Analysis of insertion mutants reveals two new genes in the pNRC100 gas vesicle gene cluster of *Halobacterium halobium*

Jeffrey G.Jones, Neil R.Hackett¹, John T.Halladay, Douglas J.Scothorn, Chin-fen Yang, Wai-lap Ng and Shiladitya DasSarma

Department of Microbiology, University of Massachusetts, Amherst, MA 01003 and ¹Department of Molecular Biology, Vanderbilt University, Nashville, TN 37235, USA

Received June 8, 1989; Revised and Accepted August 25, 1989

EMBL accession no. X15374

ABSTRACT

The archaeobacterium, *Halobacterium halobium*, achieves buoyancy through synthesis of intracellular gas-filled vesicles. The plasmid-encoded gene (*gvpA*) specifying the major structural gas vesicle protein has previously been cloned and sequenced allowing the analysis of high-frequency mutations to the vesicle negative phenotype. Among eighteen gas vesicle mutants analyzed, four were observed to contain insertion elements 0.2 to 2 kb upstream of the structural gene. To explain the phenotype of these mutants, the upstream area was analyzed by DNA sequencing and transcriptional mapping. This analysis showed the presence of two open reading frames, *gvpD* and *gvpE*, which are of opposite transcriptional orientation to *gvpA* (gene order *gvpA-D-E*). *gvpD* begins 201 nucleotides from the *gvpA* structural gene and is 1608 nucleotides long while *gvpE* begins two nucleotides from the 3'-end of *gvpD* and is 573 nucleotides long. Primer extension analysis showed the occurrence of divergent promoters in the *gvpA-gvpD* intergenic region with the transcription start sites separated by 109 nucleotides. The sites of three insertion sequences in gas vesicle mutants mapped within *gvpE* while the fourth insertion site mapped near the N-terminal coding region of *gvpD*. Homology between the *gvpDE* gene region and a chromosomal site in a *H. halobium* NRC-1 derivative and in several other *Halobacterium* strains was identified by Southern hybridization.

INTRODUCTION

Extremely halophilic archaeobacteria such as *Halobacterium halobium* flourish in hypersaline brine containing 3–5 M NaCl (1). Like other aquatic bacteria, many *Halobacterium* strains synthesize gas-filled vesicles, which give buoyancy and, thus, increase the availability of light and oxygen to cells (2–4). The gas vesicle membrane contains one or two major proteins; however, a minor protein component has recently been reported on the external surface of cyanobacterial vesicles (4–7). *H. halobium* gas vesicles are largely spindle-shaped, consisting of a central cylindrical region with conical ends (8). Synthesis is thought to begin at the conical ends and proceed by addition of subunits in the central region (9) with inflation occurring during synthesis by passive diffusion of dissolved gases across the membrane but exclusion of water (2).

Our interest in gas vesicles resulted from the extreme genetic instability of this phenotype apparent in the wild-type *H. halobium* NRC-1 strain (reviewed in ref. 10). Inflated gas vesicles diffract light, giving an opaque appearance to wild-type *H. halobium* colonies on agar plates. Gas vesicle mutants of *H. halobium* arise spontaneously at frequencies of about 1% and are readily apparent as translucent colonies (11–13). After cloning of the *gvpA* gene encoding the major structural protein in gas vesicles from *H. halobium* we mapped the wild-type *gvpA* gene to a large *H. halobium* plasmid, pNRC100 (14). A second gene (*gvpB*), homologous to *gvpA*, was subsequently cloned from the *Halobacterium*

chromosome (15). We characterized 18 independent gas vesicle-deficient mutants and 4 revertants and found three classes of gas vesicle mutants, all of which had mutations within or near the plasmid-encoded *gvpA* gene (16). Of these, class II mutants contained insertion sequences (ISH) up to 2.2 kb (kilobase pairs) upstream of *gvpA* (Fig. 1A) and had reduced (1.0 to 8.2%) gas vesicle levels compared to wild-type. One class II mutant, strain R1, contained a 1.3 kb element, ISH3, which is similar to the *Halobacterium volcanii* ISH51 element, 14 bp (base pairs) upstream of the *gvpA* transcription start site (10,16). The other two classes of mutants contained either reduced (class I) or no (class III) *gvpA* gene copies (16).

