Re-evaluation of the primary structure of *Ralstonia eutropha* phasin and implications for polyhydroxyalkanoic acid granule binding

Steven Zachary Hanley[a,*], Darryl J.C. Pappin[b], Dinah Rahman[b], Andrew J. White[a], Kieran M. Elborough[a], Antoni R. Slabas[a]

[a] *Department of Biological Sciences, University of Durham, Durham, DH13LE, UK*
[b] *Imperial Cancer Research Fund, London, WC2A 3PX, UK*

Received 2 February 1999

Abstract Sequence analysis of several cDNAs encoding the phasin protein of *Ralstonia eutropha* indicated that the carboxyl terminus of the resulting derived protein sequence is different from that reported previously. This was confirmed by: (1) sequencing of the genomic DNA; (2) SDS-PAGE and peptide analysis of wild-type and recombinant phasin; and (3) mass spectrometry of wild-type phasin protein. The results have implications for the model proposed for the binding of this protein to polyhydroxyalkanoic acid granules in the bacterium.

Key words: *Alcaligenes eutrophus*; *Ralstonia eutropha*; Polyhydroxyalkanoic acid granule; Phasin; Biodegradable polyester

1. Introduction

PHAs are carbon-oxygen storage polyesters occurring as intracellular granules in many bacteria [1]. The monomer subunits of PHA are usually 3-hydroxyalkanoic acids with a number of different side groups. The PHA polymers are biodegradable thermoplastics with the potential to replace the non-biodegradable plastics currently manufactured from the Earth's hydrocarbon reserves. Poly(3-hydroxybutyrate), the most studied member of the class, is synthesized from acetyl-CoA in a three-step pathway by the bacterium *Ralstonia eutropha* (formerly *Alcaligenes eutrophus* [2]). The copolymer poly(3-hydroxybutyrate-*co*-3-hydroxyvalerate) is produced commercially via the industrial fermentation of *R. eutropha* and is marketed as the biodegradable polythene substitute BIOPOL by Monsanto Corporation.

PHA granules in *R. eutropha* are thought to be enveloped in a predominantly protein monolayer [3]. The most abundant granule-associated protein is phasin, encoded by the *phaP* gene. Phasin appears to serve a structural role in a manner analogous to that of the oleosin proteins in oleaceous plants [4], which form a boundary layer sequestering hydrophobic triglycerides from the cytoplasm. Catalytically active granule-associated proteins such as the PHA synthase and depolymerase enzymes are also present on the granule surface but to a lesser extent (representing approximately 10–15% of the granule surface in one organism [5]). The proteins are not integral parts of a phospholipid mono- or bilayer and phospholipids cover less than 10% of the granule [4]. PHA biosynthesis in heterologous hosts in the absence of phasin results in the deposition of a single large granule rather than the numerous small granules seen in the presence of phasin and an identical phenotype is observed in *R. eutropha* cells in which transcription of the phasin gene has been disrupted by a transposon [6]. The rate of PHA biosynthesis in these mutant *R. eutropha* cells is lower than that of the wild-type, as is the specific activity of the PHA synthase enzyme. In vitro experiments in the biosynthesis of PHA show that PHA synthase turnover is increased in the presence of phasin, although reductions in specific activity or increases in product accumulation were not observed [7]. Western blot analysis of PHA granule-associated proteins from 50 bacterial strains using antibodies raised to the phasin of *R. eutropha* identified many PHA-accumulating bacteria which possessed cross-reacting proteins, while protein sequencing data from several Ralstonia strains indicated that phasin is highly conserved within the genus [8]. All of these observations would suggest that phasins have an important role in the dynamic processes of PHA accumulation and utilization.

