively (see Fig. 2). These deletions are thus similar to a set of fusions reported previously that delete DNA downstream from *trp* t but allow readthrough transcription from the *trp* promoter (Wu *et al.*, 1980). A second terminator has since been found 250 base-pairs distal to *trp* t (designated *trp* t′) and is thought to be the terminator normally used *in vivo* (Wu *et al.*, 1981). The fact that 3′ end of the mature *trp* message is coincident with *trp* t suggests that this is a processing site and not a transcription termination site.

(d) *The* lac *operator*

The W225 fusion junction in the *lac* operator deletes one more base-pair from the *lac* regulatory region than the W211 fusion. In fact, W225 fortuitously fuses an A-A-T-G sequence from *trp*A to the residual *lac*O, thus generating a single base-pair substitution that is one base-pair downstream from the W211 end-point; *lac*O = A-A-T-T, W225 = A-A-T-G (see Fig. 4). The *in vitro* affinity of the *lac* repressor for the DNA fragment carrying W211 was found to be about fivefold greater than that for the DNA fragment carrying W225 (W211 DNA binds repressor with a twofold lower affinity than does O^+ DNA, while W225 DNA binds repressor with an 11-fold lower affinity than does O^+ DNA (Reznikoff *et al.*, 1974*b*)). Biochemical studies have shown that the T·A base-pair deleted in fusion W211 is the left-most base-pair protected by the *lac* repressor from methylation by dimethyl sulfate, suggesting that it interacts directly with the *lac* repressor (Ogata & Gilbert, 1978). Crosslinking of the *lac* repressor to bromouracil-substituted DNA upon u.v. irradiation also suggests that this T·A base-pair is in close proximity to the bound *lac* repressor (Ogata & Gilbert, 1977). Three further facts strongly suggest that the W211 deletion end-point defines the left-side boundary of the *lac* operator. (1) Studies of *lac* repressor binding to synthetic operator sequences by Bahl *et al.* (1977) indicate that the minimal *lac* operator sequence that binds to the repressor with a high affinity is defined at one end by the T·A base-pair removed by W211. (2) All known O^c mutations are located downstream from this position (+4). (3) A mutation at base +1 has no effect on

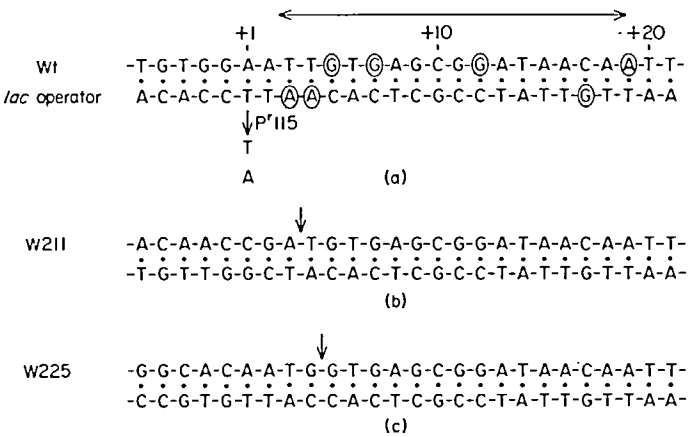

FIG. 4. Alternated *lac* operator regions in fusion deletions W211 and W225. (a) Wild-type (Wt) *lac* operator. The base-pairs are numbered relative to the site of *lacZ* transcription initiation. The arrow between base-pairs +3 and +19 indicates the synthetic operator that binds repressor with natural affinity (Bahl *et al.*, 1977). P'115 is an A·T to T·A transversion at position +1 which does not affect *lac* repressor binding. Purine residues that *lac* repressor protects from methylation by dimethyl sulfate are encircled (Ogata & Gilbert, 1978; Maquat *et al.*, 1980). (b) and (c) *lac* operator DNA generated by deletions W211 and W225. The arrows indicate the deletion junction points.

repressor binding, either *in vivo* or *in vitro* (Maquat *et al.*, 1980; Reznikoff *et al.*, 1974*b*) (see Fig. 4). Thus, it is very likely that the precise left-side boundary of the *lac* operator lies in between the A·T and T·A base-pairs that fusion W211 deleted.

(e) *The* lac *promoter*

A two-site model for the *lac* promoter was proposed based on genetic studies. One site is for the interaction of the CAP–cAMP complex, and the other for the interaction of RNA polymerase (Beckwith *et al.*, 1972). Fusions F36a and W227 have deletion end-points that lie within the CAP–cAMP complex binding site (see Fig. 2). This is consistent with the results of previous studies, which indicate that both fusions retain full promoter activity but are unresponsive to stimulation by the CAP–cAMP complex (Mitchell *et al.*, 1975).