In this paper, we report on the nucleotide sequence of the 2.3 kb region upstream of *gvpA* and the sites of insertion of ISH elements in four class II mutants. Our results show the presence of two open reading frames which are in the opposite transcriptional orientation to *gvpA*. We discuss the structure of these *gvp* genes and their possible function in gas vesicle synthesis. The occurrence of homology between the *gvpDE* gene region and the chromosome of *H. halobium* NRC-1 is also described.

MATERIALS AND METHODS

Strains and culturing

Halobacterium strains and growth conditions used have been described previously (14,16). The class II gas vesicle mutants R1, SD104, SD106, SD108, and SD120 and class III mutant SD118A, are derivatives of wild-type *H. halobium* strain NRC-1 (Fig. 1A, refs. 10,16). *Halobacterium* strain SB3 was provided by W. Goebel (17).

Cloning and DNA sequencing analysis

Gas vesicle genes were cloned from *SalI* libraries of plasmid DNA from four class II mutants of *H. halobium* (SD104, SD106, SD108, and SD120) (18,19). The libraries were constructed in either pKS⁺ (Stratagene) or pTZ18/19 (Pharmacia), and screened using a 0.9 kb *HindIII*-*EcoRI* *gvpA* gene-containing fragment previously cloned from the wild-type NRC-1 strain (16) which was ³²P-labelled by nick translation (20). Restriction fragments containing the ISH element-target site junctions were further subcloned using *BglII*, *DraI*, *EcoRI*, *SstI*, and *HindIII*, into M13mp18 and mp19 for sequence analysis.

For cloning the wild-type *gvpDE* gene region, a *SalI*-*HindIII* fragment from upstream of *gvpA* from mutant strain SD108 was used to screen an *EcoRI*-*HindIII* library of plasmid DNA from *H. halobium* NRC-1 constructed in pTZ19 (18). Several clones containing the 5.6 kb *EcoRI*-*HindIII* fragment of interest were obtained of which one, pNG1, was selected for sequence analysis. The previously described clone, pGVH1, which contains a 2.2 kb *Sau3A* partial fragment encoding the N-terminal half of *gvpD* and all of *gvpA* was cloned at the *Bam*HI site of pUC12 (21) and was also used for sequence analysis. Subclones of an independent clone, pNGHC1, containing the 19 kb *HindIII* fragment upstream of *gvpA* were used to verify the sequence of the *gvpD*-*gvpE* border region. Fragments of pNG1, pGVH1, and pNGHC1 were subcloned into M13mp18/mp19 (22) or pTZ18/pTZ19 using *Bam*HI, *Bgl*II, *Cl*aI, *Dra*I, *Hae*III, *Hind*III, *Hpa*II, *Sma*I, *Sst*I, *Stu*I, *Taq*I, and *Xho*I for sequence analysis.

DNA sequence analysis was carried out using the Sanger chain termination method (23) on single stranded template DNA. The universal primer or reverse primer were used to prime sequencing reactions with DNA polymerase I large fragment (Bethesda Research Laboratories) or Sequenase (U.S. Biochemicals). For two reactions, primers which hybridized within *gvpD*, 5'-GGCAACTCGCGTACGA-3' and 5'-GACTC-

Figure 1. (A) Sites of insertion near the major gas vesicle protein gene in five class II gas vesicle mutants of *H. halobium* strain NRC-1. The restriction map of a 13 kb region around *gvpA* (hatched region) is shown. The insertion sites are indicated by dashed lines and the insertion sizes are indicated by the shaded boxes above and below the restriction map. The strain designation for each mutant, R1, SD104, SD106, SD108, and SD120 is indicated (10,16). (B) Genetic map of the gas vesicle protein gene region and sequencing strategy. The dashed lines indicate the region from part A (above) that is enlarged in B. The sequencing strategy and restriction map are shown below the genetic map. The positions of the open reading frames, *gvpA*, *gvpD*, and *gvpE*, are shown by hatched regions. The approximate positions and direction of transcription of the divergent promoters (P_A and P_D) are shown above the genetic map.

