Ribitol dehydrogenase of *Klebsiella aerogenes*

Sequence of the structural gene

Thérèse LOVINY, Pamela M. NORTON* and Brian S. HARTLEY†

Department of Biochemistry, Imperial College of Science and Technology, London SW7 2AZ, U.K.

(Received 11 December 1984/15 March 1985; accepted 30 May 1985)

The ribitol dehydrogenase gene was cloned from wild-type *Klebsiella aerogenes* and also from a transducing phage λ prbt which expresses the *rbt* operon constitutively. The coding sequence for 249 amino acids is separated from the following D-ribulokinase gene by 31 base pairs containing three stop codons, one of which overlaps the ribosome binding site for D-ribulokinase. Three residues in the amino acid sequence differ from that predicted from the DNA sequence: Asp-212 for Asn-212 is probably a protein sequencing error, but -Ala-Val- for -Ser-Ser- at 146–147 appears to be a 'neutral mutation' that may have arisen during prolonged chemostat selection of a strain that superproduces the enzyme from which the protein sequence was determined.

K. aerogenes grows on ribitol or D-arabitol by virtue of two contiguous but independently inducible operons. The ribitol (*rbt*) operon is induced by D-ribulose (Charnetzky & Mortlock, 1974a) and contains the *rbtD* gene encoding ribitol dehydrogenase (RDH) and *rbtK* encoding D-ribulokinase (DRK). The D-arabitol operon, induced by D-arabitol (Charnetzky & Mortlock, 1974b), contains *dalD* encoding D-arabitol dehydrogenase (ArDH) and *dalK* encoding D-xylulokinase (DXK). *E. coli* C contains similar operons (Scangos & Reiner, 1978) and its ribitol dehydrogenase shows 95% homology with that of *K. aerogenes* (Altosaar & Hartley, 1976), but *E. coli* B and *E. coli* K lack these operons and cannot utilize pentitols.

Rigby *et al.* (1976) used P1 transduction to transfer these operons from *K. aerogenes* into *E. coli* K12 where they were found to map in a similar chromosomal location to that in *E. coli* C (40 min) and from this strain (*E. coli* EA) Neuberger & Hartley (1979) constructed specialized transducing phages λ prbt and λ prbt *dal*, which confer on their lysogens the ability to grow on one or both pentitols respectively. The fine structure of these two bacteriophages shows that the operons are transcribed divergently from opposite strands of the

DNA and that the intervening region encodes a *rbt* repressor (*rbt R*) and a *dal* repressor (*dal R*). Moreover, the two operons resemble a 'metabolic transposon' in that they are enclosed within a large stretch of repeating sequence.

The gene for ribitol dehydrogenase, *rbtD*, has been the subject of an extensive study of experimental enzyme evolution (Rigby *et al.*, 1974; Hartley *et al.*, 1976). Strains constitutive for synthesis of this enzyme grow on xylitol by virtue of a weak side specificity for this substrate, but this activity is growth-limiting. Hence mutants are selected in continuous culture on xylitol that either superproduce the wild type enzyme (RDH-A) or produce mutant enzymes with increased specific activity towards xylitol. Ribitol dehydrogenase-A is a tetramer of four identical subunits (Taylor *et al.*, 1974) and the previous paper (Dothie *et al.*, 1985) described the evidence for its amino acid sequence (249 residues) and those of two 'evolvants' with increased xylitol activity, RDH-B and RDH-D.

These pentitol operons exhibit several features of evolutionary interest. Their structure invites speculation that they may have arisen by an act of invert gene duplication, which would stabilize the second copy against excision by recombination and allow divergence to a new enzyme activity: recruitment of the other genes and control proteins would then have followed. The flanking repeat sequences may be a relic of this original duplication or may have acquired biological value in

Abbreviations used: RDH, ribitol dehydrogenase; bp, base pairs; kb, kilobases.

* Present address: Department of Biochemistry, St. Mary's Hospital Medical School, London W2, U.K.

† To whom reprint requests should be addressed.

facilitating interspecies transfer of pentitol metabolism, such as that achieved in the artificial transfer of the genes from *K. aerogenes* to *E. coli* K12. The DNA sequences of both operons might provide evidence for such speculation. Moreover, gene-duplication has been implicated in the super-production of ribitol dehydrogenase-A elicited by continuous culture on xylitol (Rigby *et al.*, 1974; Neuberger *et al.*, 1979) and the DNA sequences may indicate the mechanisms involved.

