REVIEW ARTICLE

Gas vesicle proteins

Anthony E. WALSBY and Paul K. HAYES

Department of Botany, University of Bristol, Woodland Road, Bristol BS8 1UG, U.K.

INTRODUCTION

One of the most extraordinary structures found in bacteria is the gas vesicle. It has the form of a shell of protein molecules enclosing a gas-filled space. The properties of this structure — its rigidity, its contrasting inner and outer surfaces, its permeability to gases and impermeability to liquid water — are all provided by its constituent proteins.

Gas vesicles provide many aquatic bacteria with buoyancy. They are confined to prokaryotes and, while they are best known in the cyanobacteria or blue-green algae (Walsby, 1981a) they also occur in many other bacterial groups (Walsby, 1981b) including two of the three major archaeobacterial lineages, the methanogens (Archer & King, 1984) and the extreme halophiles (Larsen *et al.*, 1967). In all cases the gas vesicles appear to be homologous structures. They have very similar shape and fine structure and in those which have been analysed the major proteins are homologous.

The gas vesicle has the form of a hollow cylindrical tube, closed by a hollow, conical cap at each end (Bowen & Jensen, 1965; Jost, 1965). The structure is formed by a gradual assembly process, starting as a hollow bicone which enlarges to a critical diameter and then grows in length by the extension of the cylindrical mid-section (Waaland & Branton, 1969; Lehmann & Jost, 1971). The diameter of the cylinder varies in different organisms (Walsby, 1972). In cyanobacteria it is usually within the range 60–110 nm (Hayes & Walsby, 1986; Walsby & Bleything, 1988). The most extreme form is in the halobacteria where the end caps continue to widen as the structure grows, producing a rounded, spindle-shaped bicone which may be up to 250 nm wide (Larsen *et al.*, 1967; Stoeckenius & Kunau, 1968).

Both the conical end caps and central cylinders of gas vesicles are made up of 4–5 nm-wide ribs that run at right angles to the long axis of the structure (Jost, 1965). X-ray diffraction indicates a constant rib repeat of 4.6 nm in *Anabaena flos-aquae* (Blaurock & Walsby, 1976). The ribs represent either stacks of hoops or turns of a shallow spiral, though the latter is more likely (Walsby, 1989). The gas vesicle wall has an average thickness of less than 2 nm; the value indicated by X-ray diffraction of stacked, collapsed gas vesicles is 1.95 nm (Blaurock & Walsby, 1976) and that calculated from mass, density and volume measurements is 1.80 nm (Walsby & Armstrong, 1979, as modified by Hayes *et al.*, 1986).

The isolation of gas vesicles

Gas vesicles released from lysed cells can be separated from most other cell constituents by low-speed centrifugation; they float to the surface of the preparation where they can be drawn off (Larsen *et al.*, 1967). During centrifugation the acceleration applied must be restricted so that the pressure generated in the liquid layer does not

exceed the minimum pressure at which the gas vesicles collapse (Walsby & Buckland, 1969). Very pure preparations of intact gas vesicles can be obtained by repeated centrifugation followed by rinsing, membrane filtration and detergent washing; detailed methods are given by Walsby (1974) and Hayes (1988).

A summary of the physical properties of gas vesicles

The gas vesicle fulfils its function of providing buoyancy by creating a stable, gas-filled structure of very low density. The stability of the structure is determined by a number of physical properties, which have been characterized largely by experiments with isolated gas vesicles.

The first essential property is that the hollow structure is rigid. When the *Anabaena* gas vesicle is subjected to a pressure rise of 1 bar (0.1 MPa) its volume decreases by only 1 part in 645. From this measurement the elastic modulus of the vesicle wall is calculated to be about 2.8 GPa, within the expected range for polymers like nylon or proteins (Walsby, 1982). The gas vesicle will withstand, for example, the combination of cell turgor pressure (perhaps 0.4 MPa) and hydrostatic pressure in the water column of a lake (0.2 MPa at 20 metres) with a less than 1% change in volume, and the buoyancy it provides will therefore be little diminished. At a certain 'critical pressure' the structure collapses flat. It cannot be reinflated by gas (Walsby, 1969, 1971), though ¹⁴C-labelling studies suggest the components of collapsed gas vesicles may be recycled into the new gas vesicles that are subsequently assembled (Hayes & Walsby, 1984).

The gas vesicle does not maintain its gas space by keeping gases trapped inside it: the walls of the structure are highly permeable to gases, which rapidly diffuse in and out (Walsby, 1969). The half-time for equilibration of gases such as nitrogen is a millisecond or less (Walsby, 1984) and there is therefore no possibility that the gas vesicle can store gases. The gas vesicle in a cell will contain gases constantly equilibrating with those dissolved in the surrounding water; the gas in the vesicle will normally be similar to air at atmospheric pressure. The hydrostatic pressure is not directly transmitted to the gas but is borne largely by the rigidity of the gas vesicle wall.

The gas vesicle wall is impermeable to liquid water, which is kept out by the effects of surface tension (Walsby, 1971) owing to the hydrophobic nature of the inner surface (Worcester, 1975). Partitioning experiments have shown that the outer surface of the gas vesicle is, in contrast, hydrophilic (Walsby, 1971). This is important in minimizing forces of interfacial tension on the outer, highly curved surface, which would tend to collapse the structure.

Chemical composition of the gas vesicle

The first chemical analyses of isolated halobacterial

gas vesicles indicated that protein was a component (Stoeckenius & Kunau, 1968). Protein was the only component found in gas vesicles purified from the cyanobacteria *Anabaena* (Walsby & Buckland, 1969) and *Microcystis* (Jones & Jost, 1970). No carbohydrate or lipid could be detected (Jones & Jost, 1970; Falkenberg *et al.*, 1972).