AAAGTCCTCAT-3' (positions 1333 to 1348 and 1303 to 1318, Fig. 2) were used for sequencing. Most reaction products were analyzed on 6% polyacrylamide-8.3 M urea gels. The sequence was determined for both strands over the entire region shown in Figure 2 by the strategy summarized in Figure 1B.

Primer extension analysis

Primer extension analysis was carried out using a 15-nucleotide long synthetic oligodeoxynucleotide (5'-GGTGAGCTAGATTGG-3') hybridizing to codons 4 to 9 of *gvpD* (see Fig. 2). The oligodeoxynucleotide was first labelled at the 5'-end by T4 polynucleotide kinase and [γ - 32 P]ATP (24) and then extended using AMV reverse transcriptase on *H. halobium* NRC-1 RNA template isolated from late-log phase cultures as previously described (14). Products were analyzed on a 6% polyacrylamide-8.3 M urea gel using as size standards a Sanger sequence ladder generated using the same phosphorylated primer.

Southern blotting analysis

For Southern blotting analysis (25), *H. halobium* genomic DNA (5.0 μ g per lane) was digested with *Bam*HI, *Bgl*III, *Eco*RI or *Pst*I, electrophoresed on a 1% agarose gel, denatured and transferred to nylon membranes and hybridized using the 1.1 kb *Hind*III-*Bgl*III fragment

Figure 2. Nucleotide sequence of the *gvpD* and *gvpE* genes and flanking DNA. The sequence is numbered from the transcription initiation site for *gvpD* and the orientation is opposite to that in Figure 1. The positions of transcription starts for *gvpD* and *gvpA* (14) are shown by prominent arrows. The sites of insertion of ISH elements in class II mutants R1, SD104, SD106, SD108, and SD120 are marked by underlining and overlining. The start codons for *gvpD* and *gvpE*, and the *gvpA* start codon complement are marked and underlined. The stop codons for *gvpD* and *gvpE*, are underlined. Sequences complementary to the 3' end of the small rRNA

containing *gvpD* and a small part of *gvpE* cloned in M13mp19, which had been ^{32}P -labelled by the method of Hu and Messing (26).

RESULTS

Nucleotide sequence of the gvpD and gvpE genes

Previously published Southern blotting analysis of 18 gas vesicle deficient mutants of *H. halobium* NRC-1 indicated that four mutants, SD104, SD106, SD108, and SD120, contain insertions from 0.2 to 2.2 kb upstream of the major gas vesicle protein gene (16). This prompted us to determine the complete nucleotide sequence of this region in the wild-type (Fig. 2). The sequence revealed the presence of two long open reading frames (ORFs) upstream of *gvpA* and in the opposite strand which were designated *gvpD* and *gvpE*. The codon usage for *gvpD* and *gvpE* is very similar to the previously sequenced *H. halobium* genes with a marked preference for G or C in the third position. The first ORF, *gvpD*, starts 201 bp from the *gvpA* structural gene and is 1608 bp long while the second ORF, *gvpE* starts two nucleotides beyond *gvpD* and is 573 bp long.

Analysis of nucleotide sequences across the target site-ISH element junctions in the gas vesicle-deficient mutants, SD104, SD106, SD108, and SD120, indicated that each insertion had interrupted the *gvpDE* gene region. The site of insertion in one mutant, SD106, is near the beginning of the *gvpD* gene while the other three mutants contain insertions at sites within the *gvpE* gene (Fig. 2).

Transcription initiates 71 nucleotides upstream of gvpD

The sequence suggests that the *gvpA-gvpD* intergenic region should contain 2 divergent promoters. Transcription of *gvpA* has previously been shown to originate 20 nucleotides before the start codon (position -110 in Figure 2). Primer extension analysis using an oligodeoxynucleotide hybridizing within codons 4-9 of *gvpD* (Fig. 3) showed a divergent transcript originating 71 nucleotides before the *gvpD* start codon. This result was confirmed by S1 nuclease mapping (data not shown) (27).

