Complete amino acid sequence of cyanobacterial gas-vesicle protein indicates a 70-residue molecule that corresponds in size to the crystallographic unit cell

Paul K. HAYES,* Anthony E. WALSBY* and John E. WALKER†

*Department of Botany, University of Bristol, Bristol BS8 1UG, and †M.R.C. Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, U.K.

Gas vesicles of cyanobacteria are formed by a protein called 'gas-vesicle protein' (GVP). The complete amino acid sequence has been determined of GVP from *Anabaena flos-aquae*. It is 70 residues long and has an M_r of 7388. This corresponds to the size of the repeating unit cell demonstrated by X-ray crystallography of intact gas vesicles. Details of the sequence are related to the secondary β -sheet structure of the protein and its contrasting hydrophilic and hydrophobic surfaces. Extensive amino acid sequences have also been determined for GVPs from two other cyanobacteria, species of *Calothrix* and *Microcystis*; they are highly homologous with that of *Anabaena* GVP. Electrophoretic analysis indicates that GVPs of different cyanobacteria form a variety of stable oligomers.

INTRODUCTION

Planktonic cyanobacteria derive their buoyancy from hollow cylindrical structures called gas vesicles, which are formed from a single species of gas-vesicle protein (GVP). The protein is arranged in a crystalline fashion along ribs oriented normal to the long axis of the cylinder, and it appears that the structure is self-assembled, as are virus coats [see Walsby (1972, 1978) for reviews]. We describe here the complete amino acid sequence of GVP from *Anabaena flos-aquae*, and further partial sequences of GVPs from other cyanobacteria. The *Anabaena* GVP is now shown to be only 70 residues long. This conclusion is supported by sequence analysis of sets of both peptic and tryptic digests, giving an entire overlapping sequence whose amino acid composition corresponds with that obtained by direct analysis. The 70-residue protein, which corresponds in size with the repeating unit cell demonstrated by X-ray crystallography of *Anabaena* gas vesicles (Blaurock & Walsby, 1976), is shorter than claimed previously on the basis of SDS/polyacrylamide-gel electrophoresis (Walker & Walsby, 1983).

Previous uncertainties concerning the M_r of GVP have frustrated attempts to reconcile the molecular and crystallographic structure of the gas vesicle [see review by Armstrong & Walsby (1981)]. These uncertainties stemmed from the inability to solubilize GVP in SDS, and they appeared to be resolved when conditions were found for dissolving the protein, which ran as a single mobile band of estimated M_r 21 000 on SDS/polyacrylamide electrophoretic gels (Walker & Walsby, 1983). Comparative investigations with gas vesicles of a number of cyanobacteria, which have highly homologous GVPs (Walker *et al.*, 1984), now indicate that the anomalous behaviour on SDS/polyacrylamide gels is probably a result of persistence of multimers of the GVP polypeptide.

EXPERIMENTAL

Isolation of gas vesicles

Gas vesicles were isolated from *Anabaena flos-aquae* (C.C.A.P. strain 1403/13f) as described by Walsby & Armstrong (1979). Those of *Microcystis* sp. (BC 84/1; Thomas & Walsby, 1985) and *Calothrix* sp. (PCC 7601) were isolated by centrifugally accelerated flotation of intact gas vesicles from cells suspended in 50 mM-Tris/HCl (pH 8.0)/100 mM-EDTA and lysed with lysozyme (0.5 mg·ml⁻¹) (Sigma). Unlysed cells were removed from the preparations of gas vesicles by filtration (Walker *et al.*, 1984).

Gel electrophoresis

Gel electrophoresis was carried out by using vertical slab gels prepared by the method of Laemmli (1970). Lysozyme (M_r 14 300; Sigma) was used in addition to the M_r markers described by Walker & Walsby (1983).

Amino acid and sequence analysis

Quantitative amino acid analyses with a Durrum D-500 amino acid analyser and automated amino acid sequencing on a Beckman spinning-cup sequencer were performed as described by Walker & Walsby (1983). The sequence of *Microcystis* GVP was determined with the aid of an Applied Biosystems 470B gas-phase sequencer, operated under standard conditions as recommended by the manufacturer. Polybrene (2 mg) (Tarr *et al.*, 1978) was dried on to the glass-filter disc and precycled for five cycles. Then GVP was applied as an aqueous suspension. Amino acid phenylthiohydantoin derivatives were identified by h.p.l.c.