Electrophoretic analyses of purified cyanobacterial gas vesicles run on phenol/urea/acetic acid (PUA) gels revealed only a single mobile band, though much protein remained at the origin (Jost & Jones, 1971; Falkenberg *et al.*, 1972). Jones & Jost (1971) showed that extracting the non-penetrating residue with 90% formic acid gave more of the mobile band and it was therefore concluded that the residue represented the same protein, perhaps in some denatured form. Similar assumptions were made about the results obtained with SDS gels, which also revealed a minor mobile band with a major non-mobile residue (Walker & Walsby, 1983). It is now known, however, that the material at the origin of SDS electrophoresis gels represents the main gas vesicle protein, a small but undissolved protein of about 7.5 kDa now referred to as GVPa; the mobile band represents a quite different protein, GVPC, of about 22 kDa in *Anabaena* and *Microcystis*, which forms a minor component of the structure (see below). The identity of bands on phenol/urea/acetic acid gels still requires investigation.

Until these facts were elucidated there was much confusion over the nature and molecular mass of what was assumed to be a single type of gas vesicle protein. Full details of this work were reviewed by Armstrong & Walsby (1981), who discussed the conflicting evidence suggesting the disparate values of 7.5, 15, 22 and 55 kDa. Before closing the door on this era of gas vesicle research, however, it is worth mentioning some of the principal findings and putting them into perspective.

The value of 7.5 kDa was indicated by the volume of the X-ray crystallographic unit cell (Blaurock & Walsby, 1976), further details of which are given below. The 15 kDa value was suggested by phenol/urea/acetic acid gel electrophoresis of proteins run with native positive charge at low pH [Jones & Jost, 1971; Buckland, 1971 (see Fig. 5 of Walsby, 1972); Falkenberg *et al.*, 1972; Falkenberg, 1974; Walsby, 1977; Simon, 1981]. It was known that these systems did not give a reliable estimate of molecular mass yet they continued to be used because of difficulties experienced with SDS gels.

Acid hydrolysis of entire gas vesicles yielded the principal amino acids in approximately integer ratios. By relating the amounts to the least abundant residues, Falkenberg *et al.* (1972) claimed that the ratios supported a molecular size of 15 kDa, though, as Armstrong & Walsby (1981) argued, many of the ratios were even integers and supported an empirical value of half of this, 7.5 kDa; the actual values could be simple multiples, e.g. 7.5, 15, or 22.5 kDa. The ratios are now known to reflect the composition of the major 7.5 kDa protein component, which dominates the overall composition.

The first clear evidence of a 22 kDa protein by SDS/polyacrylamide-gel electrophoresis of *Anabaena* gas vesicles was provided by Walker & Walsby (1983) though they did not recognize that this protein was a separate minor component. They found that the 22 kDa protein was present in freshly prepared gas vesicles but disappeared after prolonged storage at 5 °C. They presumed

that the loss was caused by a gradual denaturation, but it probably results from degradation of the minor surface protein, which leaves the major structure intact.

The separate identity of the two types of gas vesicle protein was finally resolved by the combination of protein and gene sequencing described below.

AMINO ACID SEQUENCING OF GVPa

The initial attempts to sequence the gas vesicle protein were made, in ignorance of the existence of two types of protein, by analysing entire gas vesicle preparations and digests of such preparations. All of the sequences obtained relate to the major 7.5 kDa GVPa proteins, which recent studies (Hayes *et al.*, 1988; Walsby & Hayes, 1988) have shown to account for 97 mol % of the protein present. The first analyses were made on *Microcystis* gas vesicles by Weathers *et al.* (1977) who obtained the sequences of the N-terminal tetrapeptide AVEK and of four other non-overlapping peptides.

Walker & Walsby (1983) determined the sequence of the first 64 residues of *Anabaena* GVPa from overlapped sequences of peptides obtained by tryptic digestion and a chemical cleavage at the single tryptophan residue. The sequence of the one remaining part, the 15-residue C-terminal peptide, was subsequently determined (Walker *et al.*, 1984); it showed a precise 8-residue overlap with residues 56–64 of the published sequence. Further detailed sequence analysis of 10 proteolytic cleavage products confirmed that the entire protein had been covered in this 70-residue sequence (Hayes *et al.*, 1986) and it was concluded that the size of the molecule, which corresponded closely to that of the crystallographic unit cell, was at variance with that suggested by SDS-gel electrophoresis.

GVPa proteins from other cyanobacteria

Automated Edman degradation permitted extensive sequencing of the principal protein present in the gas vesicles isolated from a number of other cyanobacteria (and two halobacteria). Four of these sequences are shown in Fig. 1.

A remarkable feature of the amino acid sequences is the degree of homology. Four of the cyanobacteria investigated are from different orders or families: *Microcystis*, a unicellular form; *Oscillatoria*, a non-heterocystous filament; *Anabaena*, from the Nostocaceae with intercalary heterocysts; and *Calothrix* from the Rivulariaceae with terminal heterocysts; yet the amino acid sequences of their principal gas vesicle proteins differ by only one to three residues in the 46–70 analysed.

Sequences derived through the nucleotide sequence of *gvp* genes

Our initial attempts to obtain gas vesicle protein genes from *Anabaena* by probing a genomic library with a family of synthetic oligonucleotides constructed to be homologous to the region coding for the sequence DAWVR (residues 26–31; see Fig. 1) were not successful, perhaps because clones carrying multiple copies of the highly-expressed gas vesicle genes are killed due to overexpression of the protein. The first success in cloning a gas vesicle protein gene was obtained by Tandeau de Marsac *et al.* (1985) using *Calothrix* sp., a cyanobacterium in which gas vesicles are expressed only when hormogonium production is induced.

Fig. 1. Comparison of GVPa amino acid sequences

1, *Anabaena flos-aquae* CCAP 1403/13f; 2, *Calothrix* sp. PCC 7601; 3, *Halobacterium halobium* NRL, plasmid encoded; 4, *Halobacterium halobium* NRL, chromosome encoded. Vertical lines show residues conserved in all four proteins. Residues denoted with large capitals have been determined by direct amino acid sequencing; where a small capital is used the identity of the residue has been deduced from the nucleotide sequence of the gene. The sequences listed are taken from: Oesterhelt *et al.* (1984) and Walker *et al.* (1984), protein sequence 3; Tandeau de Marsac *et al.* (1985), gene sequence 2; Hayes *et al.* (1986), protein sequences 1 and 2; P. K. Hayes, unpublished work, gene sequence 1; Surek *et al.* (1988), protein sequences 3 and 4; Horne *et al.* (1988), gene sequences 3 and 4.

