CHAPTER 4

Northern blot analysis of the *vap* genes in *Dichelobacter nodosus*

The great tragedy of Science - the slaying of a beautiful hypothesis by an ugly fact.

Thomas Henry Huxley (1825-1895).

4.1 Introduction

The *vap* regions are found in almost all virulent strains (>98%) of *D. nodosus* but are absent from most benign strains (72%), and hence may have a role in virulence(Katz *et al.*, 1991; Katz, Strugnell & Rood, 1992; Katz *et al.*, 1994). A number of genes have been identified within the *vap* regions, including an integrase gene, *intA* , *vapA*, *vapB*, *vapC*, *vapD*, *vapE*, *vapF*, *vapG*, *vapH*, *toxA*, and a number of related genes including *vapE'*, *vapA'*, *vapA''*, *toxA''* and *vapG'* (Figure 4.1).

There is at present no transformation system for *D. nodosus* which precludes direct tests for the role of these and other genes in virulence, hence past investigations have concentrated on analysis at the DNA level. Southern blot experiments indicated that *(i)* multiple copies of the *vap* genes are not required for virulence since there are virulent strains of *D. nodosus* that have only single copies of the *vap* element (Table 4.1); *(ii)* that the arrangement of the *vap* genes is similar in benign and virulent strains of *D. nodosus* in which their arrangement has been analysed (Bloomfield *et al.*, 1997); and *(iii)* that the presence or absence of certain *vap* genes alone, did not determine whether a given strain exhibited a benign or virulent phenotype, since benign strain H1204 contains the full complement of *vap* genes (Table 4.1) (Bloomfield *et al.*, 1997). Previous work showed

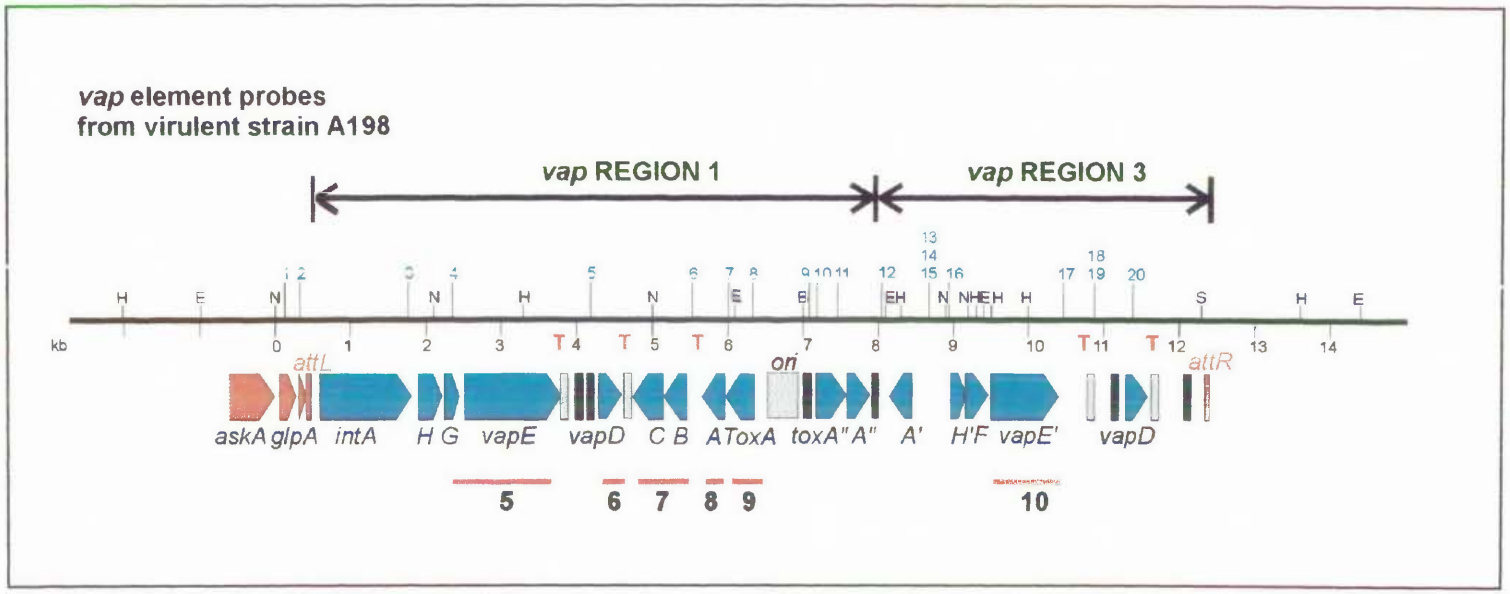

Figure 4.1: Map showing probes 1 to 6 from *vap* regions 1 and 3 from *D. nodosus* virulent strain A198, which were used in Northern blot experiments. Probe fragments are indicated by red lines below the corresponding sequences and are numbered as follows: *vapE* (1), *vapD* (2), *vapBC* (3), *vapA* (4), *toxA* (5) and *vapE'* (6). The numbers shown in black type indicate the distance in kb from the leftmost *Nru*I site. Restriction sites shown are as follows: *Bam*HI (B), *Eco*RI (E), *Hin*dIII (H), *Nru*I (N) and *Sac*I (S). The location of putative open reading frames are indicated by filled arrows. Repeated sequences shown include: putative attachment sites, *attL* or *attR* (narrow red boxes), 102 bp repeats or partial copies thereof (filled boxes), 103 bp repeats or partial copies of (grey boxes), and a putative origin of replication, *ori* (grey box). Putative promoter sequences, numbered 1 to 20 (blue type), correspond to those listed in Table 4.2. The approximate positions of putative terminators for those genes analysed in this work are indicated by T in red type, and the exact co-ordinates are given in Table 4.3.

that *vapD* is expressed at the protein level in all strains of *D. nodosus* that contain *vapD* (Katz *et al.*, 1992).

Table 4.1: Distribution of the *vap* genes in 8 strains of *D. nodosus*

	Strain of *D. nodosus*							
Gene	A198 *(v)*	C305 *(b)*	AC3577 *(i)**	B1006 *(v)*	G1220 *(v)*	H1204 *(b)*	H1215 *(v)*	1311 *(v)*
vapA	++	−	+	+	+	+	+	+
toxA	++	−	+	+	+	+	+	+
vapB	++	−	+	+	+	+	+	+
vapC	++	−	+	+	+	+	+	+
vapD	+++	−	+	++	+++	+	+	+
vapE	++	−	+	+	+	+	-	+
vapE'	+	−	+	−	−	−	+	−

Key to Table 1: present: 1 copy (+), 2 copies (++) or 3 copies (+++); absent (−); The classification of each strain is indicated as follows: benign *(b)*, intermediate *(i)* and virulent *(v)* . * strain AC3577 contains the *vap* plasmid (Billington *et al.*, 1996), and has at least one copy of all of the *vap* genes except *vapE'*. However, strain AC3577 may have more that one copy of the *vap* genes per cell since the plasmid may be in multiple copies and other work (Section 7.2.8) suggests that strain AC3577 also contains an integrated copy of the *vap* element.

There are several hypotheses to explain why benign strains of *D. nodosus* that contain the full complement of *vap* genes still have a benign phenotype: *(i)* the *vap* genes or key *vap* genes may not be expressed in the benign strains in which they are present; *(ii)* the *vap* genes may be differentially expressed in virulent and benign strains; *(iii)* the protein products of the *vap* genes may be non-functional in benign strains; *(iv)* where a benign strain of *D. nodosus* has the full complement of *vap* genes other virulence associated genes are non-functional; *(v)* the *vap* genes may be required for the acquisition of virulence genes; *(vi)* the site of integration of the *vap* regions may determine the virulence of the strain; or *(vii)* the *vap* genes have no role in virulence.

In order to eliminate some of the aforementioned possibilities, northern blot experiments were undertaken to determine whether the *vap* genes are expressed at the RNA level in the virulent and/or benign strains in which they are present. This is an important question because it may indicate whether the *vap* genes have a role in virulence.