Very little is known about the secondary and tertiary structures of the phasin protein in vivo. However, a model has been proposed for the anchoring of monomeric phasin to a phospholipid monolayer of a PHA granule, based on common motifs found in the deduced amino acid sequences of phasins of *R. eutropha* and *Rhodococcus ruber* [6]. These two proteins are of disparate sizes (reported as 24 kDa and 14 kDa, respectively) but each has been reported to contain two segments of hydrophobic and amphiphilic amino acids at the carboxyl terminus. Deletion of these regions (termed HD1 and HD2) from the *R. ruber* phasin prevented the protein from binding to PHA granules [9]. HD1 and HD2 are thought to span the granular membrane and anchor the protein, allowing the phasin to act as an amphiphilic barrier between the hydrophilic cytoplasm and the hydrophobic PHA granule. A similar process of granule binding has been assumed to take place with the *R. eutropha* phasin as it appeared to contain similarly non-hydrophilic regions at the carboxyl terminus.

In this article we report three differences in the DNA sequence of the *R. eutropha* phasin cDNA to the published data with subsequent alterations in the primary structure of the resulting protein. The result is confirmed by direct analysis of the phasin protein by SDS-PAGE, amino-terminal se-

*Corresponding author. Fax: +44 (191) 377 2417.
E-mail: s.z.hanley@durham.ac.uk

Abbreviations: PHA, polyhydroxyalkanoic acid; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; MALDI TOF MS, matrix-assisted laser desorption/ionization time of flight mass spectrometry; CAD, collision-activated dissociation

Data depositions: The nucleotide sequence reported in this paper has been submitted to the Genbank/EMBL/DDBJ databases under the accession number AF079155.

PII: S0014-5793(99)00235-5

quencing and mass spectrometry. These differences are unlikely to be a consequence of a tRNA mutation. The primary sequence of the new protein is incompatible with the current model of phasin anchorage to PHA granules. However the data in this work have allowed us to formulate an alternative model.

2. Materials and methods

2.1. Bacterial strains

Wild-type *R. eutropha* strain H16 (DSM428) was generously donated by Dr. A.J. Anderson (University of Hull, UK). *R. eutropha* strain H2275 and the plasmid pVK5000 was made available by Dr. A. Steinbüchel (Universität Münster, Germany). Freshly harvested *R. eutropha* H16 PHA granules were kindly supplied by Dr. H.E. Valentin and Dr. M. Tran (Monsanto Corporation, USA).

2.2. PCR

PCR was carried out using Vent Polymerase (New England Biolabs) according to the manufacturer's recommendations. The oligomers 5′GA24 (GAAGTTCCATATGGCCATGGTTATGATCCTCACCCC) and 3′GA24 (CAGCACCGCATATGGATCCTCAGGGCACTACCTTCATC), designed to amplify the coding region of the *R. eutropha* phasin gene (Genbank accession number X85729) (see Fig. 1), were used to amplify cDNA. The cycle conditions were as follows: hot start; 4 min at 95°C; 10 cycles of 1 min plateaux at 95°C, 45°C and 72°C; 25 cycles of 1 min plateaux at 95°C, 55°C and 72°C; 10 min at 72°C. Templates used were: plasmid pVK5000 (contains 5 kbp segment of the *R. eutropha* genome with a single ORF encoding the phasin protein) [6], *R. eutropha* strain H2275 (in which a transposon prevents transcription of the phasin gene *phaP* [6]), and *R. eutropha* strain H16.

2.3. Cloning and overexpression of phasin cDNA

The pCR-script (amp) cloning kit (Stratagene) was used to clone the 727 bp PCR products from pVK5000 and H2275. The cDNA derived from pVK5000 was subcloned into pET11a to create the vector pET-GA24, which was used to transform *E. coli* strain BL21(DE3) for overexpression of the phasin gene via IPTG-mediated induction. Plasmid was prepared for restriction analysis and sequencing using the Hybaid Recovery Plasmid Mini Prep Kit.

2.4. DNA sequencing

Dideoxy chain termination DNA sequencing was performed using the primers int1 (CCGGCTGCACGGCTGCGGCC) and int2 (CTCGAAGAACGTGCAAGCGC), and also with primers annealing to the vectors used. Both strands were sequenced using an ABI 377 Prism DNA Sequencer.