Using a single 29-mer oligonucleotide constructed against the sequence ILDKGIVIDA present in the *Anabaena* protein (residues 18–27; see Fig. 1) as a probe, Tandeau de Marsac *et al.* (1985) isolated a 709 bp *HindIII*–*HincII* restriction fragment. Nucleotide sequence analysis revealed a 213 bp open reading frame whose inferred amino-acid sequence was nearly identical to that of the *Anabaena* gas vesicle protein (Tandeau de Marsac *et al.*, 1985). Gas vesicles isolated from the same strain of *Calothrix* were analysed directly, using the gas-phase sequencer; the sequence obtained, covering the first 58 residues, was identical to the inferred sequence, but minus the *N*-terminal methionine (Hayes *et al.*, 1986). Damerval *et al.* (1987) subsequently extended the nucleotide sequence of the *gvp* operon and showed that 105 bp downstream from *gvpA* was a second gene, *gvpB* (now redesignated *gvpA2*) which encoded a protein of identical amino acid sequence. A further 215 bp downstream was a third open reading frame *gvpC*, whose protein product was subsequently identified as a different type of gas vesicle protein, described below.

We used *gvpA* from *Calothrix* to probe a genomic library of non-gas-vacuolate mutants of *Anabaena flos-aquae*, which we have found to contain three (and possibly more) copies of the *gvpA* gene in tandem repeats separated by short intergenic regions of 105 bp (Hayes & Walsby, 1988). The inferred amino acid sequence differed from the sequence that we determined directly using the *Anabaena* gas vesicles in two respects. (1) The gene sequence indicated the last four residues at the *C*-terminal end as Ala-Val-Pro-Ala, instead of Val-Pro-Ala-Ala, which was erroneously suggested by misidentification of the dansylated *C*-terminal dipeptide (see Hayes *et al.*, 1986). (2) At residue 37 the gene sequence indicated glutamic acid instead of glutamine. The correct *Anabaena* sequence shows only one residue that differs from the *Calothrix* sequence.

Other GVPa-like proteins

Damerval *et al.* (1989) have reported that upstream from the *gvpABC* operon of *Calothrix* is a gene *gvpD* whose sequence closely resembles *gvpA* and whose protein product differs from GVPa in only a few residues.

Such a protein must have a very similar function to GVPa, which forms the ribs of the structure; as discussed below it might make the ribs of the end cones.

GVPa in halobacteria

Direct amino acid sequencing of gas vesicles isolated from two species of *Halobacterium* yielded sequences of the first 26 and 43 amino acids, respectively (Walker *et al.*, 1984; Oesterhelt *et al.*, 1984). Remarkably, the sequences of these archaebacterial proteins were highly similar to those of the cyanobacterial GVPs. The major differences were next to the *N*-terminal alanine residue which was followed by three deletions and three substitutions.

The high degree of similarity suggested that the *Calothrix* gene probe might facilitate the isolation of the homologous gene in halobacteria, and this was successfully achieved for *H. halobium* by DasSarma *et al.* (1987) and by Horne *et al.* (1988).

Previous studies on halobacteria indicated that mutants lacking the ability to produce gas vesicles arose very frequently (Larsen *et al.*, 1967; Stoeckenius & Kunau, 1968; Krantz & Ballou, 1973) and it was suggested that genes for gas vesicle production might be carried on plasmids (Walsby, 1977), which are subject to frequent loss or modification. Simon (1978) demonstrated that a gas-vesicle-defective mutant of *H. salinarum* had lost a 44 MDa plasmid. Weidinger *et al.* (1979) found that in *H. halobium* there were changes in the 100 MDa plasmid of the gas-vesicle-defective mutant. The modified plasmid contained a 3.6 kbp insertion element. Analysis of such mutants indicated that gas vesicle formation was determined or controlled by plasmid genes. The story was complicated, however, by the later finding of Simon (1981) that, in late exponential phase, cultures of the gas-vesicle-defective mutants did produce gas vesicles, albeit in smaller amounts. Moreover, the gas vesicles of the mutant were predominantly cylindrical with conical ends, in contrast to those in the wild type which were mostly spindle-shaped.

Direct amino acid sequencing performed by Surek *et al.* (1988) on gas vesicles isolated from the wild type *H. halobium* showed an *N*-terminal sequence identical to

that obtained previously (Walker *et al.*, 1984; Oesterhelt *et al.*, 1984), while those of the gas-vesicle-defective strain showed two changes, Gly → Ser at residue 7 and Val → Ile at residue 28. Using Southern hybridization with a synthetic oligonucleotide Surek *et al.* (1988) showed that genes homologous to *gvpA* occurred on both the chromosome and plasmid of *H. halobium*. This was confirmed by Horne *et al.* (1988) who cloned and sequenced the two genes. The chromosomal gene encoded a 9 kDa protein of 78 amino acid residues, while the gene on the 150 kb plasmid encoded an 8.7 kDa protein of 75 residues; as well as the differences at residues 7 and 28, the plasmid gene lacked three amino acid residues (Pro-Glu-Pro) near the C-terminus. DasSarma *et al.* (1987) reported a single gene sequence from the plasmid of *H. halobium*. It evidently corresponds with the plasmid gene of Horne *et al.* (1988), but an omitted cytosine residue generated a frameshift.

The studies on plasmid mutants of *H. salinarum* and *H. halobium* reported by Simon (1981), Surek (1987) and Horne *et al.* (1988) together indicate that the smaller plasmid-encoded gas vesicle protein is constitutively produced in high amount while the chromosome-encoded protein accumulates only in the stationary phase. The mutants lacking the plasmid gene produce a preponderance of cylindrical gas vesicles rather than the normal spindle-shaped ones. The small differences in amino acid sequence of the two proteins may specify these morphological differences. Falkenberg (1974), Simon (1981) and Surek (1987) have described different proteins separated by phenol/urea/acetic acid electrophoresis from the wild-type and mutant halobacteria. In view of the mistakes made over the identity of the cyanobacterial proteins, however, it is necessary now to identify these electrophoretic bands unambiguously by *N*-terminal sequencing.