We have therefore attempted to determine the complete DNA sequence of both operons. The present paper describes the sequence of the *rbtD* gene and its control region.

Materials and methods

Materials

Restriction endonucleases were purchased from New England Biolabs, Bethesda Research Laboratories or Boehringer Mannheim. T₄ DNA ligase and the replicative form of the vectors M13 mp8 and mp9 were obtained from Biolabs. DNA polymerase I (Klenow subfragment) was supplied by Boehringer or Biolabs. The deoxy- and dideoxynucleotides were obtained from P-L Biochemicals. Purified bacterial alkaline phosphatase and T₄ polynucleotide kinase were a gift from Dr. D. Glover (Imperial College). A 17mer synthetic primer was a generous gift from Dr. G. P. Winter (MRC, Cambridge, U.K.) as were *E. coli* JM 101 strain and the replicative form of M13 mp7 vector. Two other synthetic oligonucleotides used as primers were purchased from Collaborative Research Inc. or Biolabs. Radioisotopes were

obtained from The Radiochemical Centre, Amersham, Bucks., U.K.

DNA preparations

The preparation of DNA from plasmids and from the phage λ prbt has already been described (Loviny *et al.*, 1981). DNA fragments used for sequencing were isolated on a preparative scale by digestion with suitable restriction endonucleases according to the manufacturers' protocol, separation on agarose gel and elution by simple electrophoresis (Galibert *et al.*, 1974) or electrophoresis on DEAE-cellulose paper (Dretzen *et al.*, 1981).

DNA sequencing

Two methods were used, the chemical method of Maxam & Gilbert (1977) and the chain termination method of Sanger *et al.* (1977) using a universal primer and the M13 cloning vectors developed by Messing *et al.* (1981) and Messing & Vieira (1982). Details of the isolation of end-labelled restriction fragments and sequencing by the chemical method have been described by Loviny *et al.* (1981). With the chain termination technique, a 'shotgun' approach with a 4kb isolated fragment (Fig. 1) was used essentially as described by Sanger *et al.* (1980), in combination with the more selective technique of cloning into the specific M13 mp8 and mp9 sites (Messing & Vieira, 1982). In the later stages of the work, it was very useful to select clones by hybridization according to Winter & Fields (1980). For the dideoxy method, hybridizations were achieved in 10mM-Tris/HCl, pH 8.5, containing 10mM-MgCl₂, at 100°C for 3 min, or later in the work at 55°C for

Fig. 1. Restriction map of the *rbt-dal* region of the *K. aerogenes* chromosome

The extent of *K. aerogenes* chromosomal DNA inserted in λ prbt, λ prbt dal and pJCW1 is shown by vertical arrows, with the positions of the three fragments isolated for sequencing purpose in the hatched areas. Above the restriction map, the pentitol operons are briefly summarized, the horizontal arrows representing the directions of transcription. R, repressor genes; D, dehydrogenase genes; K, kinase genes. Sizes are given in kb. (From C. Smith, unpublished results.)

20 min, followed by 15–30 min cooling to room temperature. Recent use of [α - 35 S]dATP (Biggin *et al.*, 1983) in the polymerization step of the dideoxy method improved significantly the resolution of the products on acrylamide gels.

Sequencing gels

Urea/polyacrylamide gels used routinely were as published (Sanger *et al.*, 1980; Loviny *et al.*, 1981). Drying gels before exposure to X-ray films increased considerably the resolution of products, following Garoff & Ansorge (1981). Treatment of one of the glass plates with silane (Wacker Chemie, Munchen, Germany) made easier the manipulation of gels and the 0.4 mm-thick gels were routinely dried for 90 min at 90°C after acetic acid fixation. The genome of *K. aerogenes* being G+C-rich, we often encountered important compressions in the 'sequence ladders' on the normal 6%-urea gels, where strong G·C hairpin loops were subsequently revealed. These compressions were efficiently resolved using the thermostating system of Garoff & Ansorge (1981) at 65–70°C: the light background of bands visible at high temperature all along the nucleotide ladder is of great help in ensuring complete denaturation of the DNA. Finally, buffer gradient gels (Biggin *et al.*, 1983) were the last refinement to the gel electrophoresis step.