Digestion with trypsin

Intact gas vesicles were freeze-dried and then dissolved in 99% (v/v) formic acid. This solution was dialysed at

4 °C against 2 × 5 litres of distilled water and finally against 5 litres of 2 M-urea (deionized with Amberlite MB1 and MB3 resins)/50 mM-NH₄HCO₃. Trypsin [tosylphenylalanylchloromethane ('TPCK')-treated; Sigma; 1 mg · ml⁻¹ in 1 mM-HCl] was added to the clear GVP solution in two aliquots (second addition after 4 h incubation) to give a final substrate/enzyme ratio of 40:1 (w/w). The digestion was carried out at 37 °C for 24 h. The reaction products were fractionated on a column (100 cm × 2 cm) of Sephadex G-50 (fine grade) in 50 mM-NH₄HCO₃ (Fig. 1). Three pools of fractions were used for subsequent sequence analysis. Only pool T1 could be used without further purification; pools T2 and T3 were subject to fractionation by h.p.l.c. using a Nucleosil C-18 reverse-phase column [25 cm × 0.7 cm (int. diam.), manufactured by Macherey, Nagel and Co. and supplied by Jones Chromatography] equilibrated with 0.1% trifluoroacetic acid and eluted with a linear gradient of acetonitrile. Pool T2 yielded a large number of peaks of which three peptides, T2-1, T2-2 and T2-3, were eluted at 33, 34 and 39% acetonitrile respectively; pool T3 gave a peptide that was eluted at 14% acetonitrile.

Tryptic digestion of 'succinylated' (3-carboxypropionylated) GVP

The amino groups of formic acid-denatured GVP were allowed to react with a 50-fold molar excess of succinic anhydride as described by Walker *et al.* (1980). The dialysed reaction product in 0.5% (w/v) NH₄HCO₃ was digested at 30 °C for 16 h with trypsin (as above) at a substrate/enzyme ratio of 40:1 (w/w). Digestion products were fractionated by passage through a column (100 cm × 2 cm) of Sephadex G-50 (fine grade) in 50 mM-NH₄HCO₃ (Fig. 2). Two sample pools were collected and used for sequence analysis without further purification.

Digestion with pepsin

Freeze-dried gas vesicles were dissolved in 99% formic acid and then vigorously mixed with 2 × 20 vol. aliquots of a pepsin (Sigma; from hog stomach mucosa) stock solution (12.5 µg · ml⁻¹) in 1 mM-HCl. The final substrate/enzyme ratio was 50:1 (w/w). Digestion was for

Fig. 1. Fractionation of a tryptic digest of GVP (950 nmol) on Sephadex G-50

The column (100 cm × 2 cm) was eluted with 50 mM-NH₄HCO₃. The pooled fractions containing peptides used in sequence analysis are indicated by bars.

Fig. 2. Fractionation of a tryptic digest of 'succinylated' GVP (310 nmol) on Sephadex G-50

See the legend to Fig. 1 for other details.

24 h at room temperature. Saturated NH₄HCO₃ was added to give a final pH of 7.5 and the soluble reaction products were then fractionated by passage through a column (100 cm × 2 cm) of Sephadex G-50 (fine grade) in 50 mM-NH₄HCO₃ (Fig. 3). Three pools of fractions were collected and subjected to further fractionation by h.p.l.c. (see above). H.p.l.c. fractionation of each of these pools produced complex elution profiles; the various peaks were collected, subjected to amino acid analysis and, where appropriate, sequence analysis.

Digestions with carboxypeptidase Y

Aliquots of the C-terminal tryptic peptide (see the Results section) were vacuum-dried and then redissolved in 50 mM-Mes (pH 6.5; BDH)/1 mM-EDTA. Carboxypeptidase Y (in water; Carl-Biotech) was added to a final substrate/enzyme ratio of 40:1 (w/w) and the digestion allowed to proceed at room temperature. Samples were periodically removed from the digestion mixture and any free amino acids present quantified.