2.5. DNA and protein sequence analysis

Alignments and sequence analysis were performed using the GCG program suite [10] at the SEQNET service (CCL, Daresbury Laboratory). The Protein Structure Analysis Server (http://bmerc-www.bu.edu/psa/) was used for secondary structure predictions.

2.6. Purification of phasin from R. eutropha PHA granules and recombinant E. coli

Three different procedures were utilized: (1) Phasin was purified for amino-terminal sequencing and digestion analysis from PHA granules harvested from *R. eutropha* cultures using Triton X-114 solubilization [6]. This protein is hereafter referred to as X-114-phasin. Such samples are not suitable for MALDI TOF MS due to the heteropolymeric nature of the detergent. (2) Phasin was purified from granules for MALDI TOF MS by boiling granules (approximately 50 mg wet weight) in 10% (w/v) SDS. This protein is hereafter referred to as SDS-phasin. The homogeneous nature of SDS allows these samples to be used for mass spectrometry. Amino-terminal sequencing of the initial six residues was used to confirm the identity of each phasin as authentic. (3) Protein extracts were prepared from recombinant *E. coli* by boiling cells for 10 min in SDS-PAGE loading buffer. This protein is hereafter referred to as recombinant-phasin. Proteins were electrophoresed using discontinuous Tris-glycine SDS-PAGE (15% acrylamide) [11] with a Mini-PROTEAN II apparatus (Bio-Rad). Gels were stained with Coomassie Brilliant Blue.

2.7. Protein and peptide sequencing

Protein sequencing was carried out using an ABI 477A pulsed liquid protein microsequencer with 120A on-line analyzer. Granule extract containing phasin was electrophoresed as previously reported [11]. The protein was transferred to Problott membranes (Applied Biosystems) using a Mini Trans Blot apparatus (BioRad) in accordance with the manufacturer's instructions. Cyanogen bromide cleavage was carried out as previously reported [12]. Hexadimethrine bromide was used to prevent the loss of peptide fragments during cyanogen bromide cleavage.

2.8. Mass spectrometry

Following SDS-PAGE the stained protein band was excised from the gel and digested with trypsin [13]. Following overnight digestion, 0.5 μl was sampled directly from the digest supernatant for MS analysis using a TofSpec 2E laser desorption time of flight MS (Micromass, UK). The remaining digested peptides ($>90\%$ of total digest) were extracted with washes of 5% (v/v) aqueous trifluoroacetic acid and acetonitrile, pooled and dried. Peptides were then derivatized with *N*-succinimidyl-2-morpholine acetate to improve b-ion abundance and facilitate sequence analysis by tandem mass spectrometry [14]. Dried peptide fractions were treated with 7 μl of freshly prepared, ice-cold 1% (w/v) *N*-succinimidyl-2-morpholine acetate in 1.0 M HEPES (pH 7.8 with NaOH) containing 2% (v/v) acetonitrile. After incubation for 20 min on ice the reaction was terminated by the addition of 1 μl heptafluorobutyric acid and diluted with an equal volume of water. The solution was then injected in 3×5 μl aliquots onto a capillary reverse-phase column (300 μm$\times$15 cm) packed with POROS R2/H material (Perceptive Biosystems, MA) equilibrated with 2% (v/v) methanol/0.05% (v/v) trifluoroacetic acid running at 3 μl/min. The adsorbed peptides were washed isocratically with 15% (v/v) methanol/0.05% (v/v) trifluoroacetic acid for 30 min at 3 μl/min to elute the excess reagent and HEPES buffer. Derivatized peptides were eluted with a single step gradient to 75% (v/v) methanol/0.1% (v/v) formic acid and collected in two 3 μl fractions. Eleven derivatized peptides, from 11–19 residues in length, were then fully sequenced by low-energy collision-activated dissociation (CAD) using a Finnigan MAT LCQ ion-trap MS fitted with a nanoelectrospray source [15,16]. CAD was typically performed with collisional offset voltages between -17 V and -35 V.