Pfeifer *et al.* (1989) have investigated the mutation rates of the two genes in *H. halobium*. The plasmid gene mutates at the very high frequency of 10^{-2} owing to the combined effects of integration of insertion elements and deletion events. The chromosomal gene mutates much

less frequently, at 10^{-5} , and neither insertions nor deletions are involved.

THE SECOND GAS VESICLE PROTEIN, GVPc

The possibility that there might be an additional protein component in *Anabaena* gas vesicles was suggested by the identification of peptides not derived from the GVPa sequence in proteolytic digests of the gas vesicles (Hayes *et al.*, 1988). The amino acid sequences of these peptides showed similarity to the inferred sequence of the gene product of *gvpC*, a 489 bp open reading frame in *Calothrix* (Damerval *et al.*, 1987). This gene encoded a 162-residue 19 kDa protein containing four internal repeats of 33 amino acid residues (Fig. 2). The function of this protein was unknown but Northern hybridization showed that the gene was co-transcribed with *gvpA* and *B*, and this suggested its involvement in gas vesicle formation.

By using a probe supplied by Tandeau de Marsac, a *gvpC* gene was isolated from *Anabaena* (Hayes *et al.*, 1988). The gene here was larger, 579 bp, coding for a protein of 193 residues containing five homologous 33-residue repeats flanked by sequences of 18 and 10 residues at the *N*- and *C*-termini, respectively (see Fig. 2).

It has now been demonstrated that the gene product of *gvpC* in *Anabaena* is the protein that forms the single mobile band seen when gas vesicles from this cyanobacterium are separated by electrophoresis on SDS/polyacrylamide gels. The evidence for this is as follows.

(1) The band's mobility indicated a molecular mass, 21 kDa (Hayes *et al.*, 1986) similar to that calculated for the *Anabaena gvpC* gene product, 21985.

(2) The *N*-terminal amino acid sequence of the 21 kDa protein eluted from gels (Fig. 3) corresponds with that predicted for the *gvpC* gene product (Hayes *et al.*, 1988).

(3) After cells of *Anabaena* had been uniformly labelled with ^{35}S all of the label in the gas vesicles was located in the mobile electrophoresis band (see Fig. 4). The ^{35}S label in the gas vesicles could be quantitatively accounted for

Fig. 2. Comparison of the predicted amino acid sequences for GVPc

Sequences are shown for GVPc from *Anabaena flos-aquae* CCAP 1403/13d (Hayes *et al.*, 1988) and *Calothrix* sp. PCC 7601 (Damerval *et al.*, 1987). The method of presentation emphasises the 33-amino-acid repeats (I–V in *Anabaena*, I–IV in *Calothrix*). Strict identities between repeats are boxed.

Fig. 3. *N*-Terminal amino acid sequences for 21 kDa proteins eluted from SDS/polyacrylamide gels

Proteins eluted from gels were sequenced using an Applied Biosystems 470B gas-phase sequencer. Strict identities are boxed. Small capitals denote residues the identity of which have been determined by nucleotide sequencing of the gene.

Fig. 4. Distribution of label in gas vesicles isolated from cells grown in the presence of [35 S]sulphate revealed by SDS gel electrophoresis

(a) Coomassie Blue stained gel; in addition to the mobile band, darkly staining material is also clearly visible in the well. (b) Autoradiograph of gel. The label is largely restricted to the mobile band, GVPc, with very little apparent in the well, GVPa. (Modified from Walsby & Hayes, 1988).

by the two methionine residues in GVPc (Walsby & Hayes, 1988, and see below). GVPa, whose sequence indicates no sulphur-containing amino acids, was unlabelled.

(4) The identity of the mobile band as GVPc is also confirmed immunologically. A synthetic peptide with the sequence AQA EKQAQEL, which occurs three times (four without the first A; see Fig. 2) in the GVPc sequence, was linked via a *C*-terminal cysteinyl residue to PPD (purified protein derivative of tuberculin) and used to raise polyclonal antibodies in rabbit. The antibodies used to probe a Western blot of proteins separated by SDS/polyacrylamide-gel electrophoresis recognized only the mobile band. Other polyclonal antibodies raised against the sequence AVEKTNSSSLAE, the *N*-terminal sequence of GVPa (see Fig. 1), stained only the material left in the well of the gel, and a trace that had migrated to the base of the stacking gel (see Fig. 5).

It is not known why the 7.4 kDa GVPa fails to separate; just below the band at the base of the stacking gel a series of very faint bands can be detected on some gels, suggesting a polymeric series. There is no direct evidence that GVPa molecules are covalently linked together. The

Fig. 5. Localization of GVPa and GVPc in SDS/polyacrylamide gels revealed by Western blotting

The Coomassie Blue stained gel is shown in the central panel (b). (a) Blot probed with anti-GVPa; (c) blot probed with anti-GVPc. Bound antibodies were visualized using a protein A-horse-radish peroxidase conjugate. See the text for details of antibodies.

fact that GVPa can be sequenced out to residue 66 or even 69 out of 70 (*Microcystis* sequence in Hayes *et al.*, 1986) seems anyway to preclude this. Phosphate, which might have formed bridges between the hydroxyamino acids serine or threonine, is absent from gas vesicles (Walsby & Hayes, 1988).

GVPc in other organisms

The gas vesicles of the unicellular cyanobacterium *Microcystis* sp. also contain a protein of similar mobility to that of the *Anabaena* protein. Its *N*-terminal sequence is similar to that of the *Anabaena* GVPc, but it lacks the methionine residues (Fig. 3) both at the *N*-terminus and at position 5 (Hayes *et al.*, 1988). 35 S-labelling indicates no other sulphur amino acids.