Digestion with post-proline cleaving peptidase

Dry samples of the C-terminal peptide (see the Results section) were dissolved in 200 mM-sodium phosphate, pH 7.0. Post-proline cleaving peptidase from *Flavobacterium meningosepticum* (Seikukagu Kogyo Co., Tokyo, Japan) was then added to give a final substrate/enzyme ratio of 167:1 (w/w) and the digestion

Fig. 3. Fractionation of a peptic digest of GVP (270 nmol) on Sephadex G-50

See the legend to Fig. 1 for other details.

allowed to proceed at 37 °C for 16 h. Separation of the reaction mixture by h.p.l.c. (see above) allowed the identification of six components eluted at 0, 17, 18, 26.5, 29.5 and 33% acetonitrile. The reaction products were subject to quantitative amino acid analysis and, where appropriate, their *N*-terminal residues were identified by dansylation (Allen, 1981).

RESULTS

Complete amino acid sequence of GVP from *Anabaena flos-aquae*

The amino acid sequences of the peptides described in the previous section are given in Table 1. This sequence information can be used to extend the sequence of Walker & Walsby (1983) out to residue 66 (peptide P1-1 should be aligned with the *C*-terminal portion of the published sequence).

The tryptic peptide T2-1 was found to consist of 15 amino acid residues with no lysine or arginine and may therefore be taken to represent the *C*-terminal peptide of the complete protein. Only the first eight residues could be placed in unambiguous sequence using the spinning-cup sequencer. However, alignment with peptide P1-1 (Table 1) allows this sequence to be extended to residue 11 (YAEAVGLTQSA). Digestion of the tryptic peptide with carboxypeptidase Y revealed the presence of an alanine residue at the *C*-terminus; no other information could be dependably gleaned using this approach. This left the known sequence as YAEAVGLTQSA...A with P, A and V still to be placed in sequence. Incubation of the peptide with the post-proline cleaving peptidase gave six products that could be separated by h.p.l.c. Three of these could be recognized from their amino acid compositions as parts of the existing sequence and a fourth as being the undigested peptide. Of the two remaining products, one was found to consist of equimolar amounts of proline and valine and the other to contain only alanine. Dansylation of the first of these products followed by acid hydrolysis showed that only dansylvaline was present; therefore the sequence of this dipeptide must be VP. Dansylation of the alanine-containing fraction gave dansylalanine and

dansylalanylalanine before acid hydrolysis, but only dansylalanine after hydrolysis. Taken in conjunction with the previously obtained data, the complete amino acid sequence of this peptide must be:

with the post-proline cleaving enzyme giving complete cleavage at the sites marked by * and partial cleavage at those sites marked by °. Yoshimoto *et al.* (1980) comment that this enzyme gives complete cleavage after proline and cleaves more slowly after alanine residues. When this *C*-terminal peptide is overlapped with the other peptides in Table 1 and the sequence published by Walker & Walsby (1983), the complete amino acid sequence of the GVP from *Anabaena flos-aquae* is obtained (Fig. 4).

The molar yields of the peptides used in the sequence analysis were generally low, except at the *N*-terminal end of the protein. In the tryptic digest the amount of the peptide AVEK (T3, Table 1) recovered represented 29% of the total calculated to be present in the initial digest, assuming that the M_r of GVP was 7388 (see below). The corresponding yield was 34% for the peptic digest, with this *N*-terminal sequence represented in five different fractions. The recoveries for the peptides T2-1, T2-2 and T2-3 were 6.7, 6.2 and 1.8% respectively. It is noted that the yield declined with distance from the *N*- or *C*-terminus of the protein.

In addition to the peptides given in Table 1, another 21 products of the peptic digest of GVP were purified. Each one could be aligned with a stretch of the complete sequence (Fig. 4), either by sequencing or from amino acid composition. Between them these peptic products covered the entire length of the GVP molecule.

Partial amino acid sequence of GVP from *Calothrix* and *Microcystis*

Gas-vesicle formation was induced in *Calothrix* sp. (P.C.C. 7601) by incubation of the culture in red light for 1 day, conditions that induce differentiation of gas-vacuolate hormogonia. Analysis of the GVP in isolated gas vesicles, in the Beckman sequencer, yielded the continuous sequence from residue 1 to 57 shown in Fig. 5. The sequence differs from that of *Anabaena* GVP at residue 25, which was V instead of I, and at residue 37, which was E instead of Q (residue 26 was either D, as in *Anabaena*, or N).