Computing

Computer analyses were performed with the programs (BATIN and BATOVR) supplied by Dr. R. Staden (MRC, Cambridge, U.K.) (Staden, 1980) and adapted to the Imperial College CD6500 and IBM400 computers by S. Cox (Department of Computing, Imperial College), with the help of T. Knott (Department of Biochemistry, Imperial College).

Results and discussion

Sources of DNA

Fig. 1 shows a restriction map of the *rbt dal* region in the original inducible wild-type strain of *K. aerogenes* (strain FG5, *arg, gua*) (Wu *et al.*, 1968). Plasmid pJCW1 contains a *Bst*I fragment from strain FG5 cloned by Wu (1983) into the *Bam*H1 site of plasmid pBR322 (Bolivar *et al.*, 1977). A 0.7 kb fragment (*Bgl*II–*Pst*B) and a 1.3 kb fragment (*Sal*C–*Bst*A) of this plasmid were used in some of the later sequencing studies, but most work was done with DNA from *lprbt* or *lprbt dal* (Neuberger & Hartley, 1979). This DNA derives from *K. aerogenes* strain FG5 (*arg, gua*) via strain X1 of Wu *et al.* (1968) (*arg, gua, rbt101*; constitutive for synthesis of ribitol dehydrogenase). *K. aero-*

genes strain A, which was the ancestral strain in the experimental evolution studies, is a prototrophic revertant (*rbt101*) of strain X1. DNA was transferred from *K. aerogenes* strain A into *E. coli* strain EA and thence into *lprbt* and *lprbt dal*. This lineage is relevant to our subsequent discussion.

Sequencing strategy

Fig. 2 summarizes the strategy used to sequence the *rbtD* gene of *K. aerogenes*. The start of the *rbt* operon transcription had been previously localized between *Sal*C and *Hind*A (Fig. 1) by hybridization of a *lprbt* DNA digest with pulse-labelled mRNA extracted from an RDH-superproducer mutant (Neuberger & Hartley, 1979). The first sequences, which were obtained by the chemical method with *lprbt* DNA, span this region to ascertain the correct position of the *rbtD* gene from the known protein sequence (Fig. 2a). Subsequently, a bank of sequence data for the whole *rbt* operon was created by the shotgun approach of the dideoxy method with the 4 kb *Sal*B–*Sal*C fragment isolated from *lprbt* (Fig. 2b). *Taq*I and *Hpa*II digests were cloned into *Acc*I-digested vector M13 mp7, *Sau*3A digests into *Bst*I-cut mp7 and *Alu*I digests into the *Hin*CII site of mp7 after phosphatase treatment of the cut vector (Winter *et al.*, 1981). Assembly of the data was achieved by the BATIN and BATOVR programs of Staden (1980). From the same 4 kb fragment, two *Pst*B–*Bgl*II clones were also obtained, as inserts in M13 mp8 and mp9 in opposite orientations: they were used at the final stage of the work for selection of desired clones by hybridization (Winter & Fields, 1980). The clones in Fig. 2(c) were prepared from the 0.7 and 1.3 kb fragments of pJCW1 DNA, by double digests of the fragments inserted into appropriate sites in M13 mp8 and mp9. It was necessary to fill in the sites produced by *Bst* EII and *Cla*I in order to ligate them to the blunt-ended *Sma*I site of the vectors (Deininger, 1983). Finally, a *Pst*I clone obtained from the whole pJCW1 plasmid DNA spanned the *Sal*C site.

Sequence of the *rbtD* gene

The nucleotide sequence of the *rbtD* gene is presented in Fig. 3. Most regions of the sequence were covered by two or more independent clones. Although a small region was covered by one clone only, its sequence analysis presented no ambiguity and was repeated for confirmation.