3. Results

3.1. PCR and DNA sequence analysis

Initial studies were aimed at the isolation of phasin cDNA (727 bp) from the plasmid pVK5000 using the 5′GA24 and 3′GA24 oligomers. Sequencing of several clones found consistent differences from the Genbank database deposition (Fig. 1). These differences had effects on the length of the phasin protein and also the sequence of the carboxyl terminus, which has strong significance for the model proposed for phasin granule binding [6]. The same nucleotide differences in multiple clones from different PCR events using a high-fidelity polymerase (Vent) indicated that polymerase fidelity was not responsible for the observed sequence. One possibility was that mutations had arisen in the DNA either because the *phaP* gene had not been subjected to selection pressure or due to a reverse mistranscription. Accordingly, we amplified and sequenced phasin genomic DNA from *R. eutropha* strains H16 and H2275 which confirmed the results of PCR from the pVK5000 plasmid.

The DNA sequence contained three differences from the published sequence (Genbank accession number X85729) and the deduced amino acid sequence gave rise to a protein of 192 amino acids and a calculated molecular mass of 19966.35 Da compared with previously reported values of 228 amino acids and 23978.04 Da. The sequence differences (see Fig. 1) were: (1) substitution of a cytidine residue for a guanidine residue at position 369 of the ORF, with no effect

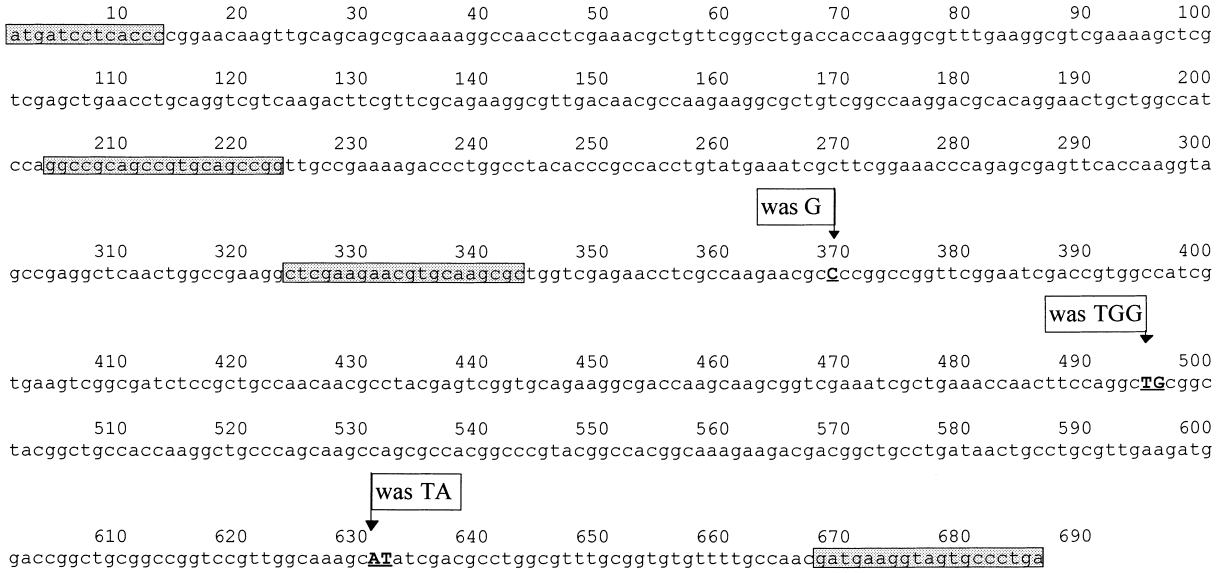

Fig. 1. Sequence of the *phaP* coding region with the differences from the published sequence (Genbank accession number X85729) in underlined boldface with boxes indicating the changes. The ORF (protein sequence shown in Fig. 3) is from 1–579 bp. Primers used in sequencing and PCR (boxed) are described in Section 2.

on the resulting amino acid residue; (2) omission of the guanidine residue reported at position 496 of the ORF, resulting in a shift of reading frame which alters the primary structure of the protein from that point forward (Gly-166 becoming Ala-166); and (3) exchange of the thymidine and adenine residues reported at positions 632 and 633 of the ORF, which falls outside of the new ORF due to the effect of the frameshift.