Electrophoresis of *Calothrix* gas vesicles reveals a 19 kDa band (A. Bees, P. K. Hayes & A. E. Walsby, unpublished work), which corresponds to the molar ratio calculated from the deduced amino acid sequence (Damerval *et al.*, 1987). Electrophoretic analysis of gas vesicles from other cyanobacteria reveals the presence of mobile bands indicating a range of apparent molecular size, 14 kDa in *Oscillatoria agardhii*, 21 kDa in *Aphanizomenon*, and two proteins in *Dactylococcopsis salina*, one of 35 kDa (Hayes *et al.*, 1986) and one of 17 kDa (A. Griffiths, unpublished work). Konopka *et al.* (1977) reported a 55 kDa protein in the gas vesicles of the

Table 1. The proportion of gas vesicle protein formed by GVPc, calculated from measurements of amino acids present exclusively either in GVPa or GVPc

Amino acid	Number of residues per molecule of		GVPc/(GVPa + GVPc) from analysis		Reference
	GVPa	GVPc	Mass %	Mol %	
Proline	1	0	7.8	2.8	Hayes <i>et al.</i> (1988)
Tryptophan	1	0			
Phenylalanine	0	12	8.2	2.9	Hayes <i>et al.</i> (1988)
Histidine	0	3			
Methionine	0	2	8.9	3.3	Walsby & Hayes (1988)
Mean value			8.3	3.0	

heterotrophic bacterium *Ancylobacter* (*Microcyclops*) *aquaticus*. It is possible that the different molar ratios are explained by different numbers of the 33-amino acid repeats. Hayes *et al.* (1988) point out that there is a positive correlation between the size of these proteins and the diameter of the gas vesicles from which they are extracted. Using the *Calothrix gvpC* as a probe Damerval *et al.* (1989) detected *gvpC* by Southern hybridization in 12 strains of cyanobacteria all of which make gas vesicles abundantly. They failed to detect *gvpC* in a further 14 strains that were said to make less abundant gas vesicles, and they conclude that GVPc may not be an essential component of the gas vesicle. Another possibility is that in these strains a functionally similar protein is produced from a gene that shows insufficient similarity to the *Calothrix gvpC* to allow detection under the conditions used. The presence or absence of such proteins can only be determined by direct electrophoretic analysis of the isolated gas vesicles.

In *Halobacterium halobium* Damerval *et al.* (1988) reported a gene, located downstream from *gvpA*, whose protein product has certain features in common with GVPc. The deduced amino acid sequence indicates a periodic structure, with seven internally-homologous repeats of 32–39 residues, a high phenylalanine content and a hydrophilic nature. It shows no sequence similarity to the cyanobacterial GVPc and has apparently arisen by convergent evolution. The possibility that this is the protein which migrates on phenol/urea/acetic acid electrophoresis gels should be investigated by *N*-terminal sequencing.

The proportion of GVPc in *Anabaena* gas vesicles

Comparison of the sequences of GVPa and GVPc in *Anabaena* (Figs. 1 and 2) revealed that there were two amino acids that occurred exclusively in GVPa and another three that were exclusive to GVPc (see Table 1). Measurements of the amounts of these amino acids in preparations of entire gas vesicles provided a means of calculating the proportion of the total which each of the GVPs formed. For example, analysis of hydrolysates showed that phenylalanine accounted for 0.5 mol % of the total amino acid residues. Since there is no Phe in GVPa, this was accounted for by the 12 residues of Phe per molecule of GVPc, and the contribution of the remaining 181 residues of GVPc could therefore be determined. From this it was calculated that GVPc formed 2.9 mol % of the total gas vesicle protein, which

is equivalent to 8.2 % of the mass (Hayes *et al.*, 1988). A similar calculation was made from the analysis of proline, which occurs only in GVPa (Table 1). The proportion of GVPc was also calculated by measuring the content of methionine (two residues per molecule of GVPc) by a ³⁵S-labelling technique (Walsby & Hayes, 1988). The average GVPc content indicated by the three methods was 3.0 mol % (8.3 % by mass; see Table 1).

THE ROLE OF GVPa AND GVPc IN THE GAS VESICLE

When gas vesicles of *Anabaena* and *Microcystis* are rinsed with solutions of SDS at concentrations of 1 % (w/v) or greater, GVPc is removed but the gas vesicles remain intact. ³⁵S-labelling experiments indicate that at least 94 % of the GVPc is removed from the *Anabaena* gas vesicles and no further GVPc can subsequently be detected on the gas vesicles by electrophoresis. Electron microscopy reveals no obvious changes in morphology after the removal of GVPc. The conical end caps remain attached to the cylinders and the 4.6 nm ribs can be seen throughout the structure. It is concluded, therefore, that GVPc does not form the caps but must be located at the outer surface where it can be removed by the detergent without otherwise disturbing the structure (Walsby & Hayes, 1988).

Clearly the main structure of the ribs is formed by GVPa (and/or GVPd where this exists). As Hayes *et al.* (1986) point out, the size of this 70-residue protein, 7397 Da, corresponds closely to the size of the crystallographic unit cell that is shown by X-ray diffraction to repeat along the ribs of the structure (Blaurock & Walsby, 1976). The dimensions of this cell are defined by the rib width (4.57 nm), the periodicity along the rib (1.15 nm) and the mean wall thickness (1.8 nm; see Hayes *et al.*, 1986). The volume is then 9.46 nm³ and this will accommodate a protein molecule (density 1.32 g·cm⁻³) of 7.5 kDa, very close to the actual value of GVPa.