The nucleotide sequence is in good agreement with the protein sequence except for three amino acids which are boxed in Fig. 3, the change Ala-Val for Ser-Ser at residues 146–147 being the most significant. The change Asn-Gly for Asp-Gly at residues 212–213 is probably a protein sequencing

Fig. 2. Clones used in establishing the DNA sequence of the *rbtD* gene

(a) By the chemical method, using the whole λ *prbt* DNA; (b) by the chain-termination method, using the 4 kb fragment isolated from λ *prbt* DNA; (c) by the chain-termination method, using the 0.7 and 1.3 kb fragments isolated from pJCW1 DNA and the whole pJCW1 plasmid DNA. The arrows show the directions of the clones and their lengths indicate the exact length of the sequences determined. The stars * represent the ends of the DNA fragments which were labelled in the chemical technique.

error, since this sequence readily deamidates during manipulations at alkaline pH. The region of the DNA sequence where these changes have occurred has been thoroughly checked using the DNA from λ *prbt* as well as the DNA from plasmid pJCW1. As mentioned earlier, λ *prbt* DNA is constitutive for the *rbt* operon and has been through many genetic manipulations and replication cycles in *E. coli*. In contrast pJCW1 DNA is inducible for the operon and was recently prepared directly from the wild type *K. aerogenes* FG5 genome and replicated in *E. coli*. No difference at all was observed between the sequences obtained with the two different DNAs for the RDH protein.

A possible ribosome attachment site (S/D) is present 7 bp before the *rbtD* initiation codon (Shine & Dalgarno, 1974), and 29 bp separate the last *rbtD* coding triplet from the initiator of the next gene *rbtK*, recognized from a sequenator analysis of purified D-ribulokinase (Neuberger *et al.*, 1979) which gave Met-His-Asn-Asp-Thr-Glx-Asn-Ile-Ile-Gly-Val-Asp (Neuberger *et al.*, 1981). This intergenic region contains three termination codons in different reading frames and a ribosome binding site (S/D) for the *rbtK* gene immediately 5' to the *rbtK* initiation codon. The two genes are not of the same reading frame.

Discussion

Base composition and codon usage in *rbtD*

The total base composition of the *rbtD* gene of *K. aerogenes* is: T, 19.4%; C, 30.4%; A, 18.9%; G, 31.3%, which gives a G + C content of 61.7%. The evolution of the *K. aerogenes* genome towards a total G + C content higher (56%) than *E. coli* (51%) and *S. typhimurium* (52%) has been documented (Sober, 1979), and Nichols *et al.* (1981) found a 64.8% G + C content in their *trpA* nucleotide sequence of *K. aerogenes*.

This high G + C content of the *rbtD* gene is reflected in the very high frequency of codons (79%) ending in either G or C. Table 1 shows the codon frequencies in three different structural genes of *K. aerogenes*: R, *rbtD* gene, the present work; T, *trpA* gene (Nichols *et al.*, 1981); D, *dalD* gene (Knott, 1982); the number of residues in each respective protein is 249, 269 and 454. They are compared with a percentage use of codons compiled from 25 non-regulatory genes of *E. coli* by Konigsberg & Godson (1983). Notable differences in *K. aerogenes* are the preference of GUC and GUG for Val, UCG for Ser, GCC and GCG for Ala, CAG for Gln, GAG for Glu, UGC for Cys, CGC for Arg and GGC + GGG for Gly. The

Fig. 3. Sequence of the *rbtD* gene of *K. aerogenes*

Table 1. *Codon frequencies in K. aerogenes*
The codon usage in ribitol dehydrogenase (R) (this work), *trp A* (T) (Nichols *et al.*, 1981) and D-arabitol dehydrogenase (D) (Knott, 1983) are averaged as a percentage for each alternative (K) and compared with a similar average for 25 non-regulatory genes of *E. coli* (E) (Konigsberg & Godson, 1983).