Homology of gvpDE to the chromosome of H. halobium (NRC-1) and other Halobacterium species

H. halobium contains a *gvpA* homolog, *gvpB*, at a site on the chromosome (15). To determine whether the *gvpDE* gene region is also present on the chromosome, we carried out Southern blotting analysis of several strains lacking the pNRC100 gas vesicle genes using a probe specific for *gvpD* and a short region of *gvpE*. The strains analyzed include SD118A, a class III mutant with a deletion of the entire gas vesicle locus from pRNC100 (W.-L.N. and S.D., unpublished results) and *Halobacterium* species SB3 which lacks a pNRC100-like plasmid. The data shown in Figure 4 clearly indicates the presence of a homologous region in both SD118A and SB3. The hybridizing fragments were of identical size for both strains using each of four restriction enzymes (*Bam*HI, *Bgl*II, *Eco*RI, and *Pst*I).

DISCUSSION

In this paper, we have reported on the region upstream of the major gas vesicle protein structural gene. We have shown by nucleotide sequencing and transcription mapping

upstream of *gvpD* and *gvpE*, are shown bolded. The predicted amino acid sequences for *gvpD* and *gvpE*, are indicated below the nucleotide sequence. The nucleotide binding site consensus in the *gvpD* amino acid sequence is underlined. The recognition sites for *Bam*HI, *Hind*III, *Sma*I, *Bgl*II, *Clal*, and *Dra*I sites are marked and underlined. A *Sau*3A site in the *gvpD* gene which was used for cloning pGVH1 (14) is also marked (see text). An inverted repeat sequence in the promoter region is indicated by half-arrows.

Figure 3. Primer extension analysis of *gvpD* mRNA. A 15-nucleotide long synthetic oligodeoxynucleotide was used to prime cDNA synthesis on yeast RNA (lane 1) and *H. halobium* NRC-1 RNA (lane 2) using AMV reverse transcriptase. A major 96-nucleotide product, specific to *H. halobium* RNA, is observed in lane 2. The oligodeoxynucleotide was also used to generate Sanger sequencing ladders on single-stranded DNA template (lanes G, A, T, and C) for determination of the precise location of the transcription start. The promoter region sequence is indicated on the side with a prominent arrow indicating the transcription start site and half-arrows indicating an inverted repeat sequence.

experiments the presence of two ORFs, *gvpD* and *gvpE*, which are in opposite transcriptional orientation from *gvpA*. We also described the sites of ISH element insertions (three in *gvpE* and one in *gvpD*) which reduce the synthesis of gas vesicles.

Our results suggest the possibility that the *gvpD-gvpE* gene region is organized as an operon. Transcription starts 71 nucleotides before the translational start of *gvpD* (Fig. 3). The direction of transcription is opposite to *gvpA* with 109 bp separating the two divergent transcription starts (Fig. 2). The *gvpA-gvpD* intergenic region contains an inverted repeat sequence, 5'-CGTACTT-3', centered 28–29 bp upstream of the *gvpD* transcription start site (marked by half-arrows in Fig. 2). The spacing between repeats and their location suggest a regulatory role for this inverted repeat sequence. This region is A + T-rich (56%) and shows similarity to proposed halobacterial promoter sequences (28,29).

Interestingly, the organization of the *gvp* gene cluster is extremely compact, with the stop codon for *gvpD* and the start codon for *gvpE* in close proximity. Properly positioned Shine–Dalgarno (30) sequences are present for each of these genes, including *gvpD* and *gvpE* (bold letters in Fig. 2). Previously, several other *H. halobium* genes which are translated from the first AUG in the message were found to lack a properly positioned Shine–Dalgarno sequence (14,28,31–35). More recently, studies on the genetic

Figure 4. Southern hybridization of *Halobacterium* DNA using the *gvpDE* gene-specific probe. Total DNA of *H. halobium* SD118A (a,d), or *Halobacterium* SB3 (e,h) was digested with *Bam*HI (a,e), *Bgl*III (b,f), *Eco*RI (c,g) or *Pst*I (d,h). The sizes of markers in kb are shown to the left.

organization of a ribosomal protein gene cluster and the RNA polymerase operon in *H. halobium* showed the occurrence of Shine–Dalgarno sequences and very short intergenic regions (36,37). These findings suggest the possibility that *H. halobium* ribosomes may require the presence of a properly positioned binding site for recognition of translational start codons located internally but not for recognition of start codons located very close to the 5' end of messages. Thus, translation initiation in *H. halobium* may have similarity both to mammalian mitochondrial systems (38) in choosing 5' proximal start codons and also to prokaryotic systems (30) in selecting internal starts (39). However, two highly expressed genes, *gvpA* and *glyC*, contain substantial untranslated leaders but lack Shine–Dalgarno sequences (14,34). Analysis of other genes and operons is necessary to determine the general features of translation signals in halobacteria.