(f) *Efficiency of* lac*Z translation initiation for readthrough transcripts*

The synthesis of the *lacZ* gene product from the different *trp–lac* fusions varies within a 60-fold range, even though the transcription is initiated from the same promoter (the derepressed *trp* promoter) (Mitchell *et al.*, 1975). The experiments by Reznikoff *et al.* (1974*a*) suggest that this is primarily due to different frequencies of translation initiation at the start site of *lacZ* for different fusions. Several hypotheses were proposed to explain the depressed initiation of the *lacZ* gene from fusion mRNA. (1) The deletions may have removed some nucleotides that normally code for part of the *lacZ* ribosome binding site. (2) Translational

readthrough from *trp* structural genes might compete with the correct binding of new ribosomes at the initiation site of *lacZ*. (3) Abnormal RNA sequences have been added onto the *lac* mRNA close to the *lacZ* ribosome binding site. It is possible that these sequences alter the configuration of the *lacZ* ribosome binding site or make it structurally inaccessible, thereby reducing the affinity of ribosomes for the site (Reznikoff *et al.*, 1974*a*). The determination of the fusion end-points in six fusion strains by sequence analysis made it possible to examine more closely the causes of repressed initiation of *lacZ* from fusion mRNA in five of these strains.

In fusion strain W200, the *lac* deletion end-point is in the purine-rich region (A-G-G-A) known as the Shine–Delgarno sequence, which is complementary to the 3′ end of 16 S ribosomal RNA of the *E. coli* ribosome and has been shown to play a role in translation initiation (Grunberg-Manago, 1979). The substitution of A-G-G-A by G-A-A-A in W200 may reduce the strength of the *lacZ* ribosome binding site. Another factor possibly contributing to the low level of translation initiation from *lacZ* in W200 is the translational readthrough from the *trpB* gene, which terminates in a stop codon 41 nucleotides past the initiation codon of *lacZ* (see Fig. 2). Readthrough-translation competition with the correct binding of new ribosomes has been postulated by Platt *et al.* (1972) to explain the necessity for a chain-termination codon prior to an initiation codon in order for initiation to occur. On the other hand, studies by Schumperli *et al.* (1982) have shown that for the *gal* operon, translation initiation of a downstream gene is approximately the same when upstream translation stops either prior to or well into the downstream gene.

W200 is the only deletion that we sequenced in which translation readthrough is across the translational initiation site of *lacZ*. The other three readthrough-translation events from *trpA* (W211, W225 and W227) would stop at the same UGA codon 30 nucleotides prior to the initiation site of the *lacZ* gene (see Fig. 2). The β-galactosidase content of two fusion strains, F36a and W227, is extremely low relative to that of the other fusion strains. For example, the amount of β-galactosidase found in W227 is more than 15-fold lower than that found in W211, whereas the amount of thiogalactoside transacetylase found in W227 is only about twofold lower than that found in W211 (Mitchell *et al.*, 1975). From the sequence data, these two strains have virtually the same deletion end-points in *trpA*. However, W211 deleted the whole RNA polymerase binding site of the *lac* promoter, which was retained in W227 (see Fig. 2). Since the deletion end-points in *lac* of fusions F36a and W227 are 80 to 100 nucleotides preceding the *lacZ* translation initiation site, it is appropriate to assume that no sequence signals required for translation initiation from *lacZ* have been deleted. One thing in common to F36a and W227 is that their fusion mRNAs have 50 to 70 more nucleotides in the 5′ end than other fusion *lac* mRNAs (W211 and W225). We believe that the presence of the fused abnormal RNA (transcripts from the *lac* promoter region) alters the accessibility of the *lacZ* ribosome binding site. Figure 5 illustrates a possible secondary structure of the 5′-terminal segment of the fusion mRNAs in strains F36a and W227. The Shine–Dalgarno sequence of *lacZ* in the

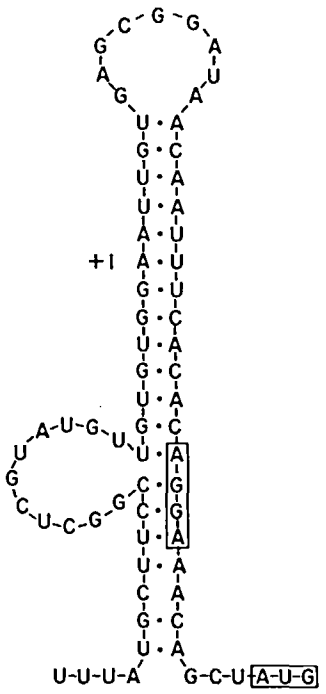

FIG. 5. A possible secondary structure for the fusion mRNA in strains F36a and W227. The initiation codon AUG and Shine–Dalgarno sequence of *lacZ* are indicated by boxes. The *lacZ* transcription initiation site is indicated as +1. The thermodynamic stability of this structure ($\Delta G = -16{\cdot}6$ kcal (1 kcal = 4184 J)) is estimated following the rules of Tinoco *et al.* (1973).

fusion mRNAs is locked into a stable stem-and-loop structure; therefore, the accessibility of ribosomes to it is greatly reduced.