These studies could also confirm that the *vap* genes are transcribed as predicted from analysis of the DNA sequences (Cheetham *et al.*, 1995b; Katz *et al.*, 1992; Katz *et al.*, 1994). The results suggest that the *vap* genes (in general) are expressed in the virulent and benign strains in which they are present, and although differential expression of the *vap* genes was observed, the differences were not related to the virulence of the isolates. Analysis at the RNA level also confirmed that *vapB* and *vapC* are expressed as an operon, as are *toxA* and *vapA*. In addition, *vapE* was found to be transcribed from two transcriptional start sites, and one of these transcripts includes *vapH* and *vapG*. The arrangement and DNA sequence of *vapHG* is similar to the immunity region from bacteriophage P4.

4.2 Results

4.2.1 Identification of putative promoter sequences from *vap* regions 1 and 3 from *D. nodosus* strain A198

Prior to expression studies the previously-determined sequence of *vap* regions 1 and 3 from *D. nodosus* strain A198 was analysed for potential −35 and −10 promoter sequences matching the consensus promoter recognised by the *E. coli* RNA polymerase σ^{70} subunit, as described previously (Hawley & McClure, 1983) (Table 4.2).

Table 4.2: Promoter sequences in *vap* regions 1 and 3 from *D. nodosus* strain A198

P#	Promoter Position	Promoter sequence -35 -10 t(c/g)TTGACa tg*TAtAaTg	Gene
P1	76-106R	TTGACGgaggtttttttttcggTATAAT	*tRNA-ser*
P2	281-309R	TTCTATaaaaacatagtattaagTATATT	*intA*
P3	1731-1763R	TTGACAagccaagagcttttagcTATAGT	*vapG, vapE, orf54*
P4	2139-2170R	TTGCCTatcatttaacgcaagcgTTAAAT	*vapE*
P5	4176-4204	TTGAAActcaattattgataacaGATAAT	*vapD* (Reg.1)
P6	5373-5345L	TTCGAAtaaaaaactaccttgcTAAAAT	*vapBC*
P7	6001-5973L	TTGACTttaaagatggcaatgccTACATT	*vapA*
P8	6281-6253L	TTGAACcccgtcacgttacgcgaTACAAT	*toxA*
P9	7060-7088R	TTGAAAccctattgcggaaaaatCATTAT	*toxA″*
P10	7152-7180R	TTCTCGtttatcgataaacgataTAAAAT	*toxA″*
P11	7432-7460R	TTCCGTtgggatagtggcgcgcaTGATGT	*vapA″*
P12	8011-7984L	TTGCAAcgttcataaatggcatTAATAT	unknown
P13	8551-8579R	TTGACTtttcggcgctaattcatTAATAA	*vapG′, orf77*
P14	8537-8508L	TTGACTgtagccaaatggctacctGATTAT	*vapA′*
P15	8600-8572L	TTGCGCgctgcgataggtaaattATTAAT	unknown
P16	8985-9013R	TTGCCTatcatttaacgcaagcaTGCTTT	*vapF, vapE′*
P17	10447-10419L	TTCTTTatccagccaaggcaaaaTAATAT	unknown
P18	10843-10871R	TTGCATatgttgcacttactaaaTATTAT	unknown
P19	10896-10866L	TTCTTTatttatcaatgtattaatTATAAT	unknown
P20	11356-11384R	TTGAAActcaattattgataacaGATAAT	*vapD* (Reg.3)

Note: the direction of the promoter on a map of *vap* regions 1 and 3 from strain A198 is indicated as follows: rightward (R), and leftward (L). Reg. 1 - *vap* region 1. Reg. 3 - *vap* region 3. The sequence of *vap* regions 1 and 3 of strain A198 (GenBank accession number L31763).

Table 4.3: Predicted and experimentally determined transcript sizes for *vap* genes of *D. nodosus* strain A198

Gene	Co-ordinates 5'-3- (nt)	Putative promoter	Position of putative promoter	Position and sequence of putative terminator	Predicted transcript size (kb)	Experimental transcript size (kb)
vapA	5936-5628	7 8	6001-5973L 6281-6253L	(5582) AATGCCGCGtcatCGCGGTATT (5568)	0.41 0.69	none 0.76
toxA	6227-5949	8	6281-6253L	(5582) AATGCCGCGtcatCGCGGTATT (5568)	0.69	0.68
vapB	5292-5062	6	5373-5345L	(4615) AAGCAGCCtatttGGCTGCTT (4595)	0.75	0.70
vapC	5065-4658	6	5373-5345L	(4615) AAGCAGCCtatttGGCTGCTT (4595)	0.75	0.70
vapD	4255-4536 11435-11716	5 20	4176-4204R 11356-11384R	(4592) ATGAAGCAGCCaaataGGCTGCTTTTT (4617) (11772) ATGAAGCAGCCaaataGGCTGCTTTTT (11798)	0.41 0.41	0.44 0.44
vapE	2304-3617	3 4	1731-1763R 2139-2170R	(3657) AAAACGAGCttcggctGCTTTTTTA (3680)	1.92 1.51	1.45 2.00 0.88 0.50
vapE'	9393-10234	16	8985-9013R	(10750) ATGAAGCAGCtaaatagGCTGCTTTTT (10776)	1.76	1.20

The positions of putative promoters and terminators were determined using the sequence of *vap* regions 1 and 3 from *D. nodosus* strain A198 (GenBank accession number L31763).

All potential promoter sequences are shown irrespective of whether they are thought to be required for transcription of a *vap* gene or otherwise. In addition, putative Rho-independent RNA polymerase terminator sequences from within *vap* regions 1 and 3 were identified using the method of Brendel and Trifonov (Brendel & Trifonov, 1984). The putative promoters, terminators and predicted transcript sizes for *vap* genes *A-D*, *toxA* and *vapE'* analysed in this study are shown in Table 4.3.

4.2.2 Northern blot analyses of the *vap* genes in benign and virulent strains of *D. nodosus*

Northern blots were prepared using total RNA isolated from 8 strains of *D. nodosus*, including virulent strains A198, B1006, G1220, H1215, intermediate strain AC3577 and benign strains C305 and H1204. Northern blots were then probed with DNA fragments corresponding to each of the *vap* genes listed in Table 4.1. The DNA fragments used as probes were derived from the *vap* regions of *D. nodosus* virulent strain A198 (Figure 4.1). The results obtained are summarised in Table 4.4.

***Table 4.4*:** Expression of the *vap* genes in 8 strains of *D. nodosus*

	Strain of *D. nodosus*							
Gene	**A198** *(v)*	**C305** *(b)*	**AC3577** *(i)*	**B1006** *(v)*	**G1220** *(v)*	**H1024** *(b)*	**H1215** *(v)*	**1311** *(v)*
vapA	0.76	–	0.76	0.76	0.76	0.76	0.76	0.76
toxA	0.68	–	0.68	0.68	0.68	0.68	0.68	0.68
vapBC	0.70	–	0.70	0.70	0.80	0.70	0.70	0.80
vapD	0.44	–	0.44	0.44	0.44	0.44	0.44	0.44
vapE	2.00 1.45 0.88 0.50	–	2.00 1.45 0.88 0.50	2.00 1.45 0.88 0.50	2.00 1.45 0.88 0.50	2.00 1.45 0.88 0.50	–	2.00 1.45 0.88 0.50
vapE'	–	–	–	–	–	–	–	–

Where a transcript was detected the size in kb is shown. Where no transcript was detected it is indicated by a dash (–); The classification of each strain is indicated as follows: benign *(b)*, intermediate *(i)* and virulent *(v)*.

No gene probe was available which could be utilised as a control for constitutive RNA expression. The possibility that the polynucleotide phosphorlyase gene (*pnpA*) could be utilised for this purpose was investigated, however, no expression of *pnpA* was detected. Thus to ensure that equivalent amounts of total RNA were loaded on RNA gels, and in order to assess the integrity of the RNA after running the denaturing gel, samples were separated on the denaturing gel in duplicate. One set of samples was stained with ethidium bromide, whilst the unstained sample was used for northern transfer. An example of this is shown in Figure 4.2.