Codon	R	T	D	K (%)	E (%)
Phe UUU	1	4	6	31	44
Phe UUC	6	8	10	69	56
Leu UUA	0	1	1	2	6.1
Leu UUG	2	0	3	5	8
Leu CUU	4	2	3	10	9
Leu CUC	5	2	5	13	7
Leu CUA	1	0	2	3	2
Leu CUG	15	22	23	66	69
Ile AUU	4	4	13	34	37
Ile AUC	11	15	14	66	62
Ile AUA	0	0	0	0	1
Met AUG	6	5	12	—	—
Val GUU	1	0	5	8	38
Val GUC	10	7	13	39	13
Val GUA	2	0	1	4	23
Val GUG	18	8	11	49	27
Ser UCU	2	1	1	10	27
Ser UCC	4	5	2	28	26
Ser UCA	0	0	1	3	8
Ser UCG	4	3	5	30	11
Ser AGU	0	0	0	0	6
Ser AGC	5	3	4	30	22
Pro CCU	1	1	3	9	9
Pro CCC	0	2	2	8	6
Pro CCA	1	1	2	8	20
Pro CCG	8	16	16	75	65
Thr ACU	2	0	2	8	24
Thr ACC	8	8	21	73	51
Thr ACA	0	0	2	4	6
Thr ACG	1	2	5	16	20
Ala GCU	2	1	2	4	28
Ala GCC	13	24	21	47	19
Ala GCA	2	1	4	6	23
Ala GCG	12	18	24	44	30
Tyr UAU	3	6	7	62	41
Tyr UAC	0	1	9	38	59
Ter UAA	0	0	1	33	88
Ter UAG	0	0	0	0	4
Ter UGA	1	1	0	67	8
His CAU	1	2	9	48	39
His CAC	4	4	5	52	61
Gln CAA	1	1	3	10	27
Gln CAG	9	12	22	90	73
Asn AAU	2	3	5	26	24
Asn AAC	8	12	19	74	76
Lys AAA	8	5	9	71	77
Lys AAG	2	3	4	29	23
Asp GAU	4	7	14	46	51
Asp GAC	10	7	12	54	49
Glu GAA	6	8	11	45	73
Glu GAG	4	8	18	55	27
Cys UGU	0	0	1	9	42
Cys UGC	2	3	5	91	58
Trp UCG	3	0	8	—	—
Arg CGU	3	0	5	17	58
Arg CGC	7	12	18	77	35
Arg CGA	0	0	1	2	2
Arg CGG	0	0	1	2	3
Arg AGA	0	0	0	0	1
Arg AGG	1	0	0	2	0.2
Gly GGU	2	3	5	14	48
Gly GGC	14	12	8	49	41
Gly GGA	0	4	5	13	5
Gly GGG	4	2	11	24	7
Total	249	269	454	—	—

general trend is the preferred use of G and C in the third position in codon families that are 3- or 4-fold degenerate, and a pattern of rarely and frequently used codons consistent with the hypothesis of Grantham *et al.* (1981) relating 'differentially expressed' genes to the availability of isoaccepting tRNAs.

Neutral mutations in continuous culture of K. aerogenes?

The sequence difference of Asn-212 in the DNA sequence for Asp-212 in the amino acid sequence is possibly a protein sequencing error, but that of Ser-Ser at residues 146–147 for Ala-Val in the amino acid sequence cannot be so explained: the sequence evidence is strong on both sides (Dothie *et al.*, 1985). We are therefore forced to treat this as a genuine evolutionary change and ask where it might have arisen.

At first sight one would suspect the many manipulations of the *rbtD* gene that have occurred from the original wild-type *K. aerogenes* strain FG5, through the selection of the RDH-constitutive strain X1 to the prototrophic revertant strain A; then the P transduction into *E. coli* K12 and curing of phage to give *E. coli* strain EA (Rigby *et al.*, 1976). From strain EA the gene was transferred to phage λ to give λ pdal and λ prbt dal (Neuberger & Hartley, 1979) and from these into M13 phage for DNA sequencing. However the DNA sequences derived from plasmid pJCW1 are identical in this region, and this plasmid was recently constructed directly from a restriction digest of the wild type *K. aerogenes* DNA strain FG5. Hence the DNA sequence shown in Fig. 3 presumably represents that of the wild-type *rbt* gene.

Why then does the amino acid sequence of ribitol dehydrogenase-A differ from that encoded by the *rbtD* gene? Residues 146–147 lie within the tryptic peptide T13, and the sequence evidence was obtained from enzyme purified from *K. aerogenes* strain A211. Unfortunately, as discussed by Dothie *et al.* (1985), there is no firm evidence for this region in RDH-A prepared from other strains, but a Ser-Ser sequence at residues 146–147 might fit the composition of peptide T13 prepared from strain D (which produced RDH-D with increased xylitol activity arising from a change of Ala-196 to Pro-196). Moreover, a change of Gly-186 for Pro-186 was detected in a sample of enzyme from another strain A7 which superproduces RDH-A, and Morris *et al.* (1974) have offered evidence for Gln-60 in place of Glu-60 in enzyme from strain A11. All of these strains have been through several hundred generations of continuous culture on xylitol, which exerts selective pressure to maximize the xylitol dehydrogenase activity of this enzyme. In some cases (such as RDH-D) a dramatic rise in

specific activity arises from a single point mutation. In others (such as A11, A211 and A7) there is massive superproduction of an apparently unchanged RDH-A, up to 30% of total protein. It would not be surprising if mutations were selected in the coding sequence to allow this massive protein synthesis.