It should be noted that the stability of the structure pictured in Figure 5 is primarily due to the extensive 2-fold symmetry found from base-pairs −7 to +28 in the *lac* controlling elements and that the existence of this symmetry element has no known function (with the exception of the region between base-pairs +3 and +19, which dictates repressor binding). Perhaps it exists in order to inhibit *lacZ* translation on transcripts that fortuitously read through the *lac* controlling elements. If this were true, one would assume that this same property might be manifest in other gene systems. In fact, inefficient translation initiation for *gal*E on *λ-gal*E readthrough transcripts has been reported and a similar mRNA secondary structure has been postulated to play a role in *gal*E translation initiation inhibition (Merril *et al.*, 1978,1981).

This work was supported by grant GM-19670 from the National Institutes of Health, grant PCM 79-10686 from the National Science Foundation and a grant from the 3M Foundation. L.M.M. was in part supported by predoctoral training grant GM07215 from

the NIH. Y.X.-M. was in part supported by the People's Republic of China and a University of Wisconsin Graduate School fellowship.

Department of Biochemistry
College of Agricultural and Life Sciences
University of Wisconsin-Madison
Madison, WI 53706, U.S.A.

YU XIAN-MING
LIANNA M. MUNSON
WILLIAM S. REZNIKOFF

Received 7 June 1983, and in revised form 6 October 1983

REFERENCES

Bahl, C. P., Wu, R., Stawinsky, J. & Narang, S. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 966–970.

Bassford, P., Beckwith, J., Berman, M., Brickman, E., Casadaban, M., Guarente, L., Saint-Girons, I., Sarthy, A., Schwartz, M., Shuman, H. & Silhavy, T. (1978) In *The Operon* (Miller, J. H. & Reznikoff, W. S., eds), pp. 245–261, Cold Spring Harbor Laboratory, New York.

Beckwith, J. R., Grodzicker, T. & Arditti, R. (1972). *J. Mol. Biol.* **69**, 155–160.

Dickson, R. C., Abelson, J. M., Barnes, W. M. & Reznikoff, W. S. (1975). *Science,* **182**, 27–35.

Grunberg-Manago, M. (1979). In *Ribosomes: Structure, Function, and Genetics* (Chambliss, G., Craven, G. R., Davies, J., Davis, K., Kahan, L. & Normura, M., eds), pp. 445–477, University Park Press, Baltimore.

Maquat, L. E., Thornton, K. & Reznikoff, W. S. (1980). *J. Mol. Biol.* **139**, 537–560.

Maxam, A. M. & Gilbert, W. (1980). In *Methods in Enzymology* (Grossman, L. & Moldave, K., eds), vol. 65, pp. 499–560, Academic Press, New York.

Merril, C., Gottesman, M., Cowit, D. & Adhya, S. (1978). *J. Mol. Biol.* **118**, 241–245.

Merril, C., Gottesman, M. & Adhya, S. (1981). *J. Bacteriol.* **147**, 875–887.

Miller, J. H. (1978). In *The Operon* (Miller, J. H. & Reznikoff, W. S., eds), pp. 31–88, Cold Spring Harbor Laboratory, New York.

Miller, J. H., Reznikoff, W. S., Silverstone, A. E., Ippen, K., Singer, E. R. & Beckwith, J. R. (1970). *J. Bacteriol.* **104**, 1273–1279.

Mitchell, D. H., Reznikoff, W. S. & Beckwith, J. R. (1975). *J. Mol. Biol.* **93**, 331–350.

Mitchell, D. H., Reznikoff, W. S. & Beckwith, J. R. (1976). *J. Mol. Biol.* **101**, 441–457.

Nichols, B. P. & Yanofsky, C. (1979). *Proc. Nat. Acad. Sci., U.S.A.* **76**, 5244–5248.

Ogata, R. & Gilbert, W. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 4973–4976.

Ogata, R. & Gilbert, W. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 5851–5854.

Platt, T., Weber, K., Ganem, D. & Miller, J. H. (1972). *Proc. Nat. Acad. Sci., U.S.A.* **69**, 897–901.

Reznikoff, W. S., Michels, C. A., Cooper, T. G., Silverstone, A. E. & Magasanik, B. (1974*a*). *J. Bacteriol.* **117**, 1231–1239.

Reznikoff, W. S., Winter, R. B. & Hurley, C. K. (1974*b*). *Proc. Nat. Acad. Sci., U.S.A.* **17**, 2314–2318.

Schumperli, D., McKenney, K., Sobieski, D. A. & Rosenberg, M. (1982). *Cell,* **30**, 865–871.

Smith, A. J. H. (1980). In *Methods in Enzymology* (Grossman, L. & Moldave, K., eds), vol. 65, pp. 560–580, Academic Press, New York.

Tinoco, I. Jr, Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973). *Nature (London),* **246**, 40–41.

Wu, A. M., Chapman, A. B., Platt, T., Guarente, L. P. & Beckwith, J. H. (1980). *Cell,* **19**, 829–836.

Wu, A. M., Christie, G. G. & Platt, T. (1981). *Proc. Nat. Acad. Sci., U.S.A.* **78**, 2913–2917.

Edited by M. Gottesman