4.2.2.1 *vapB* and *vapC*

vapB and *vapC* are in an operon-like arrangement since the last 3 nt of *vapB* overlaps with the 5'-end of the *vapC* gene. From sequence data, a single transcript of approximately 0.75 kb was predicted for *vapB* and *vapC* (probe 3, Figure 4.1). In northern blot analysis all strains except strain C305 hybridised to a single transcript of 0.70-0.80 kb (Table 4.4). This is consistent with the predicted size of the transcript, 0.75 kb, from putative promoter P6 in *vap* region 1 from strain A198. The northern blot for strains B1006, C305, H1204 and H1215 is shown below (Figure 4.2). These results therefore confirm that *vapB* and *vapC* are transcribed as an operon.

In the northern blot shown (Figure 4.2), it is apparent that transcripts detected for strain H1204 and H1215 hybridise more intensely to the probe than the transcripts detected for strain B1006. This is not due to there being more RNA loaded for strains H1204 and H1215 (Figure 4.2, left panel) than for B1006. The observed differential expression may be due to a higher level of transcription in strains H1204 and H1215 compared to B1006, or alternatively, due to the fact that these strains of *D. nodosus* are not isogenic, and so their rate of growth and therefore the level of expression may differ dependent upon how these genes are regulated during growth. The levels of expression observed are not due to

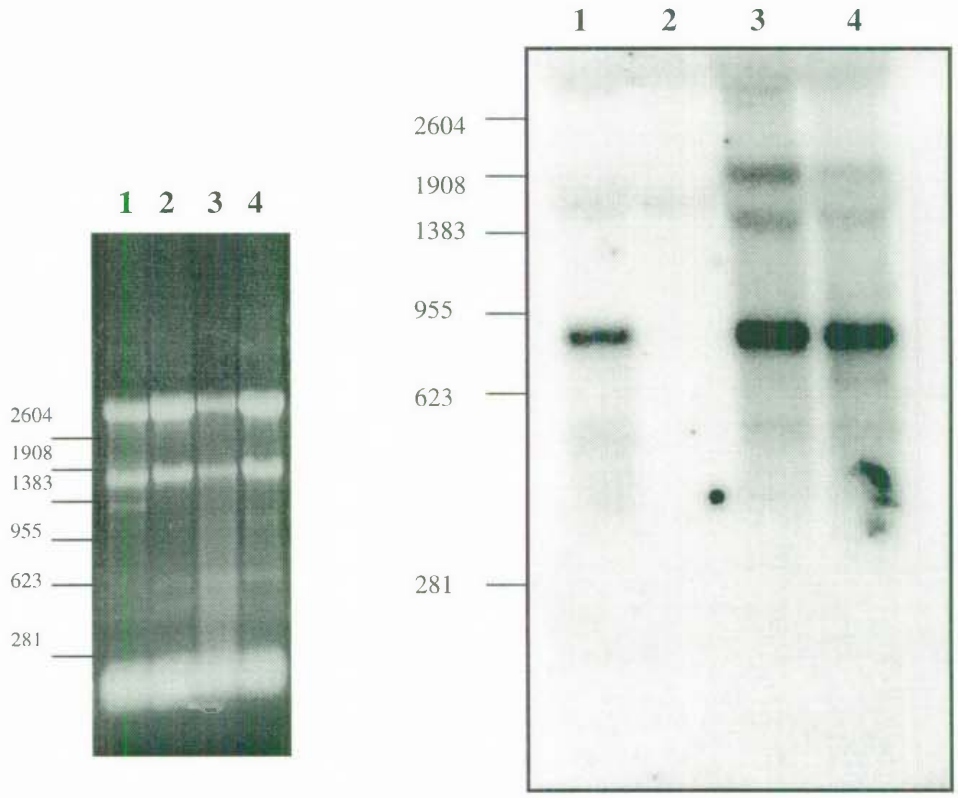

Figure 4.2: Northern blot of *vapBC* in *D. nodosus*. Total RNA was prepared from *D. nodosus* strains as follows: virulent strain B1006 (lanes 1), benign strain C305 (lanes 2), benign strain H1204 (lanes 3), and virulent strain H1215 (lanes 4), and separated on a denaturing gel in duplicate. One set of RNA samples was stained with Ethidium bromide as a control (left panel) whilst the other was used for northern transfer. The results of the northern blot experiment are shown (right panel). The size of RNA markers in bp are indicated at the left of each panel. Ribosomal RNA bands are faintly visible at 1500 and >2604 bp respectively.

copy number either, since strains B1006, H1204 and H1215 each have a single copy of *vapB* and *vapC* (Table 4.4). Irrespective, the differential expression observed above is not related to virulence, since strains B1006 and H1215 are virulent strains, whilst H1204 is a benign strain. Nor was there any correlation between the levels of *vapBC* expression in strain A198, AC3577, G1220 and 1311, and number of copies of *vapBC*.

4.2.2.2 *vapD*

From sequence data, a single transcript of 0.41 kb was predicted for the *vapD* message, since the copies of *vapD* in *vap* regions 1 and 3 of A198 would generate a message of approximately the same size given that the positions of the promoter sequences and terminator sequences relative to the *vapD* gene are identical in both region 1 and region 3 (Table 4.3). In northern blot experiments the *vapD* probe hybridised to a single band of 0.44 kb for all strains of *D. nodosus* , except for strain C305 which does not contain a copy of *vapD* (Table 4.4), and so the results correspond well with predictions. For strains B1006, H1204 and H1215 *vapD* was observed to be expressed at an equivalent level (data not shown). However, the amount of *vapD* transcript was much greater for strains A198 and AC3577 compared to strains G1220 and 1311 (Figure 4.3).

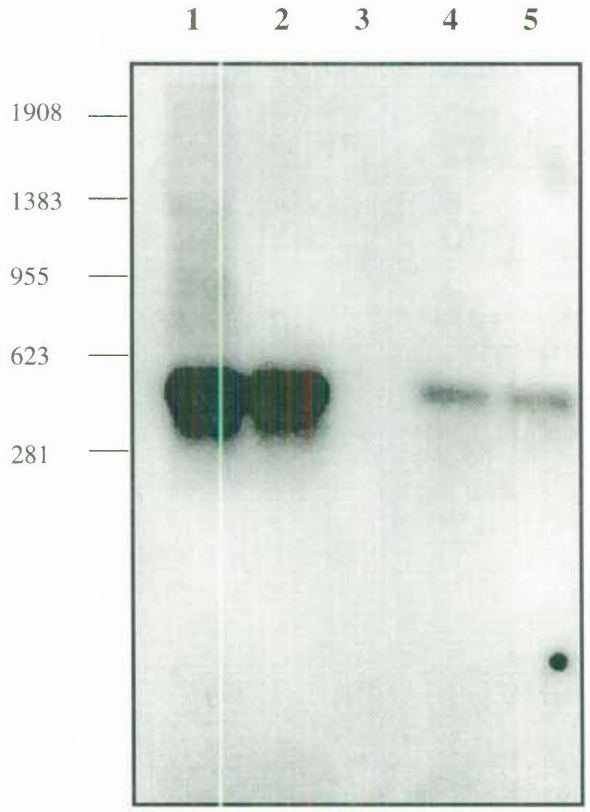

Figure 4.3: Northern blot of *D. nodosus* strains using *vapD* (probe 2, Figure 4.1) as a probe. Strains are as follows: virulent strain A198 (lane 1), intermediate strain AC3577 (lane 2), benign strain C305 (lane 3), virulent strain G1220 (lane 4), and virulent strain 1311 (lane 5). The size of RNA markers in bp are indicated at the left of the panel. Ribosomal RNA bands are faintly visible in panel.

This differential expression was observed when the experiment was repeated using freshly-extracted RNA samples. *vapD* appears to be expressed more highly in *D. nodosus* strains A198 and AC3577 compared to strain G1220 and 1311 (Figure 4.3). However, since all four strains are virulent this differential expression is not associated with the virulence of the strain.

The difference observed might be due to copy number since *D. nodosus* strain A198 contains 3 copies of *vapD* whilst, whilst strain 1311 contains only a single copy of *vapD*. The number of copies of *vapD* in strain AC3577 is unknown, since it contains the *vap* plasmid (Section 1.6.3) and may contain an integrated *vap* element (Section 7.2.8), but is likely to be greater than one copy. However, it was previously noted that strain G1220 contains at least two and possibly three copies of *vapD* (McDonagh, 1994), which is not consistent with the hypothesis that the higher level of expression visible is due to copy number of the *vapD* gene.