A region of the *H. halobium* NRC-1 chromosome is homologous to the *gvp* gene cluster of pNRC100. The chromosomal homolog of the *gvpA* gene, *gvpB*, was previously reported (15). We find that the *gvpDE*-specific probe hybridizes to a fragment of the *H. halobium* NRC-1 chromosome as well as identical size fragments in the genome of *Halobacterium* species GRA, GRB, SB3, and GN101 (Fig. 4 and our unpublished results). Preliminary results show that the chromosomal homolog of the *gvpDE* region is present on the same restriction fragment as *gvpB* (data not shown). The size of the *Pst*I fragment hybridizing to the *gvpD* probe in Figure 4 (8.7 kb) agrees with the size of the *Pst*I fragment containing *gvpB* (15). These results suggest that the gas vesicle genes may be mobilizable between the plasmid and chromosome, and perhaps between various bacterial strains and species. In addition, the partially gas vesicle-deficient phenotype of the class II mutants studied might be explained if the chromosomal genes are expressed and able to partially complement the plasmid *gvpD* and *gvpE* insertion mutations.

Translation of the *gvpD* and *gvpE* gene sequences indicate gene products of 59000 and 21000 predicted molecular weights. Analysis of the sequence for *gvpD* indicates that a

canonical nucleotide binding site (40) (underlined in Fig. 2) is present near the N-terminus. The sequence, L-I-N-G-A-P-G-T-G-K-T, is similar to sequences in the yeast adenylate kinase (L-I-G-P-P-G-A-G-K-G) (41) and thymidylate kinase (L-I-E-G-L-D-R-T-G-K-T) (42), and the *E. coli rho* transcription termination factor (L-I-V-A-P-P-K-A-G-K-T) (43). Another region with the sequence I-I-L-Y is similar to a second conserved region of several nucleotide binding proteins (40). The presence of a nucleotide binding site in *gvpD* in addition to the fact that an insertion into *gvpD* dramatically lowers the gas vesicle content suggests the possibility that the *gvpD* gene product plays a role in an energy requiring process such as assembly of gas vesicles. Other possible functions of the *gvpD* and also *gvpE* gene products include structural or regulatory roles.

The gas vesicle gene cluster on the *H. halobium* plasmid pNRC100 has grown from a single gene to at least four genes, *gvpA*, *C*, *D*, and *E* (7,10,14,15,unpublished data). In addition, homologs of *gvpA* and probably *gvpD* and *gvpE* are present on the chromosome. The possibility of involvement of other uncharacterized genes has not been ruled out. The results now indicate that gas vesicle synthesis is more complex than originally expected and may be similar to phage head morphogenesis in the number of genes involved. Further experimentation is necessary to determine the function of these new genes in gas vesicle synthesis.

ACKNOWLEDGEMENTS

This work was supported by NSF grant DBM-8703486 and NIH grant R01 GM41980-01 to (S.D.) and NIH grant R29 GM39887-01 to (N.R.H.).