X-ray crystallography indicates that β -sheet is a prominent feature of the ribs. The β -chains run at an angle of about 35° to the principal axis of the gas vesicle, i.e. 55° to the axis of the rib. There are two layers of β -sheet within the thickness of the wall and four β -chains, two in each layer, within each unit cell (Blaurock & Walsby, 1976). Probably the adjacent β -chains within

Fig. 6. The possible arrangement of GVPa molecules in the ribs of the gas vesicle

(a) A gas vesicle (diameter 84 nm), showing the 4.6 nm-wide ribs. (b) The possible arrangement of parts of two adjacent GVPa molecules, which form the ribs. The β -chains shown by X-ray diffraction to lie at $\sim 55^\circ$ to the rib axis (Blaurock & Walsby, 1976) can be accounted for by a structure with alternating antiparallel chains of eight and seven dipeptides (solid lines) joined by β -turns, which would generate an angle of 54° to the rib. Hydrogen bonds (broken lines) would lie at the complementary angle of 36° (E. D. T. Atkins & A. E. Walsby, unpublished work). (c) Details of the β -turn showing the dipeptide repeat ($2h = 0.69$ nm) and interchain distance ($a = 0.472$ nm) used to calculate the chain angle. (d) Projections of two adjacent GVPa molecules; within the thickness of the wall there are two layers of β -sheet which may run either in the same sense (shown here) or in opposite senses (as shown in e). The molecules produce a periodicity of 1.15 nm along the rib. It is emphasized that a number of details are unproven: it is not known how the two layers of β -sheet are connected; the arrangement shown would accommodate only 60 of the 70 amino acid residues per GVPa molecule. Drawing reproduced with permission from Walsby (1989).

each layer run antiparallel, in which case they would be joined by a β -turn (see Fig. 6).

The angle between the β -chains and the rib axis must be determined by the way that polypeptide chain folds. A possible structure is that of a β -chain of eight dipeptides leading, via a β -turn, to an antiparallel chain of seven dipeptides, with the ends of the chains in horizontal register at the base of the rib. The chains would then slope to one side (see Fig. 6b) making an angle $\alpha = \tan^{-1}(2h/2a)$ to the rib axis; $2h$ is the dipeptide repeat (0.69 nm) and a is the perpendicular distance between chains (0.472 nm), so that $\alpha = 53.8^\circ$ (E. D. T. Atkins & A. E. Walsby; see Walsby, 1987). In such a structure the runs of β -sheet and positions of the β -turns would be

determined by the amino acid sequence. High resolution crystallographic studies are required to show the actual folding arrangement.

The cylindrical part of the gas vesicle is probably formed from two identical halves joined back-to-back at the centre (Walsby, 1972), where a distinctive junction can be seen (Waaland & Branton, 1969). Within each half the interactions between the GVPa molecules of all the ribs are probably uniform. In the cones, however, the molecules must make a different sort of interaction with a sloping, rather than horizontal, contact between adjacent ribs (Walsby, 1978). It is possible, then, that the cones may be formed from a modified form of GVPa, or from GVPd with its slightly different amino acid sequence

(Damerval *et al.*, 1989). This could be tested by isolating the juvenile biconical stages of gas vesicles and analysing their amino acid sequence directly. Konopka *et al.* (1975) reported a mutant of *Ancylobacter* (*Microcystis*) that made only the biconical stage; unfortunately it has proved difficult to isolate gas vesicles from this organism. Irrespective of which protein forms the caps, the protein must have some considerable flexibility because the ribs have a progressively smaller diameter as they approach the cone apex.

Since GVPa is known to form the essential ribbed core of the structure and GVPc is located somewhere on the outside, it is worth asking how these two proteins might interact. The most striking feature of GVPc is its 33-residue repeats, five in *Anabaena* and four in *Calothrix*. We have considered the possibility that the GVPc polypeptide might act as a tie, crossing the ribs and holding them together (Hayes *et al.*, 1988). Similar sequences in each of the 33-amino acid repeats might then make similar interactions with GVPa molecules in the successive ribs that are crossed. Secondary structure predictions suggest that much of each repeat is in the α -helical conformation. The length of a 33-amino acid α -helix, 4.95 nm, would be sufficient to cross a 4.6 nm rib.

PHYSICAL PROPERTIES OF THE GAS VESICLE PROVIDED BY THE PROTEINS

As stated in our Introduction, the properties of the gas vesicle must be provided by the constituent proteins. We are now in a position to relate these properties to what we know of the protein structure.

Rigidity and strength

There are two features of GVPa that are relevant to the stability of the structure under pressure. The first is that the molecules stack to form ribs that provide ring-stiffening support to the structure, of the sort which is engineered into pressure vessels. The second is the angle formed by the β -chains. The hydrogen bonds that link the β -chains together run at about 55° to the long axis of the cylinder. This is very close to the ideal angle for fibre-reinforcing a pressure vessel, $\tan^{-1}\sqrt{2}$ or 54.7° . In fibres at this angle the transverse and longitudinal stresses in the cylinder wall will be equalized (see Walsby, 1987).

There is evidence that GVPc adds to the strength of the structure, perhaps by tying the ribs together in the way proposed. When GVPc is removed from intact gas vesicles of *Microcystis* or *Anabaena* by treatment with SDS, their critical collapse pressures fall by a factor of three or more (Walsby & Hayes, 1988). There is a possibility that the weakening is not directly caused by the removal of GVPc itself but by some other change wrought by SDS on the remaining GVPa structure, but this seems unlikely as that structure is not demonstrably changed. We have not, so far, been able to reconstitute entire gas vesicles by adding GVPc back to the GVPa shells.

A survey of gas vesicles isolated from eight species of cyanobacteria has shown that there is an inverse correlation between critical collapse pressure and cylinder diameter as expected for rigid, thin-walled vessels (Hayes & Walsby, 1986; Walsby & Bleything, 1988). The wide gas vesicles provide buoyancy with more efficiency and are found in organisms occupying habitats where the pressures generated by cell turgor pressure and the depth

Fig. 7. Hydropathic profiles of (a) GVPa and (b) GVPc from *Anabaena*.

The profiles are calculated by the method of Kyte & Doolittle (1982) using a seven-residue window.

of the water column are less. The width of the gas vesicle, which determines its suitability for different habitats, must be determined by features of the constituent proteins.