Gas vesicles isolated from *Microcystis* sp. (BC 84/1) analysed in the gas-phase sequencer gave continuous sequence from residue 1 to 64; beyond that, the identities of three of the next five residues were uncertain, but the penultimate residue of the molecule was obtained, according to alignment with the *Anabaena* GVP sequence (Fig. 5). Only three of the 67 identified residues differed from those of *Anabaena* GVP.

Amino acid analysis of GVPs from *Anabaena* and *Microcystis*

In Table 2 the results of quantitative amino acid analysis of *Anabaena* GVP [from Table 1 of Walker & Walsby (1983)] are compared with the composition calculated from the amino acid sequence given in Fig. 4. It is seen that there is good correspondence between the two values in most cases. Valine and isoleucine are underestimated somewhat in the analysis, but this is to be expected because the sequence contains three regions (at residues 14–15, 23–25, and 45–47) where these amino

Table 1. Sequence of peptides isolated from proteolytic digests of GVP

Peptide R2 was not subject to sequence analysis, but its sequence could be deduced from its amino acid composition. Peptide T1 contained sequences corresponding to cleavage of the intact protein at the first lysine and arginine residues. Where the sequence of a peptide is incomplete, the amino acid composition of the unsequenced portion is given in parentheses.

Peptide	Sequence
R1	ILDKGIVIDAWVRVSLVGIQLLAIEARIVIASVE TYLK YAEA...
R2	AVEKTNSSSSLAVIDR
T1	TNSS... and ILDK...
T2-1	YAEAVGLT (S ₁ E ₁ P ₁ A ₃ V ₁)
T2-2	TNSSSSLAVIDR
T2-3	IVIASVET (L ₁ Y ₁ K ₁)
T3	AVEK
P1-1	VGLTQSA (P ₁ A ₂ V ₁)
P1-2	AEVIDRILDKG
P2	TYLK YAE (A ₁ V ₁)

Fig. 4. Complete amino acid sequence of GVP from *Anabaena flos-aquae*

The Figure shows the alignment of peptides produced by the proteolytic cleavage of GVP with the *N*-terminal sequence (S1 in the Figure) published by Walker & Walsby (1983).

	10	20	30	
(a)	AVEKTNSSSSLAVIDRILDKGIVDAWVRVSLVG			
(b)	AVEKTNSSSSLAVIDRILDKGIVDAWVRVSLVG			
(c)	AVEKTNSSSSLAVIDRILDKGIVDAWVRVSLVG			
	40	50	60	70
(a)	IQLLAIEARIVIASVETYLKYAEAVGLTQSAVPAA			
(b)	IQLLAIEARIVIASVETYLKYA-----			
(c)	IQLLAIEARVIVIASVETYLKYAEAVGLTQXAXXA-			

Fig. 5. Sequence homology between GVPs isolated from (a) *Anabaena flos-aquae*, (b) *Calothrix* sp. and (c) *Microcystis* sp.

'X' denotes the position of a residue of uncertain identity; - indicates that no data was available at that position.

acids are juxtaposed and would therefore resist acid hydrolysis. In Table 2 the amino acid analysis of *Microcystis* GVP is also compared with that calculated from the *Microcystis* GVP sequence given in Fig. 5, in which the unidentified residues 65, 67, 68 and 70 are assumed to be the same as those in *Anabaena* GVP. Again the correspondence between the two values is generally good, but with the same expected underestimate of isoleucine and valine; additionally the underestimate in serine and threonine can be explained by losses on hydrolysis, as in this sample no extrapolation was made to zero time. It is noted that, as previously, a trace of phenylalanine is recorded in the analyses for both GVPs; it is equivalent to less than a third of a residue in the 70 amino acid protein.

Electrophoretic analysis

Electrophoresis of *Anabaena* gas vesicles dissolved by heating at 100 °C for 2 min in 1% SDS gave a single band whose mobility suggested an M_r of 21 000 (see Fig. 6), as reported previously. Good results were consistently obtained with freshly isolated gas vesicles; samples of intact vesicles that had been stored at 5 °C for more than a few weeks failed to give detectable bands when run on the electrophoretic system. This suggested some change

occurred in the protein; the gas vesicles become weaker, but otherwise remain unchanged.