The differential expression observed may be attributable to the fact that these strains of *D. nodosus* are not isogenic and consequently their growth characteristics may be quite different, as previously observed in growth experiments with *D. nodosus* strains A198, C305, 1311 and 1311A (Section 3.2.9). If the *vapD* gene is regulated differentially during growth, then a difference in expression may be expected.

4.2.2.3 *vapA* and *toxA*

toxA overlaps the first 13 nt of *vapA*, and consequently *toxA* and *vapA* are arranged as an operon. In addition, since these proteins are thought to comprise a protein toxin and antidote plasmid maintenance system (Bloomfield *et al.*, 1997) (Section 1.6.4.1), the expression of these genes is likely to be functionally linked. Consequently, *toxA*/*vapA* were expected to be expressed as a single transcript of about 0.69 kb (Table 4.2) from the putative

promoter, P8 (Figure 4.1). In northern blot experiments, a single transcript of between 0.68-0.76 kb was faintly detected with both *toxA* (probe 5) and *vapA* (probe 4) probes in all of the strains (A198, AC3577, B1006, C1220, H1204, H1215 and 1311) that contained *toxA* and *vapA*. No hybridisation to strain C305 was observed since this strain contains no *vap* genes (Table 4.1). A northern blot for *toxA* and *vapA* for four strains of *D. nodosus* is shown in Figure 4.4.

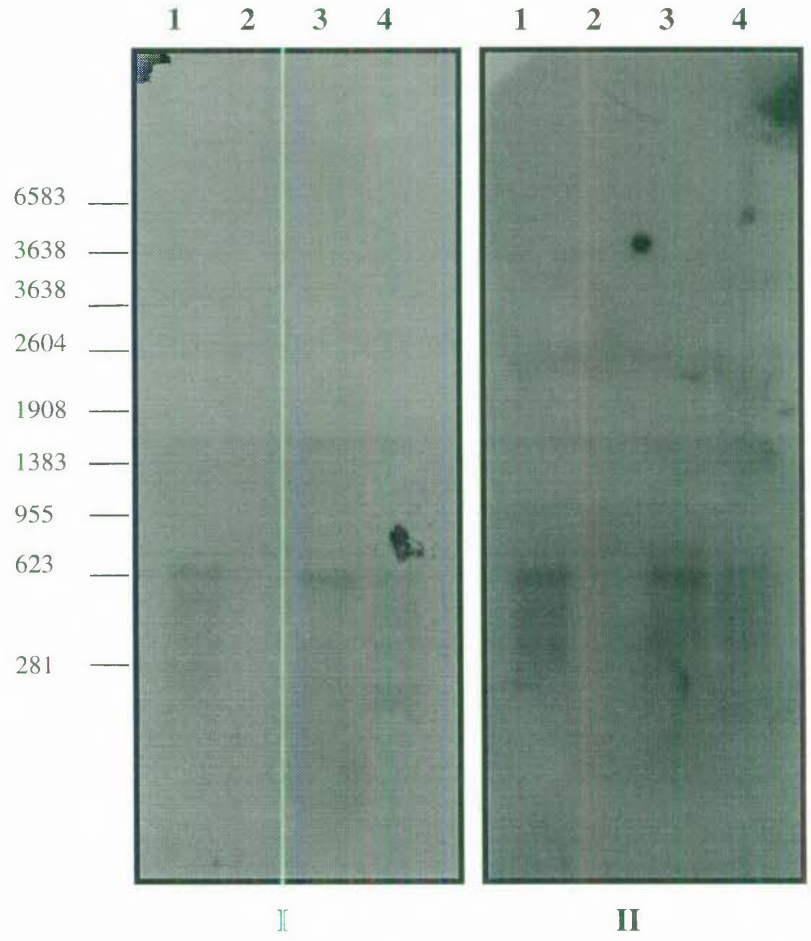

Figure 4.4: Northern blot of *D. nodosus* strains using *toxA* (probe 5, Figure 4.1) and *vapA* (probe 4, Figure 4.1) as a probes in panel I and panel II respectively. Strains are as follows: virulent strain B1006 (lane 1), benign strain C305 (lane 2), benign strain H1204 (lane 3), and virulent strain H1215 (lane 4) The size of RNA markers in bp are indicated at the left of panel I. Ribosomal RNA bands are faintly visible in panel.

4.2.2.4 *vapE*

Based on the position of the P4 promoter (Figure 4.1) and putative transcriptional terminator in *vap* region 1 (Table 4.2), a single transcript of 1.50 kb was predicted for *vapE*. However, in northern blot analyses, in addition to the predicted transcript of 1.45 kb, transcripts of 2.0 kb, 0.88 kb and 0.50 kb were detectable in all strains that contain a copy of *vapE* (A198, AC3577, B1006, G1220, H1204 and 1311), but no bands were visible for strains C305 and H1215 which lack *vapE* (Table 4.3, Figure 4.5).

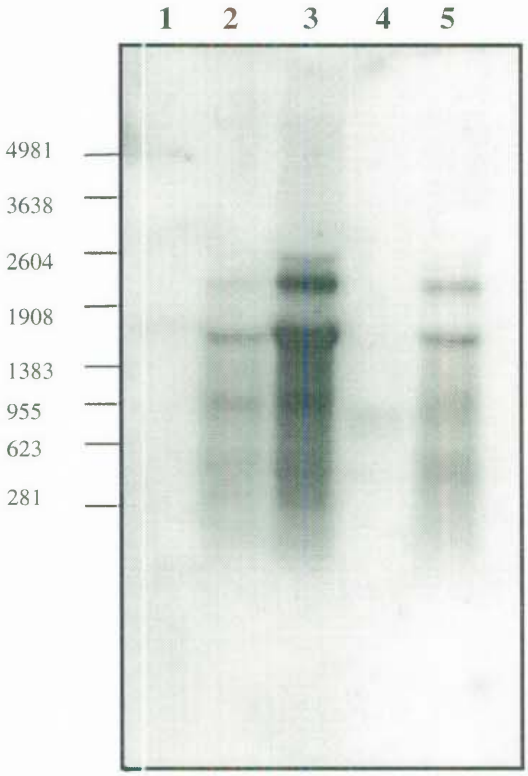

Figure 4.5: Northern blot of *D. nodosus* strains using *vapE* (probe 1, Figure 4.1) as a probe. Strains are as follows: benign strain C305 (lane 1), virulent strain G1220 (lane 2), benign strain H1204 (lane 3), virulent strain H1215 (lane 4), and virulent strain 1311 (lane 5). The sizes in bp of RNA markers are indicated at the left of the panel.

Since one transcriptional product (2.0 kb) is much larger than that predicted for *vapE*, these results suggest that there may be two transcriptional start sites for *vapE*. If *vapG* and *vapH* were transcribed with *vapE* from P3, which is just upstream of P4 (Figure 4.1), an mRNA molecule of 1.9 kb would be produced. Thus, the northern blot

results correlate well with transcription of *vapE* from both P3 and P4. Alternatively, the longer transcript may be due to antitermination. To resolve these possibilities, further experiments need to be undertaken. A probe specific for *vapH* and *vapG* should be utilised as a probe in northern blot experiments in an effort to determine whether these genes are expressed as a part of a large 2.0 kb transcript.

The significance of the smaller, more diffuse bands of 0.88 kb and 0.50 kb is not understood, but may be attributable to processing of *vapE* transcripts. It is interesting that the sum of these two small bands is approximately equally to the size of the transcript that would be produced from P3 (1.45 kb), though it may of course be coincidental. It is not likely these small bands are due to cross hybridisation to *vapE'*, since *vapE* and *vapE'* do not cross hybridise in Southern blot experiments. Furthermore, these smaller bands are not likely to be due to partial hybridisation to a separate *vapG* transcript since a faint band should be visible for strain H1215 which contains a copy of *vapG*, and no such band is visible.