REFERENCES

1. Kushner,D.J. (1985) in The Bacteria, vol. VIII. The archaeobacteria, eds. Woese,C.R. and Wolfe,R.S. (Academic Press, Orlando, FL), pp.171–214.
2. Walsby,A.E. (1978) Symp. Soc. Gen. Microbiol., 28, 327–358.
3. Stoekenius,W. and Kunau,W. (1968) J. Cell Biol., 38, 337–357.
4. Krantz,J.J. and Ballou,C.E. (1973) J. Bacteriol., 114, 1058–1067.
5. Simon,R.D. (1981) J. Gen. Microbiol., 125, 103–111.
6. Walsby,A.E. and Hayes,P.K. (1988) J. Gen. Microbiol., 134, 2647–2657.
7. Damerval,T., Castets,A.-M., Guglielmi,G., Houmard,J. and Tandeau de Marsac,N. (1989) J. Bacteriol., 171, 1445–1452.
8. Cohen-Bazire,G., Kunisawa,R. and Pfennig,N. (1969) J. Bacteriol., 100, 1049–1061.
9. Waaland,J.R. and Branton,D. (1969) Science, 163, 1339–1341.
10. DasSarma,S. (1989) Can. J. Microbiol., 35, 65–72.
11. Larsen,H., Omang,S. and Steensland,H. (1967) Archiv. fur Mikrobiol., 59, 197–203.
12. Weidinger,G., Klotz,G. and Goebel,W. (1979) Plasmid, 2, 377–386.
13. Pfeifer,F., Weidinger,G. and Goebel,W. (1981) J. Bacteriol., 145, 375–381.
14. DasSarma,S., Damerval,T., Jones,J.G. and Tandeau de Marsac,N. (1987) Mol. Microbiol., 1, 365–370.
15. Horne,M., Englert,C., and Pfeifer,F. (1988) Mol. Gen. Genet., 213, 459–464.
16. DasSarma,S., Halladay,J.T., Jones,J.G., Donovan,J.W., Giannasca,P.J. and Tandeau de Marsac,N. (1988) Proc. Natl. Acad. Sci. USA, 85, 6861–6865.
17. Ebert,K., Goebel,W. and Pfeifer,F. (1984) Mol. Gen. Genet., 194, 91–97.
18. Maniatis,T., Fritsch,E.R. and Sambrook,J. (1982) Molecular Cloning (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
19. Bergquist,P.L. (1987) in Plasmids: a practical approach, ed. Hardy,K.G. (IRL Press, Oxford), pp. 37–78.
20. Rigby,P.W.J., Dieckmann,M., Rhodes,C., and Berg,P. (1977) J. Mol. Biol., 113,237–251.
21. Vieira,J. and Messing,J. (1982) Gene, 19,259–268.
22. Messing,J. (1982) Meth. Enzymol., 101, 20–78.
23. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA, 74, 5463–5467.
24. Maxam,A.M. and Gilbert,W. (1980) Methods Enzymol., 65:499–560.

25. Southern, E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.
26. Hu, N.-T. and Messing, J. (1982) *Gene*, **17**, 271–277.
27. Berk, A.J. and Sharp, P.A. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 1274–1278.
28. DasSarma, S., RajBhandary, U.L. and Khorana, H.G. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 125–129.
29. Thomm, M. and Wich, G. (1988) *Nucleic Acids Res.*, **16**, 151–163.
30. Shine, J. and Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 1342–1346.
31. Simsek, M., DasSarma, S., RajBhandary, U.L. and Khorana, H.G. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 7268–7272.
32. Betlach, M., Friedman, J., Boyer, H.W. and Pfeifer, F. (1984) *Nucleic Acid Res.*, **12**, 7949–7959.
33. Hegemann, P., Blanck, A., Vogelsang-Wenke, H., Lottspeich, F. and Oesterhelt, D. (1987) *EMBO J.*, **6**, 259–264.
34. Lechner, J. and Sumper, M. (1987) *J. Biol. Chem.*, **262**, 9724–9729.
35. Leong, D., Pfeifer, F., Boyer, H. and Betlach, M. (1988) *J. Bacteriol.*, **170**, 4903–4909.
36. Leffers, H., Gropp, F., Lottspeich, F., Zillig, W. and Garrett, R.A. (1989) *J. Mol. Biol.*, **206**, 1–17.
37. Shimmin, L.C. and Dennis, P.P. (1989) *EMBO J.*, **8**, 1225–1235.
38. Kozak, M. (1983) *Microbiol. Rev.*, **47**, 1–45.
39. Montoya, J., Ojala, D. and Attardi, G. (1981) *Nature*, **290**, 465–470.
40. Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) *EMBO J.*, **1**, 945–951.
41. Tomasselli, A.G., Mast, E., Janes, W. and Schlitz, E. (1986) *Eur. J. Biochem.*, **155**, 111–119.
42. Jong, A.Y.S., Kuo, C.L. and Campbell, J.L. (1984) *J. Biol. Chem.*, **259**, 11052–11059.
43. Pinkham, J.L. and Platt, T. (1983) *Nucleic Acids Res.*, **11**, 3531–3545.