The hydrophilic outer surface and hydrophobic inner surface

Hydropathy plots show that GVPa is predominantly hydrophobic whereas GVPc is strongly hydrophilic (see Fig. 7). These properties of the two proteins could account for the observed contrasting properties of the inner and outer surface, and they probably explain why GVPc can be readily dissolved while GVPa cannot.

The outer surface of the gas vesicle remains hydrophilic after the GVPc is removed by 1% SDS; the treated gas vesicles continue to partition into the aqueous phase from a water/oil emulsion, even after passage through an SDS affinity column (A. E. Walsby, unpublished work). Unless this behaviour is explained by SDS remaining bound to the structure, this observation suggests that the GVPa molecule also presents a hydrophilic sequence, such as that which occurs at the *N*-terminal end, at the outer surface.

The hydrophobic inner surface of the gas vesicle can certainly be explained by the highly hydrophobic sequences in other parts of the GVPa molecule. It is likely also that the two layers of β -sheet within the gas vesicle wall are held together by hydrophobic bonding and so, in a sense, it might be expected that, of the four surfaces, three — the inner gas facing surface, and the two apposed surfaces within the wall — are hydrophobic.

Self assembly and turnover

In some unknown manner the protein molecules form a hollow shell. It is possible that other molecules may be involved as chaperons or templates in this assembly process. Damerval *et al.* (1987) have speculated that GVPc, with its 33-residue repeating sequences, might act in this way. It is likely, however, that a template molecule would be recycled, unless it were needed subsequently to maintain the structure. Although GVPc is only a minor component, 8% of the total GVP mass, gas vesicles themselves may account for 10% of the cell protein (see Hayes & Walsby, 1984; Oliver & Walsby, 1984; Thomas & Walsby, 1985), so that GVPc may represent 1% of the total protein, an amount which might otherwise be profitably saved.

Radioactive labelling studies with ^{14}C suggest that, after collapse, gas vesicle proteins are recycled into new gas vesicles, and that this increases the rate of gas vesicle production (Hayes & Walsby, 1984). Gas vesicles can be stored for years at 5 °C in an intact state and GVPa appears to be very stable. It is noted that the N- and C-terminal residues are both alanine, one of the amino acids that provides maximum resistance to attack by exopeptidases that recycle short-lived proteins in all types of cell.

EPILOGUE: MOLECULAR GENETICS AND GAS VESICLE RESEARCH

The isolation of the gas vesicle genes, pioneered by Tandeau de Marsac and her colleagues at the Institute Pasteur, has permitted major advances in the understanding of the gas vesicle proteins. It resolved uncertainties in the sequence of the main protein GVPa and pointed to the existence of other protein components. Without the *gvpC* gene sequence it is hard to see how the amino acid sequence of GVPc, with its five-fold repeats in *Anabaena*, could ever have been solved by the conventional techniques relying on overlapping peptide sequences.

The discovery of tandem repeats of the *gvpA* gene (Damerval *et al.*, 1987) provided direct evidence for two (or more) cistrons. This had been suspected previously in the heterotrophic bacterium *Prosthecomicrobium pneumaticum* (Walsby, 1976) and from *Anabaena flos-aquae* (Walsby, 1977), on the evidence that both gave mutants that produced about half the gas vesicle content of the wild type, insufficient to provide buoyancy. Apparently, transcription from two cistrons is needed to generate GVPa in sufficient abundance. In *Calothrix*, *gvpA1* and *A2* differ in 18 nucleotide bases, but all of them are 'silent' giving rise to no changes in amino acid sequence of the protein. There must be a constraint on the independent evolution of the two genes. Perhaps even single amino acid differences between the two proteins would result in an unevenness of structure in the gas vesicle, weakening it.

With knowledge of the amino acid sequences of the gas vesicle proteins the foundation has now been laid for an understanding of structure. There are now two important goals to be achieved. The first is to obtain expression of gas vesicle proteins and assembly of gas vesicles in a foreign host such as *Escherichia coli*. In this way all of the factors required for gas vesicle production can be defined and, by mutating specific sites in the

sequences of the proteins, the factors that determine gas vesicle morphology can be determined. The second goal is to determine the three-dimensional structure of the gas vesicle proteins by X-ray crystallography.

Part of this work was supported by grants from the SERC. We thank Mr. Tim Colborn for drawing the figures.