4.2.2.5 *vapE'*

Northern blot analysis did not detect expression of *vapE'* in any of the eight strains tested. This was unexpected since strains A198 and H1215 contain this gene (Table 4.4). It may be that northern blot analysis is not sensitive enough to detect low level expression of this gene, and consequently, RT-PCR analyses should be undertaken in order to determine more conclusively whether or not *vapE'* is expressed. Whether *vapE'* is expressed or not is an interesting question, because in some strains of *D. nodosus* such as virulent strain H1215, *vapE'* is found in place of *vapE*, between *intA* and *vapD* (Figure 1.9), and it was previously suggested that *vapE* and *vapE'* may perform a similar function (Bloomfield *et al.*, 1997). If *vapE'* is not expressed, then *vapE'* is probably not functional and so unable to perform the same function as *vapE*.

4.2.3 *vapHG* have similarity to a bacteriophage immunity region

Northern blot analysis of *vapE* suggested that there may be two transcriptional start sites upstream of *vapE* (Section 4.2.2.4). One of these transcripts is thought to be initiated at P3 (Figure 4.1), and so would include approximately 540 nucleotides upstream of the *vapE* start codon, suggesting that these upstream sequences may be functionally important.

In previous work (Cheetham *et al.*, 1995b), two open reading frames designated *vapH* and *vapG* were identified in these upstream regions (Figure 4.1) in *vap* region 1 of *D. nodosus* strain A198. It was unclear whether *vapH* encoded a protein, because a putative Shine-Dalgarno sequence was not identified upstream of *vapH*, no putative promoter sequence was identified for *vapH*, and in the *vap* plasmid there is a small deletion of 6 nucleotides together with 18 base substitution mutations from within the putative orf (Billington *et al.*, 1996). The observation that regions containing *vapH* and *vapG* are transcribed suggests that the *vapH* and *vapG* genes may have a function. As a result, in this work further analysis of these upstream regions was undertaken.

The putative proteins encoded by *vapH* and *vapG* have no amino acid similarity to sequences in the GenBank databases. However, a search at the nucleotide level in this work indicated that *vapH* has 57.4% identity to the first 296 nt of *orf179* (539 nt) of *Shigella flexneri* (Faubladier & Bouche, 1994) (Table 1.2) which has similarity to the immunity region of satellite bacteriophage P4 (Deho *et al.*, 1992; Faubladier & Bouche, 1994; Ghisotti *et al.*, 1992; Lindqvist, Deho & Calendar, 1993) (Figures 4.6 and 4.7). This immunity region of bacteriophage P4 contains *orf199*.

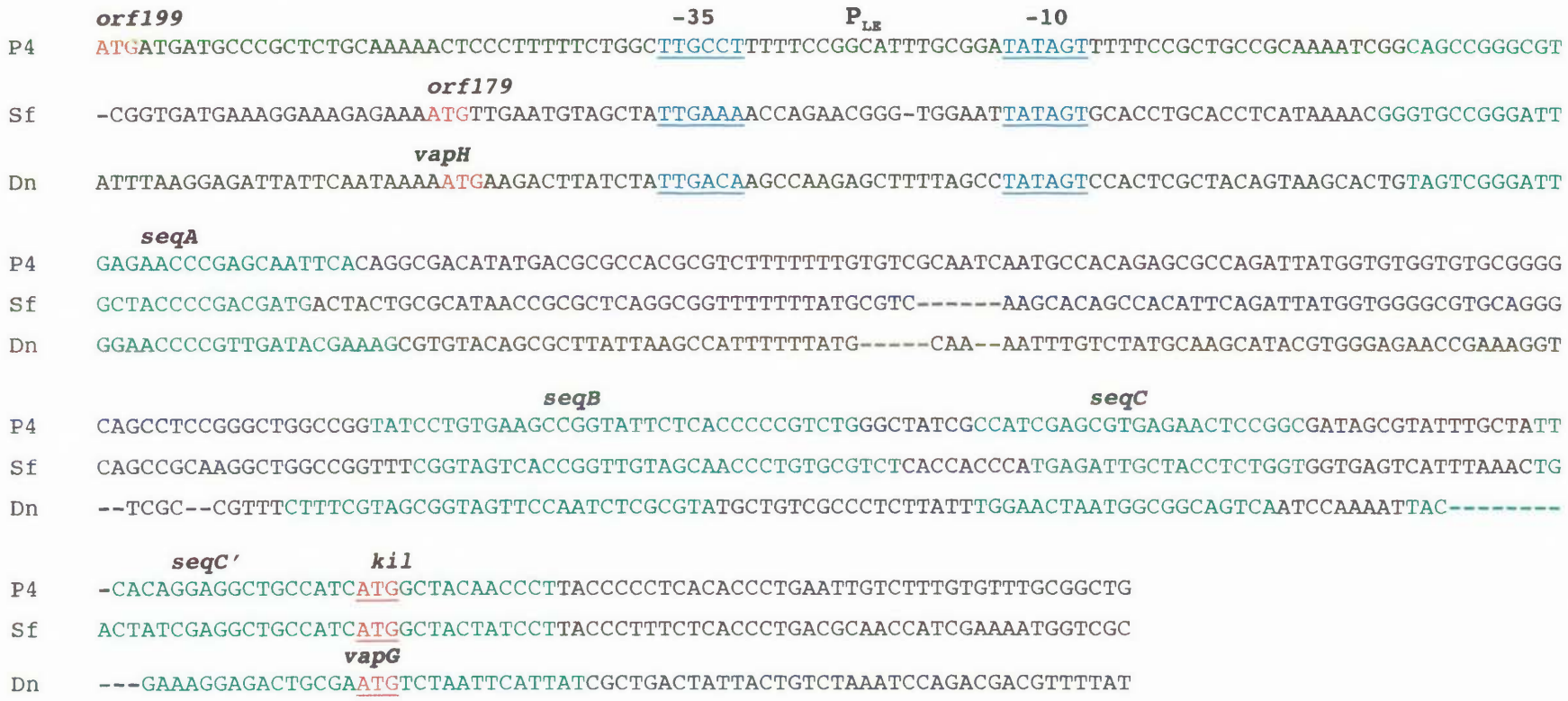

Figure 4.6: Alignment of DNA sequences from the immunity region from bacteriophage P4 (P4) containing *orf199* with regions containing *orf179* from *Shigella flexneri* (Sf) and *vapH* and *vapG* from *D. nodosus* (Dn). Features indicated are as follows: start codons (red type); sequences that can form *seqA*, *seqB*, *seqC* and *seqC'* which interact to form stem-loop structures (green type). The consensus sequence for the P_{LE} promoter of bacteriophage P4 and aligning promoter regions from *S. flexneri* and *D. nodosus* respectively (blue type). Accession numbers are as follows: P4 (MYP4CG), Sf (Z23101) and Dn (L31763).

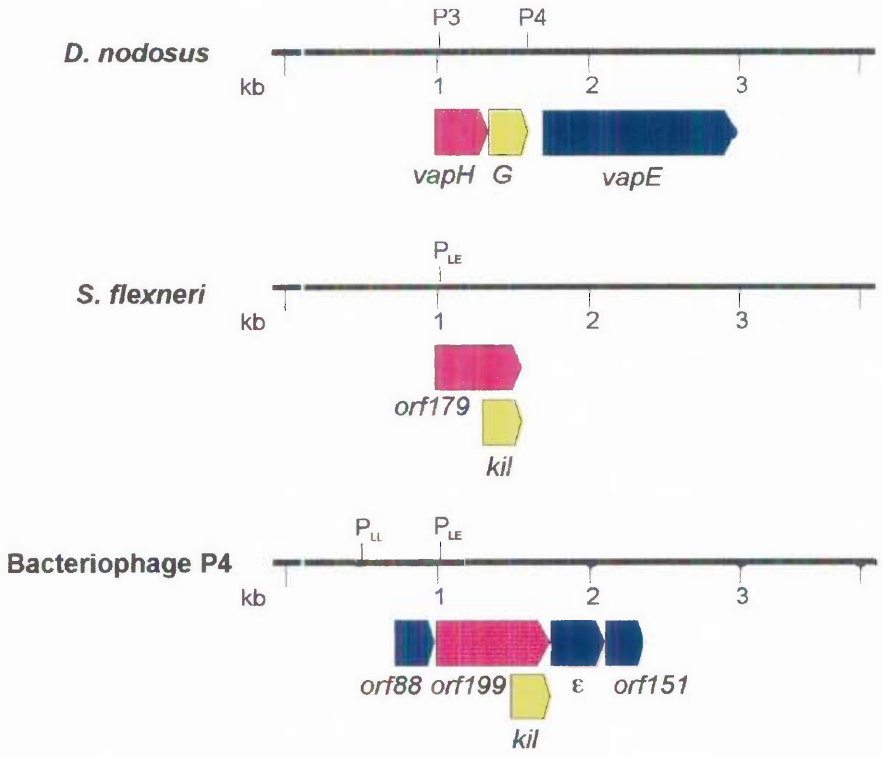

Figure 4.7: Comparison of *D. nodosus* sequences containing *vapH* and *vapG* with related *orf179* from *Shigella flexneri* and *orf199* from satellite bacteriophage P4. Promoter sequences P3 and P4 from *D. nodosus* are shown as are P_{LE} and P_{LL} promoters. The numbers indicate the distance in kb.