References

- Altosaar, I. & Hartley, B. S. (1976) *Abstr. Int. Congr. Biochem. 10th*, p. 200, abstr. no. 04-6-319
- Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3963-3965
- Bolivar, F., Rodriguez, R. L., Betlach, M. C. & Boyer, H. W. (1977) *Gene* **2**, 75-94
- Charnetzky, W. T. & Mortlock, R. P. (1974a) *J. Bacteriol.* **119**, 162-169
- Charnetzky, W. T. & Mortlock, R. P. (1974b) *J. Bacteriol.* **119**, 170-175
- Deininger, P. L. (1983) *Anal. Biochem.* **129**, 216-223
- Dothie, J. M., Giglio, J. R., Moore, C. H., Taylor, S. S. & Hartley, B. S. (1985) *Biochem. J.* **230**, 569-578
- Dretzen, G., Bellard, M., Sassone-Corsi, P. & Chambon, P. (1981) *Anal. Biochem.* **112**, 295-298
- Galibert, F., Sedat, J. & Ziff, E. (1974) *J. Mol. Biol.* **87**, 377-407
- Garoff, H. & Ansorge, W. (1981) *Anal. Biochem.* **115**, 450-457
- Gantham, R., Gautier, C., Gony, M., Jacobzone, M. & Mercier, R. (1981) *Nucleic Acids Res.* **9**, r43-r74
- Hartley, B. S., Altosaar, I., Dothie, J. M. & Neuberger, M. S. (1976) in *Structure-Function Relationships in Proteins* (Markham, R. & Horne, R. W., eds.), pp. 191-200, North-Holland, Amsterdam
- Knott, T. J. (1982) Ph.D. Thesis, University of London
- Konigsberg, W. & Godson, G. N. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 687-691
- Loviny, T., Neuberger, M. S. & Hartley, B. S. (1981) *Biochem. J.* **193**, 631-637
- Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560-564
- Messing, J. & Vieira, J. (1982) *Gene* **19**, 269-276
- Messing, J., Crea, R. & Seeburg, P. H. (1981) *Nucleic Acids Res.* **9**, 309-321
- Morris, H. R., Williams, D. H., Midwinter, J. J. & Hartley, B. S. (1974) *Biochem. J.* **141**, 701-713
- Neuberger, M. S. & Hartley, B. S. (1979) *J. Mol. Biol.* **132**, 435-470
- Neuberger, M. S., Patterson, R. A. & Hartley, B. S. (1979) *Biochem. J.* **183**, 31-42
- Neuberger, M. S., Hartley, B. S. & Walker, J. E. (1981) *Biochem. J.* **193**, 513-524
- Nichols, B. P., Blumenberg, M. & Yanofsky, C. (1981) *Nucleic Acids Res.* **9**, 1743-1755
- Rigby, P. W. J., Burleigh, B. D. & Hartley, B. S. (1974) *Nature (London)* **251**, 200-204
- Rigby, P. W. J., Gething, M. J. & Hartley, B. S. (1976) *J. Bacteriol.* **125**, 728-738
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463-5467
- Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161-178
- Scangos, G. A. & Reiner, A. M. (1978) *J. Bacteriol.* **134**, 491-500
- Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1342-1346
- Sober, H. A. (1979) in *Handbook of Chemistry: Selected Data for Molecular Biology*, 2nd edn., pp. H83-H85, CRC Press, West Palm Beach
- Staden, R. (1980) *Nucleic Acids Res.* **8**, 3673-3694
- Taylor, S. S., Rigby, P. W. J. & Hartley, B. S. (1974) *Biochem. J.* **141**, 693-700
- Winter, G. & Fields, S. (1980) *Nucleic Acids Res.* **8**, 1965-1974
- Winter, G., Fields, S., Gait, H. J. & Brownlee, G. C. (1981) *Nucleic Acids Res.* **9**, 237-245
- Wu, J. C. (1983) M.Phil. Thesis, University of London
- Wu, T. T., Lin, E. C. C. & Tanaka, S. (1968) *J. Bacteriol.* **96**, 447-456