Gas vesicles of *Microcystis* solubilized and electrophoresed in an identical fashion gave three bands, a major one at a M_r 21 000, indistinguishable from *Anabaena* GVP, one at about 14 500 (just behind lysozyme marker) and a third running close to the dye front (Fig. 6). The fact that the sequence analysis of *Microcystis* GVP indicates an M_r of about 7400 suggests that the smallest of these three bands represented the GVP monomer, whereas the others represented the dimer and trimer. This also infers that the single band of M_r 21 000 produced by *Anabaena* GVP is a trimer. It has been noted previously that, after denaturing *Anabaena* GVP in 80% formic acid, electrophoresis indicates that an oligomeric series of bands forms (Walker & Walsby, 1983).

Preliminary investigations by electrophoresis with gas vesicles of other cyanobacteria gave a variety of results. Those of *Aphanizomenon* gave a single band of M_r 21 000, *Oscillatoria* a single band at M_r 14 000, *Dactylococcopsis* a single band at M_r 35 000, and *Calothrix* a major band at M_r 14 000, with a minor one at about M_r 17 000. It is noted that, with the exception of the last value, these are all multiples of 7000. This indicates that the GVP monomers have a tendency to remain firmly bound in small oligomers.

DISCUSSION

The completely overlapped sequence through to the C-terminus indicates that *Anabaena* GVP is a protein of 70 amino acid residues. This is confirmed by the correspondence between the deduced amino acid composition and that given by direct analysis. From the amino acid sequence the M_r is calculated to be 7388. The fact that *Microcystis* GVP gave sequence up to the corresponding penultimate residue of *Anabaena* GVP indicates that it is a molecule of similar size.

In the electrophoretic analysis of *Anabaena* gas vesicles, Walker & Walsby (1983) demonstrated that the band of M_r 21 000 had a similar composition to that of the total protein in the gas vesicle, and hence also that of the sequenced GVP. Since no band with an M_r in the region of 7400 was found, it is suggested that the GVP is present as a trimer.

The fact that only one *N*-terminal sequence was found

Table 2. Comparisons of the amino acid composition of GVPs derived from protein sequence data with those measured after acid hydrolysis of the protein

Tryptophan could not be measured in the acid hydrolysates and its presence was ignored when calculating the molar percentages. In the acid-hydrolysis data for GVP from *Anabaena flos-aquae*, the values for threonine and serine were extrapolated to zero time. The values for isoleucine and leucine were measured after 72 h hydrolysis (Walker & Walsby, 1983). When calculating the composition of GVP from *Microcystis* sp. using the sequence data, it was assumed that residues 65, 67, 68 and 70 had the same identity as those in the *Anabaena* GVP. Abbreviation used: n.d. not determined

Amino acid	Composition (mol/100 mol)			
	<i>Anabaena</i> GVP		<i>Microcystis</i> GVP	
	Measured in acid hydrolysate	Calculated from sequence	Measured in acid hydrolysate	Calculated from sequence
Asp	5.8	5.8	6.3	5.8
Thr	5.2	4.3	4.0	4.3
Ser	10.8	10.1	8.7	10.1
Glu	12.9	10.1	9.2	10.1
Pro	1.4	1.4	1.6	1.4
Gly	4.4	4.3	4.0	4.3
Ala	16.4	15.9	17.8	17.4
Cys	n.d.	0	n.d.	0
Val	11.8	14.5	12.0	14.5
Met	0	0	0	0
Ile	8.2	11.6	8.2	10.1
Leu	10.7	10.1	11.1	10.1
Tyr	2.9	2.9	2.8	2.9
Phe	0.5	0	0.5	0
His	0	0	0	0
Lys	4.6	4.3	6.7	4.3
Arg	4.1	4.3	7.1	4.3

when entire gas vesicles were analysed by the spinning-cup and gas-phase sequencers suggests that GVP is the only type of protein present in the structure, but it does not rigorously exclude the possibility of a second component with a blocked *N*-terminus. The yield of the *N*-terminal residues, up to 35%, lies within the usually encountered range, but again does not rigorously eliminate the possibility that there is another component of very similar composition.