Three important features of this immunity region have been described and are as follows:

(i) transcripts that are initiated at the P_{LE} promoter are constitutively transcribed but are not translated, and terminate prematurely (within about 400 nt) (Deho *et al.*, 1992). Within the transcript initiated at the P_{LE} promoter, there are three sequences designated *seqA*, *seqB* and *seqC* respectively which form *seqA-seqB* or *seqB-seqC* secondary structures and a stable *seqB* RNA (*cI* RNA) in P4 lysogens (Lindqvist *et al.*, 1993). It is the formation of these *seqA-seqB* or *seqB-seqC* secondary structures that results in the premature termination of transcription and increases the rate of degradation by RNAses (Deho *et al.*, 1992; Ghisotti *et al.*, 1992; Lindqvist *et al.*, 1993).

(ii) orf199 can be translated if it is transcribed from a different promoter P_{LL} which is located upstream from the P_{LE} promoter (Figure 4.7). In contrast to the P_{LE} promoter, P_{LL} is not involved in lysogeny. During lysogeny, the P_{LE} promoter is constitutively expressed, and due to premature termination produces numerous short transcripts of about 90 bp that bind the *seqB* and *seqC* regions of any transcripts that are produced by the P_{LL} promoters of incoming bacteriophage P4. This binding of complete *orf199* transcripts increases the rate of degradation of *orf199* mRNA and prevents a new P4 particle in the autonomously replicating plasmid state from expressing genes that are required for the lytic cycle and replication. Hence, the P_{LE} transcripts prevent superinfection of bacteriophage P4, and thus are involved in bacteriophage immunity, and maintenance of lysogeny (Ghisotti *et al.*, 1992). It may be by such a mechanism that the *vap* plasmid is maintained extrachromosomally in *D. nodosus* strain AC3577.

(iii) in bacteriophage P4 the 3'-end of *orf199* can be expressed as an internal in-frame *kil* gene whose product inhibits cell division. Since transcripts initiated at P_{LE} terminate before the ribosome binding site of the *kil* gene, the *kil* gene, like *orf199* is only expressed and translated from P_{LL} (Figure 4.7).

Although the *vapH* and *vapG* sequences share nucleotide identity to *orf179* of *S. flexneri*, no such similarity to the nucleotide sequence of the bacteriophage P4 immunity region is detected in database searches. However, the *D. nodosus vapH* and *vapG* region has many of the aforementioned features that are characteristic of the immunity region of bacteriophage P4.

The P3 promoter of *D. nodosus* is almost identical to the P_{LE} promoter of satellite Bacteriophage P4 (Figures 4.6 and 4.7) which is responsible for transcription of the P4 immunity region, which is transcribed but untranslated in the lysogenic state (Deho *et al.*, 1992; Ghisotti *et al.*, 1992; Lindqvist *et al.*, 1993). Furthermore, the start codon of *vapH* is

located upstream of the P_{LE}-like promoter, as are the start codons of *orf199* and *orf179* of bacteriophage P4 and *S. flexneri* respectively (Figure 4.6). Although the start codon of *orf199* is located twelve codons upstream of the P_{LE} promoter, the start codons of *orf179* of *S. flexneri* and *vapH* of *D. nodosus* are both four codons upstream of the P_{LE}-like promoter region (Figure 4.6).

It has been noted that the *orf199* transcript is able to form a stable secondary structure which prevents the translation of the *kil* gene during lysogeny (Lindqvist *et al.*, 1993). Using the mFold program of Zucker (Zucker, 1989), potential secondary structures of the *vapH* transcript were predicted, the most stable of which is shown (Figure 4.8; $\Delta G = -79.4$ kcal/mol), and shows that there is a very high degree of complementarity between *seqA-seqB* and *seqC-seqC'* regions. This structure shows that despite the divergence between *orf199* and *vapH* at the both nucleotide and protein levels, the complementary sequences designated *seqA*, *seqB*, *seqC* and *seqC'* that are present in the P4, are also present in approximately the same position in *vapH* of *D. nodosus* (Figures 4.6 and 4.8). Therefore, despite considerable sequence divergence, these results suggest that *vapHG* has a high degree of similarity to the immunity region bacteriophage P4.

Since *vapH* is similar to the immunity region of P4 which is transcribed but not translated during lysogeny, it is likely that whilst the *vap* region is integrated, *vapH* is similarly transcribed but not translated. One could resolve this hypothesis using western blot analyses. Since it seems that the sequences may be transcribed (Section 4.2.2.4), it also plausible that *vapH* transcripts may have some regulatory role in maintaining an integrated *vap* element. Furthermore the lack of a reasonable Shine-Dalgarno sequence upstream of *vapH* is consistent with the hypothesis that the transcript is produced and not translated.

It is also interesting that the secondary structure predicted for *vapH* overlaps with the Shine-Dalgarno and start codon for *vapG* (Figure 4.8). This would be likely to sterically hinder the translation of *vapG*. In addition, if the formation of the secondary structure itself increases the rate of degradation of the *vapHG* transcript, as was proposed for the P4 immunity region (Deho *et al.*, 1992; Faubladier & Bouche, 1994; Ghisotti *et al.*, 1992; Lindqvist *et al.*, 1993), this too would hinder the translation of *vapG*. An overlap of the Shine-Dalgarno sequence and start codon of *kil* has been noted previously in bacteriophage P4 (Lindqvist *et al.*, 1993).

In bacteriophage P4 the *kil*-encoded protein is involved in the inhibition of cell division (Ghisotti *et al.*, 1992), and only expressed from P_{LL} during the lytic cycle (Lindqvist *et al.*, 1993). Although *vapG* does not share sequence similarity with *kil*, *vapG* may have a similar function, since results herein would suggest that *vapG* is not translated. One could test this hypothesis by determining whether *vapG* inhibits cell division in a strain of *E. coli* which is known to be susceptible to bacteriophage P4. However, the absence of a positive result may not answer the question, since the effects of *vapG* may be specific to the *D. nodosus* cell division machinery, which may be different to that of *E. coli*.