REFERENCES

- Archer, D. B. & King, N. R. (1984) *J. Gen. Microbiol.* **130**, 167–172
- Armstrong, R. E. & Walsby, A. E. (1981) in *Organization of Prokaryotic Cell Membranes, II* (Ghosh, B. K., ed.), pp. 95–129, CRC Press, Cleveland, OH
- Blaurock, A. E. & Walsby, A. E. (1976) *J. Mol. Biol.* **105**, 183–199
- Bowen, C. C. & Jensen, T. E. (1965) *Science* **147**, 1460–1462
- Buckland, B. (1971) Ph.D. Thesis, University of London, England
- Damerval, T., Houmard, J., Guglielmi, G., Csiszár, K. & Tandeau de Marsac, N. (1987) *Gene* **54**, 83–92
- Damerval, T., Guglielmi, G., Castets, A. M., DasSarma, S., Houmard, J. & Tandeau de Marsac, N. (1988) in *VI International Symposium on Photosynthetic Prokaryotes* (Mur, L. R. & Burger-Wiersma, T., eds.), p. 266, Nordwijkerhout
- Damerval, T., Castets, A. M., Guglielmi, G., Houmard, J. & Tandeau de Marsac, N. (1989) *J. Bacteriol.* **171**, in the press
- DasSarma, S., Damerval, T., Jones, J. G. & Tandeau de Marsac, N. (1987) *Mol. Microbiol.* **1**, 365–370
- Falkenberg, P. (1974) Thesis, Institutt for Teknisk Biokjemi Norges, Tekniske Hogskole, Universitetet i Trondheim
- Falkenberg, P., Buckland, B. & Walsby, A. E. (1972) *Arch. Mikrobiol.* **85**, 304–309
- Hayes, P. K. (1988) *Methods Enzymol.* **167**, 213–222
- Hayes, P. K. & Walsby, A. E. (1984) *J. Gen. Microbiol.* **130**, 1591–1596
- Hayes, P. K. & Walsby, A. E. (1986) *Br. Phycol. J.* **21**, 191–197
- Hayes, P. K. & Walsby, A. E. (1988) in *VI International Symposium on Photosynthetic Prokaryotes* (Mur, L. R. & Burger-Wiersma, T., eds.), p. 267, Nordwijkerhout
- Hayes, P. K., Walsby, A. E. & Walker, J. E. (1986) *Biochem. J.* **236**, 31–36
- Hayes, P. K., Lazarus, C. M., Bees, A., Walker, J. E. & Walsby, A. E. (1988) *Mol. Microbiol.* **2**, 545–552
- Horne, M., Englert, C. & Pfeifer, F. (1988) *Mol. Gen. Genet.* **213**, 459–464
- Jones, D. D. & Jost, M. (1970) *Arch. Mikrobiol.* **70**, 43–64
- Jones, D. D. & Jost, M. (1971) *Planta* **100**, 277–287
- Jost, M. (1965) *Arch. Mikrobiol.* **50**, 211–245
- Konopka, A. E., Staley, J. T. & Lara, J. C. (1975) *J. Bacteriol.* **122**, 1301–1309
- Konopka, A. E., Lara, J. C. & Staley, J. T. (1977) *Arch. Mikrobiol.* **112**, 133–140
- Krantz, M. J. & Ballou, C. E. (1973) *J. Bacteriol.* **114**, 1058–1067
- Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132
- Larsen, H., Omang, S. & Steensland, H. (1967) *Arch. Mikrobiol.* **59**, 197–203
- Lehmann, H. & Jost, M. (1971) *Arch. Mikrobiol.* **79**, 59–68
- Oosterhelt, D., Kerscher, L., Mathews, S. & Lottspeich, F. (1984) in *Proceedings of the Symposium on Frontiers in Bioorganic Chemistry and Molecular Biology*, Moscow–Alma Ata, Elsevier, Amsterdam
- Oliver, R. L. & Walsby, A. E. (1984) *Limnol. Oceanogr.* **29**, 879–886

- Pfeifer, F., Blaseio, U. & Horne, M. (1989) *Can. J. Microbiol.* **35**, in the press
- Simon, R. D. (1978) *Nature (London)* **273**, 314–317
- Simon, R. D. (1981) *J. Gen. Microbiol.* **125**, 103–111
- Stoeckenius, W. & Kunau, W. H. (1968) *J. Cell Biol.* **38**, 336–357
- Surek, B. (1987) Ph.D. Thesis, University of Würzburg, Germany
- Surek, B., Pillay, B., Rdest, U., Beyreuther, K. & Goebel, W. (1988) *J. Bacteriol.* **70**, 1746–1751
- Tandeau de Marsac, N., Mazel, D., Bryant, D. A. & Houmard, J. (1985) *Nucleic Acids Res.* **13**, 7223–7236
- Thomas, R. H. & Walsby, A. E. (1985) *J. Gen. Microbiol.* **131**, 799–809
- Waaland, J. R. & Branton, D. (1969) *Science* **163**, 1339–1341
- Walker, J. E. & Walsby, A. E. (1983) *Biochem. J.* **209**, 809–815
- Walker, J. E., Hayes, P. K. & Walsby, A. E. (1984) *J. Gen. Microbiol.* **130**, 2709–2715
- Walsby, A. E. (1969) *Proc. R. Soc. London Ser. B* **173**, 235–255
- Walsby, A. E. (1971) *Proc. R. Soc. London Ser. B* **178**, 301–326
- Walsby, A. E. (1972) *Bacteriol. Rev.* **26**, 1–32
- Walsby, A. E. (1974) *Methods Enzymol.* **31A**, 678–686
- Walsby, A. E. (1976) *Arch. Microbiol.* **109**, 135–142
- Walsby, A. E. (1977) *Arch. Microbiol.* **114**, 167–170
- Walsby, A. E. (1978) in *Relations between Structure and Function in the Prokaryotic Cell* (Stanier, R. Y., Rogers, H. J. & Ward, J. B., eds.), pp. 327–358, Cambridge University Press
- Walsby, A. E. (1981*a*) in *The Prokaryotes* (Starr, M., Stolp, H., Trüper, H., Balows, A. & Schlegel, H. G., eds.), pp. 224–235, Springer-Verlag, New York
- Walsby, A. E. (1981*b*) in *The Prokaryotes* (Starr, M., Stolp, H., Trüper, H., Balows, A. & Schlegel, H. G., eds.), pp. 441–447, Springer-Verlag, New York
- Walsby, A. E. (1982) *Proc. R. Soc. London Ser. B* **216**, 355–368
- Walsby, A. E. (1984) *Proc. R. Soc. London Ser. B* **223**, 177–196
- Walsby, A. E. (1987) in *The Cyanobacteria* (Fay, P. & Van Baalen, C., eds.), pp. 377–392, Elsevier, Amsterdam
- Walsby, A. E. (1989) in *International Symposium on Supersaturation and Bubble Formation in Fluids and Organisms* (Brubakk, A. & Sundnes, G., eds.), Royal Norwegian Society of Sciences and Letters, Trondheim, in the press
- Walsby, A. E. & Armstrong, R. E. (1979) *J. Mol. Biol.* **129**, 279–285
- Walsby, A. E. & Bleything, A. (1988) *J. Gen. Microbiol.* **134**, 2635–2647
- Walsby, A. E. & Buckland, B. (1969) *Nature (London)* **224**, 716–717
- Walsby, A. E. & Hayes, P. K. (1988) *J. Gen. Microbiol.* **134**, 2647–2657
- Weathers, P. J., Jost, M. & Lamport, D. T. A. (1977) *Arch. Biochem. Biophys.* **178**, 226–244
- Weidinger, G., Klotz, G. & Goebel, W. (1979) *Plasmid* **2**, 377–386
- Worcester, D. L. (1975) *Brookhaven Symp. Biol.* **27**, III 37–57