3.2. *Overexpression of phasin cDNA in E. coli and comparison of wild-type and recombinant forms*

Recombinant-phasin protein overexpressed from the pET-GA24 vector in *E. coli* strain BL21(DE3) was compared by SDS-PAGE with phasin from *R. eutropha* strain H16. The six amino-terminal residues of recombinant-phasin were determined by amino-terminal sequencing as Met-Ile-Leu-Thr-Phe-Glu, an exact match to the six amino-terminal residues of *R. eutropha* phasin [6]. The heterologously expressed protein comigrated with X-114-phasin in numerous SDS-PAGE experiments (example in Fig. 2). Recombinant-phasin protein was produced from the transcription translation of pET-GA24 vector DNA and was therefore as shown in Fig. 3 (not as previously published). No additional amino- or carboxyl-terminal residues were introduced via the cloning and overexpression. Comigration of recombinant-phasin with wild-type X-114-phasin from *R. eutropha* was further evidence supporting the hypothesis that the sequence of the phasin cDNA was an exact copy of the wild-type *R. eutropha phaP* ORF.

R. eutropha phasin protein did not migrate according to the predicted mass using standard protein molecular weight markers; Fig. 2 shows recombinant-phasin and X-114-phasin migrating equally at approximately 22 kDa compared to the reference masses. The calculated size of the phasin protein

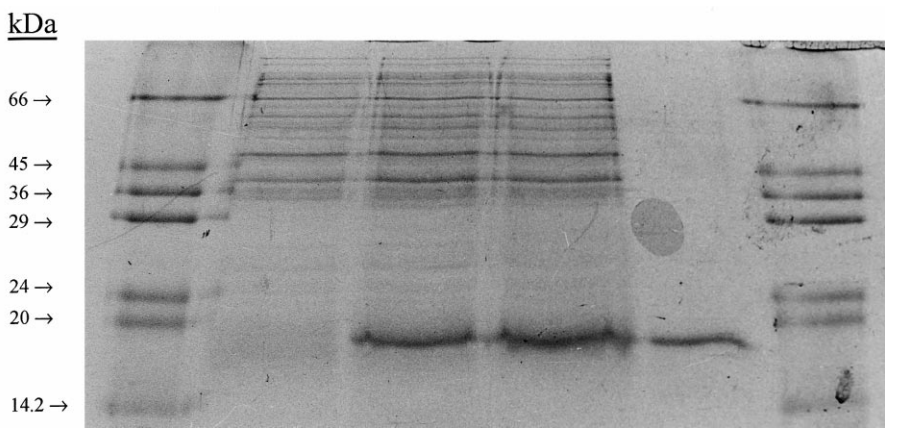

Fig. 2. SDS-PAGE 15% acrylamide gel of phasin from heterologous expression and from *R. eutropha* PHA granules. Lane 1, protein size standards (66, 45, 36, 29, 24, 20 and 14.2 kDa). Lane 2, recombinant-phasin pre-induction control. Lane 3, recombinant-phasin. Lane 4, recombinant-phasin mixed with X-114-phasin. Lane 5, X-114-phasin. Lane 6, protein size standards as above.

from the previously reported DNA sequence was 24 kDa whereas the new DNA sequence would encode a protein of mass 20 kDa. This anomalous electrophoretic mobility was observed in preparations of phasin from PHA particles regardless of which detergent (Triton X-114 or SDS) was used in the phasin extraction process and was not therefore due to the blocking of SDS-binding sites in Triton-treated samples. Since the native and heterologous proteins migrated together it was assumed that their masses, while indeterminate by this method, were equivalent. Anomalous electrophoretic mobility was not seen in published SDS-PAGE gels of the *R. ruber* phasin [9,17] but is not uncommon (e.g. acyl carrier protein from *E. coli*, which migrates at 18 kDa despite a mass of 9 kDa [18]). This can be due to hydrophobicity, conformation or acidity of a protein.