The protein of the gas vesicle is arranged along 4.57 nm-wide ribs that are oriented normal to the long axis of the cylindrical structure. Blaurock & Walsby (1976) demonstrated a 1.15 nm repeating unit cell along the rib, and showed that collapsed membranes stacked with a periodicity of 1.95 nm. The product of these three measurements gave the overall volume of the unit cell, 10.25 nm³; multiplying by the calculated density of GVP (1.32 g·cm⁻³) and the Avogadro number gave the equivalent M_r 8158. This value is in approximate agreement with the calculated M_r of 7388. Walsby & Armstrong (1979) pointed out that the collapsed gas vesicle walls will stack together in a manner that leaves intervening spaces and that the stacking periodicity will therefore slightly exceed the average thickness, t , of the wall layer. They described a method of determining the average thickness, from measurements of the weight of protein and volume of gas space in a sample of gas vesicles, and from measurements of the three-dimensional geometry of the gas vesicle. Walsby & Armstrong (1979) assumed in their calculations that the cylinder radius (r) of the *Anabaena* gas vesicle was identical with that of *Microcystis* (35.6 nm). We have now demonstrated that the actual value of r is 42 nm (P. K. Hayes & A. E. Walsby, unpublished work). By using this value we

calculate, by the equations of Walsby & Armstrong (1979), that the average wall thickness $t = 1.8$ nm and hence that the unit cell has a volume of 9.46 nm³ and will accommodate a protein of M_r 7500, very close to the actual value for *Anabaena* GVP.

Now that the repeating unit cell is known to be equivalent to the individual GVP molecule (rather than some repeating unit of a larger molecule) it is possible to examine how features of the amino acid sequence may match with features of the molecular structure. Blaurock & Walsby (1976) demonstrated that there are two layers of β -sheet within the thickness of the gas vesicle wall. In each layer the β -chains run at an angle of 55° to the direction of the rib. In each GVP unit cell there are two pairs of chains, one pair in each layer. A possible interpretation of the structure [Fig. 2 in Walsby (1978)] is that it is like house bricks laid on the slant (as in herringbone brickwork, but with adjacent rows in the same, rather than alternating, directions). Walsby (1978) demonstrated that, at the 55° angle, the dipeptides in adjacent antiparallel chains and in adjacent molecules would be in horizontal register, and that the length of the β -chain across the rib (4.5 nm/sin 55°) would be about 5.5 nm. This would just accommodate 16 amino acid residues (16 × 0.345 nm = 5.52 nm). Alternate chains might be only 14 residues long in order to generate the staggered arrangement (E. D. T. Atkins & A. E. Walsby, unpublished work). The GVP polypeptide must wind from one wall layer to the next, where again there would be runs of 16 or 14 residues. The sequence shown in Fig. 4 may permit such a folded structure, e.g. with turns between residues 21–22, 35–36, 51–52 and 67–68.

The outer (water-facing) surface of the gas vesicle is hydrophilic (Walsby, 1971), whereas the inner (gas-facing)

Fig. 6. Polyacrylamide-gel electrophoretogram of GVP from *Anabaena flos-aquae* and *Microcystis* sp.

SDS/polyacrylamide slab gel showing *Anabaena* GVP (track 1) and *Microcystis* GVP (track 2). Markers: GPDH, D-glyceraldehyde-3-phosphate dehydrogenase (M_r 35000); SOD, superoxide dismutase (22700); myoglobin (17200); lysozyme (14300).

surface is hydrophobic (Worcester, 1975; Walsby, 1978). It would therefore be expected that residue side chains exposed at the outer surface should be hydrophilic and those at the inside hydrophobic. In antiparallel β -sheet, the side chains point alternately up (above the plane of the sheet) and down (below the plane). It might therefore be expected that in the outer β -chains the residues would be alternately hydrophilic (pointing up and out) and hydrophobic (pointing down and towards the middle of the wall). However, no such alternation is seen in the sequence. The sequence EKTNSSSS near the *N*-terminus is strongly hydrophilic and may lie across the outer surface of the wall. It would be expected that the residues of the inner β -chain would all be hydrophobic, as they would point alternately to the hydrophobic, gas-facing surface and to the hydrophobic middle of the wall. There are several hydrophobic reaches in the sequence and particularly in the C-terminal half of the molecule, but even here a few serine and threonine residues and acids or bases break the hydrophobic continuity. There are, moreover, two residues of tyrosine and one of tryptophan, which have bulky side chains and are not usually found in β -sheet structure.