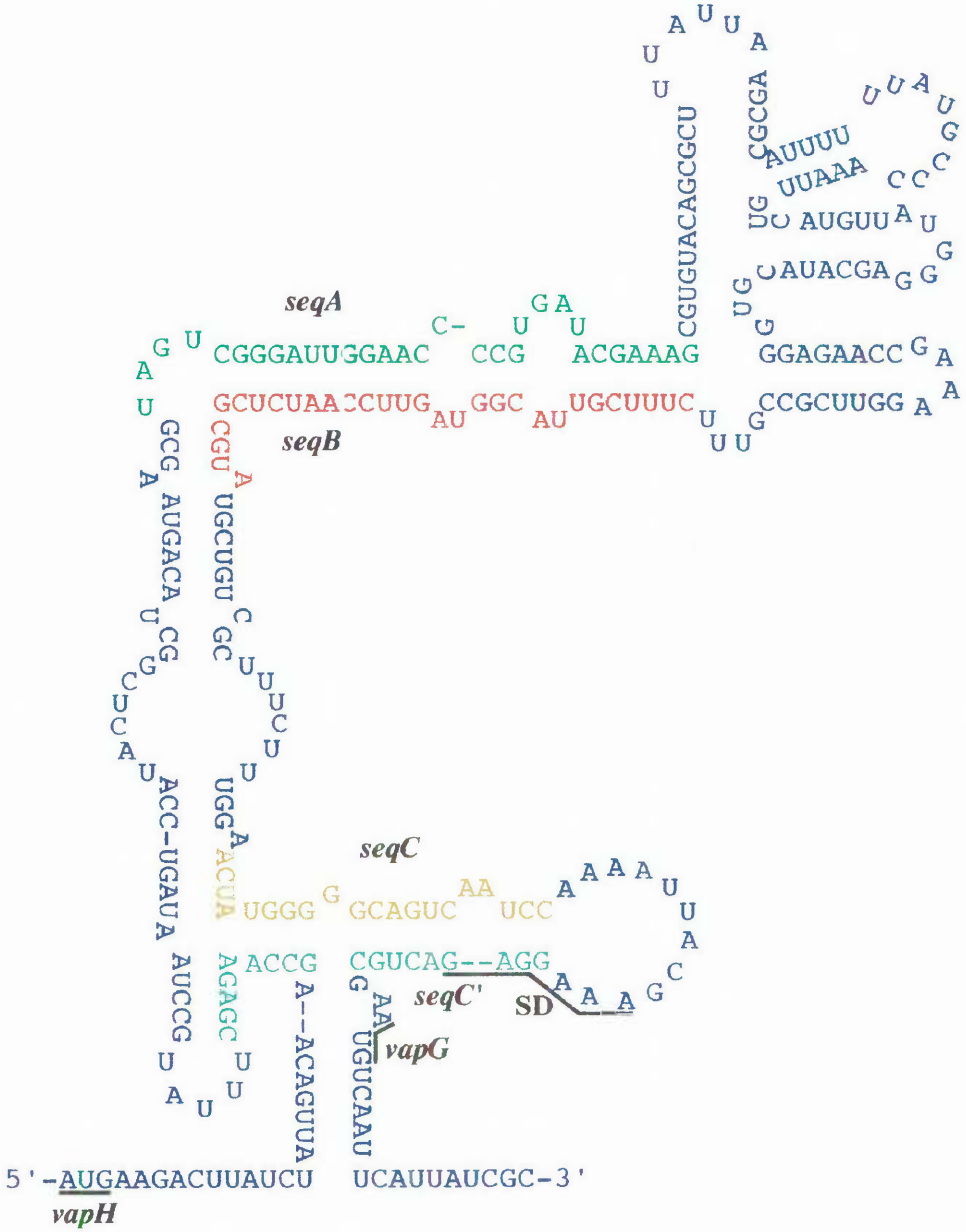

Figure 4.8: A putative secondary structure of RNA (1731-2021) transcribed from P3 promoter. The -35 consensus sequence is located 16 nucleotides downstream from the start codon of *vapH*. Sequences corresponding to *seqA* (light green), *seqB* (red), *seqC* (orange) and *seqC"* (dark green) in the *D. nodosus* RNA molecule are indicated. The start codons of *vapH* and the start codon and Shine-Dalgarno sequence of *vapG* are also indicated.

4.3 Discussion

4.3.1 The *vap* genes are expressed

Northern blot analyses of the *vap* genes in eight strains of *D. nodosus* including benign, intermediate and virulent isolates indicates that, in general, the *vap* genes are expressed in both virulent and benign strains in which they are present. Benign strain H1204 contains a single copy of all *vap* genes except *vapE'* and the results indicate that these genes are expressed. The only *vap* gene for which expression at the RNA level was not detected was *vapE'*, though this lack of expression should be confirmed using a more sensitive method, such as RT-PCR.

There is no evidence that virulent strains of *D. nodosus* exhibit higher levels of expression than benign strains that express the *vap* genes at the RNA level. Benign strain H1204 contains the full complement of *vap* genes (Table 4.1), and all of these genes were shown to be transcribed at levels at least equivalent to those observed for virulent strains of *D. nodosus*. Although higher levels of *vapBC* and *vapD* expression were detected, there was no correlation between higher levels of expression and virulence, nor between higher levels of expression and copy number. It seems most probable that the differential levels of expression are most likely due to the fact that the strains of *D. nodosus* analysed are not isogenic and grow at differing rates during culture (see Section 3.2.9, *(iv)*).

Since results show that the *vap* genes are expressed in both virulent and benign strains of *D. nodosus*, and virulence-related differential expression was not observed, several explanations remain as to why benign strains that contain the *vap* regions have a benign phenotype.

Bacterial pathogenesis is typically multifactorial, and consequently the absence of any one virulence determinant may render a strain avirulent. Thus, usually only a small

number of strains within a species tend to cause disease (Falkow, 1981). It may be that in benign strains, such as strain H1204 which contains the full complement of *vap* genes, which are also transcribed, that other virulence-associated genes may be non-functional, hence making strain H1204 benign.

Alternatively, although the *vap* genes are transcribed in benign strains like H1204, the proteins produced by these genes may be non-functional due to the presence of key mutations that would only be detectable at the DNA or amino acid sequence levels. In the absence of a transformation system it is difficult to test the later hypothesis without laboriously sequencing the *vap* regions in such benign strains.

It is also possible that the *vap* genes may be required for the acquisition of virulence-associated genes (Figure 4.9). It has been proposed (Bloomfield *et al.*, 1997) that an ancestral benign strain of *D. nodosus* may have evolved into different strains of *D. nodosus*, some of which acquired the *vap* element. Those strains which had acquired the *vap* regions may then have been able to acquire other virulence determinants. Since the *vap* element contains a *vapA/toxA* encoded plasmid maintenance system (Bloomfield *et al.*, 1997), once a strain has acquired the *vap* element it can rarely be lost (Figure 4.9).

Alternatively, the specific site in which the *vap* region/s integrate in the *D. nodosus* chromosome, may affect the virulence of the isolate by modulating the expression of genes adjacent to the integration site. This possibility was the subject of further investigations that are discussed in Chapter 7.

It is also possible that *vap* genes have no role in virulence, though this seems unlikely. There is one report of a small number (<2%) of virulent strains (Rood *et al.*, 1996) which appear to lack at least part of the *vap* regions (*vapA-D*), however due to inviability of the stored stocks we have been unable to obtain these strains for analysis. The preferential distribution of the *vap* element in >98% of virulent strains does support

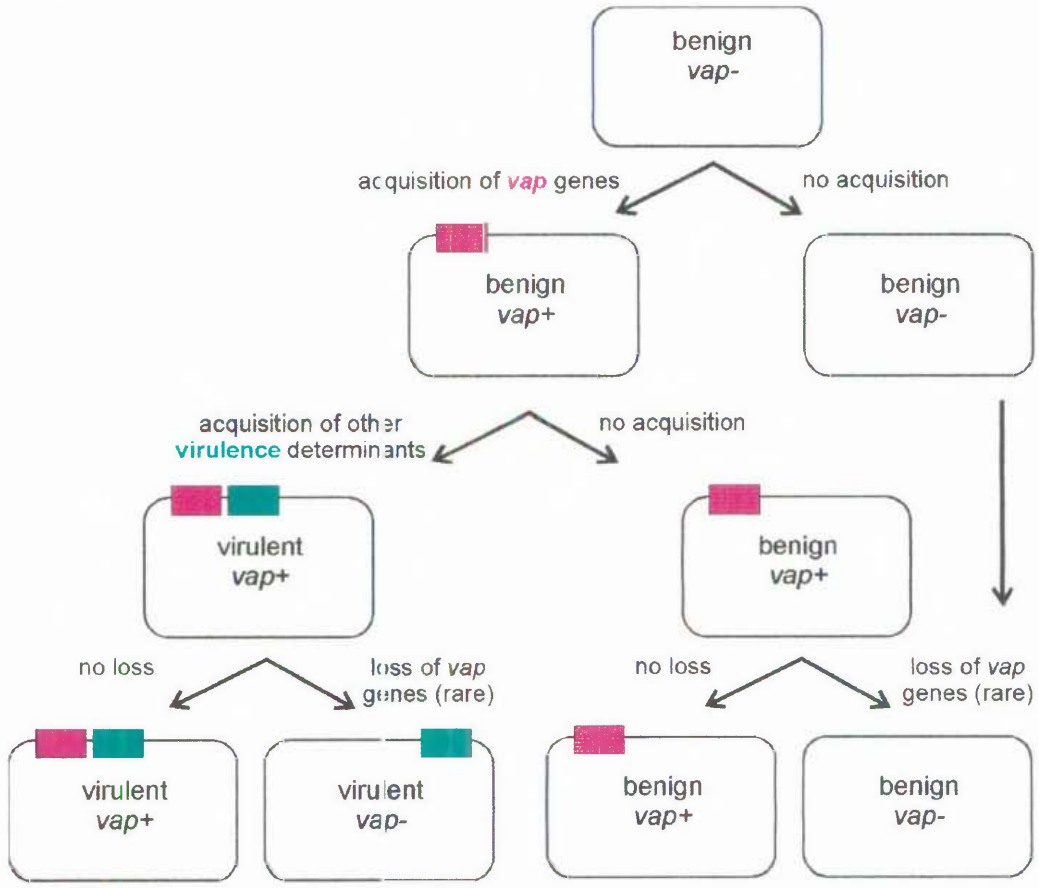

Figure 4.9: Model of how the *vap* element may be required for the acquisition of virulence-determinants in *D. nodosus* (B. Cheetham).

the hypothesis that these regions may have some direct or indirect role in virulence in *D. nodosus*.

The transcript sizes predicted from the positions of putative promoters and Rho-independent terminators, correlated very well with the actual transcript sizes observed for specific *vap* genes in northern blot analyses. The results also confirmed that *vapB* and *vapC* are transcribed as an operon as are *toxA* and *vapA*. Unexpectedly, it was found that there are likely to be two transcriptional start sites for *vapE*, from P3 and P4. The analysis of sequences that are likely to be included in the P3 transcript led to the identification of a region that is likely to be transcribed and not translated, and has similarity to regions encoding the immunity determinant of defective bacteriophage P4.

4.3.2 The *vap* element is a bacteriophage P4-like element?

During lysogeny most temperate bacteriophages directly or indirectly inhibit the expression of lytic functions *via* the negative regulation of operons that encode proteins that perform functions required for the lytic cycle or replication (Deho *et al.*, 1992). The maintenance of lysogeny usually involves the production of a repressor molecule which negatively regulates a given promoter, and this condition is referred to as superinfection immunity. Examples of superinfection immunity have been described in numerous bacteriophages including Lambda (Gussin *et al.*, 1983), and lambda-like bacteriophages, P1 (Yarmolinski & Sternberg, 1988), P2 (Bertani & Six, 1988), P22 (Poteete, 1988) and Mu (Howe, 1987). In contrast, in the integrated state, bacteriophage P4 does not encode a repressor protein and instead a small RNA is the immunity factor (Deho *et al.*, 1992; Ghisotti *et al.*, 1992; Lindqvist *et al.*, 1993). Furthermore, replication functions are inhibited by premature termination or processing of primary transcripts produced by a constitutive promoter (P_{LE}) (Lindqvist *et al.*, 1993).

The identification of a region within the *vap* element (*vapHG* region) of *D. nodosus* which may also encode proteins that perform an immunity function like that described for P4 (Section 4.2.3) again raises the possibility that *vap* regions are part of a defective bacteriophage. It was previously proposed that the *vap* regions may comprise part of a prophage or integrated plasmid (Cheetham *et al.*, 1995b) (Section 1.6.3).

Although the *vap* regions contain an integrase gene (Table 1.2) (Cheetham *et al.*, 1995b) that is most closely related to integrase genes from the lambda family of integrases, of which the bacteriophage P4 integrase is one, the other genes located within the *vap* region, including *vapA-D*, *toxA* and *vapE* are most similar to genes that are found on plasmids. In addition, sequence analyses strongly suggest that *vapA-D* and *toxA* are all involved in plasmid maintenance functions, though this does not eliminate the possibility that a defective bacteriophage has acquired plasmid maintenance genes. No other

bacteriophage-associated genes have been identified within the *vap* regions, whereas in the P4 genome genes involved in capsid size determination are present (Lindqvist *et al.*, 1993). It is difficult to resolve the precise nature of the *vap* element, however the parallels between P4 and the *vap* element are of interest.

Both the *vap* element and P4 integrate site-specifically into the 3'-end of tRNA genes ($tRNA_{ser}$ and $tRNA_{leu}$ respectively), are flanked by a 19 bp attachment site, and no excisionase has been identified in either element (Bloomfield *et al.*, 1997; Cheetham *et al.*, 1995a; Lindqvist *et al.*, 1993). Extracellularly, bacteriophage P4 exists as a virion which is able to inject its DNA into *E. coli* and other Gram-negative bacteria (Lindqvist *et al.*, 1993). The P4 genome lacks the genetic determinants required to generate capsid and tail proteins, and those proteins required for lysis of the host cell (Six & Klug, 1974). Consequently it is only when P4 infects a cell containing the helper phage P2 that it may enter either the lytic or lysogenic cycle (Six & Klug, 1974). Where the helper phage is not present, P4 may integrate into the host genome (Six & Klug, 1974), or may replicate as a multicopy plasmid (Deho *et al.*, 1984; Goldstein, Sedivy & Ljungquist, 1982). Consequently it has been suggested that P4 may be an episomal element that has evolved the ability to utilise a helper bacteriophage for horizontal transfer using a novel specialised transduction mechanism (Goldstein *et al.*, 1982; Lindqvist *et al.*, 1993). It is possible that, like P4, the *vap* element is a plasmid-like element which has evolved the ability to utilise a similar mechanism for horizontal transfer.

The majority of cells which P4 infects contain P4 in the integrated state, whilst only about 1% contain P4 plasmids (Deho *et al.*, 1984). Furthermore, cells that contain P4 as a plasmid may also contain P4 in the integrated state (Deho *et al.*, 1984). This is unusual because in most bacteriophages the lysogenic (immune) and lytic patterns of gene expression are mutually exclusive, the integrated (immune) state preventing the lytic/replicating state and vice versa (Lindqvist *et al.*, 1993). After infection and

subsequent replication, P4 can be stably maintained as an integrated prophage or, if it contains mutations that still allow replication functions to be expressed, it is also maintained as a multicopy plasmid (Birge, 1994; Lindqvist *et al.*, 1993).

Analyses of the *vap* regions in many strains of *D. nodosus* suggests that the *vap* element is also predominantly found in the integrated state (Bloomfield *et al.*, 1997; Cheetham *et al.*, 1995b; Katz *et al.*, 1991; Katz *et al.*, 1992; Katz *et al.*, 1994; Rood *et al.*, 1996), since only one *D. nodosus* strain (AC3577) has been identified which contains the plasmid state (Billington *et al.*, 1994). This same strain appears to contain the *vap* element in the integrated state (Section 7.2.8). Therefore, perhaps the presence of certain mutations have allowed the *vap* element to be maintained as a plasmid as well as in the integrated form in strain AC3577.

Cells that contain P4 in the plasmid form yield a high number of long filamentous cells, since the plasmid P4 appears to interfere with cell division and also affects the growth rate of plasmid carrying clones (Deho *et al.*, 1984). This is interesting because strains of *D. nodosus* that contain the *vap* genes also tend to have more elongated cells and in general grow less well than benign strains that contain the *vap* regions (Cheetham, Whittle & Katz, 1998).

The suggestion that *vapH* and *vapG* comprise an immunity region in the *vap* element is interesting given that in *D. nodosus* strain A198 there are three copies of the *vap* element, *vap* regions 1, 2 and 3. One would expect that the expression of an immunity region would inhibit the further integration of the same element into the host genome. It is therefore significant that although *vapH* and *vapG* of *vap* region 1 share identity with *vapG'* and regions upstream of *vapG'* in *vap* region 3, and cross hybridise in Southern blots, these sequences are not identical. These subtle differences in nucleotide sequences may be necessary if multiple copies of the *vap* element are to be acquired by the one host.

The *vapGH* segment of *vap* region 2 has not been sequenced and so it is unknown whether the corresponding sequence in *vap* region 2 differs from the segment in *vap* region 1 and *vap* region 3. It is therefore unclear whether the *vap* regions of *D. nodosus* strain A198, are truly part of the same genetic element, or are different but related elements, since *vap* regions 1, 2 and 3 of A198 (Figure 1.7) are all distinguishable.