The amino acid sequence data do not, therefore, seem to be compatible with a very simple structure composed entirely of β -sheet. It is emphasized that the X-ray-diffraction data available do not necessarily indicate

that the β -sheet extends over the whole width of both layers; if it did, then the maximum number of residues that could be accommodated within the two pairs of β -chains would be 60 compared with 70 in the GVP molecule. At least part of the molecule must therefore be outside the β -sheet.

The gas vesicles of different prokaryotes, although of very similar structure, do vary in width. The narrow ones provide buoyancy with less economy than wide ones, but they are stronger and will withstand greater pressures (Walsby, 1972). It appears that, in each organism, natural selection has produced a gas vesicle of the maximum width compatible with the strength required to withstand the hydrostatic pressure in its particular environment. If no other molecules are involved in shaping the gas vesicle, then its size and strength must be determined by a unique amino acid sequence; it would then follow that the natural selection of different gas-vesicle types might involve the evolution of only the structural gene that codes for GVP. We have synthesized a mixture of 17-base oligonucleotides that code for the amino acid sequence around residue 28 of *Anabaena* GVP. After originally submitting this paper we found a unique restriction fragment of the *Anabaena* DNA that hybridizes with the oligonucleotide probe, and we have cloned it into a pUC9 plasmid. We have also learnt that Tandeau de Marsac *et al.* (1985) have constructed a 29-base oligonucleotide on the basis of our previously published amino acid sequence and have used it to clone the GVP gene in *Calothrix*. The gene sequence confirms the *Calothrix* amino acid sequence shown in Fig. 5 (residue 26 is D) and corroborates the 70-residue size of the GVP molecule.

We thank Mr. F. D. Northrop, Mr. M. J. Runswick and Mr. S. J. Powell for assistance with sequencing, and Miss Annette Bees for assisting with gas-vesicle isolation from *Calothrix*. This work was supported by a grant from the Science and Engineering Research Council to A. E. W.

REFERENCES

- Allen, G. (1981) Lab. Tech. Biochem. Mol. Biol. **9**, 168–173
- Armstrong, R. E. & Walsby, A. E. (1981) in Organisation of Prokaryotic Cell Membranes (Ghosh, B. K., ed.), vol. 2, pp. 95–129, CRC Press, Cleveland, OH
- Blaurock, A. E. & Walsby, A. E. (1976) J. Mol. Biol. **105**, 183–199
- Laemmli, U. K. (1970) Nature (London) **227**, 680–685
- Tandeau de Marsac, N., Mazel, D., Bryant, D. A. & Houmard, J. (1985) Nucleic Acids Res. **13**, 7223–7236
- Tarr, G. E., Beecher, J. F., Bell, M. & McKean, D. J. (1978) Anal. Biochem. **84**, 622–627
- Thomas, R. H. & Walsby, A. E. (1985) J. Gen. Microbiol. **131**, 799–809
- Walker, J. E. & Walsby, A. E. (1983) Biochem. J. **209**, 809–815
- Walker, J. E., Carne, A. F., Runswick, M. J., Bridgen, J. & Harris, J. I. (1980) Eur. J. Biochem. **108**, 549–565
- Walker, J. E., Hayes, P. K. & Walsby, A. E. (1984) J. Gen. Microbiol. **130**, 2709–2715
- Walsby, A. E. (1971) Proc. R. Soc. London Ser. B **178**, 301–326
- Walsby, A. E. (1972) Bacteriol. Rev. **36**, 1–32
- Walsby, A. E. (1978) Symp. Soc. Gen. Microbiol. **28**, 327–358
- Walsby, A. E. & Armstrong, R. E. (1979) J. Mol. Biol. **129**, 279–285
- Worcester, D. L. (1975) Brookhaven Symp. Biol. **27**, 37–57
- Yoshimoto, T., Walter, R. & Tsuru, D. (1980) J. Biol. Chem. **225**, 4786–4792