3.3. Lack of a carboxyl-terminal peptide fragment in X-114-phasin

Cyanogen bromide cleavage was performed on X-114-phasin to determine whether a second methionine residue was present within the primary structure (a clear difference between the original and new sequences; see Fig. 1). No secondary peptide was observed in direct Edman degradation of the sample after cleavage, a result consistent with the new protein sequence. Hexadimethrine bromide was used to prevent the loss of the expected small peptide (sequence Lys-Val-Val-Pro) during the cleavage reaction. The lack of this cleavage fragment could not confirm that a methionine residue does not exist close to the carboxyl terminus of the protein but the negative evidence was strongly suggestive.

3.4. Confirmation of protein sequence by MALDI TOF and ion-trap MS

Definitive evidence of the sequence of the phasin protein required the alignment of multiple digestion fragments with the proposed sequence. The protein was digested with trypsin and the resulting fragment masses determined by MALDI TOF MS (Table 1 and Fig. 3). Of the 20 possible tryptic fragments predicted from the sequence, 14 were matched by mass measurement alone. The remaining six peptides were all less than 500 Da and were not observed owing to interference from low-mass ions associated with the alpha-cyano-4-hydroxycinnamic acid matrix. Mass errors between observed and calculated masses were generally <0.1 Da. Masses corresponding to partial digest products were also observed for peptides T3+T4 and T5+T6. Eleven peptides of mass >1000 Da were derivatized with succinimidyl-2-morpholine acetate and sequenced by low-energy CAD using an ion-trap MS fitted with a nanoelectrospray source. Full sequences were obtained for each peptide (Table 1 and Fig. 3). These masses were in complete agreement with the newly deduced sequence, with the restriction that low-energy collisional fragmentation does not allow discrimination between leucine and isoleucine. MS data obtained for peptide T16, showing complete b-ion and y-ion coverage, is shown in Fig. 4. This sequence is of significance as it covers the region where the previously reported sequence diverges from that reported here.

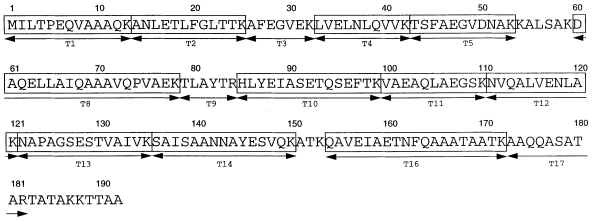

Fig. 3. Sequence and mass coverage for the deduced sequence of phasin. The protein. Tryptic peptides matched by mass measurement are underlined (↔). Peptide sequences obtained following low-energy CAD are boxed. Peptide designations are in sequence order.

Table 1
Summary of MS data for tryptic peptides of phasin

Peptide	Residues	Calculated protonated mass (Da)	Observed protonated mass (Da)	Δmass (Da)	Sequence obtained (low-energy CAD)
T1	1–13	1399.76	1399.72	−0.04	M[LI][LI]TPEQVAAAQK
T2	14–25	1307.72	1307.69	−0.03	AN[LI]ET[LI]FG[LI]TTK
T3	26–32	779.39	779.27	−0.12	–
T4	33–42	1154.71	1154.64	−0.07	[LI]VE[LI]N[LI]QVVK
T5	43–53	1138.54	1138.46	−0.08	TSFAEGVDNAK
T6	54–54	147.11	–	–	–
T7	55–59	489.30	–	–	–
T8	60–78	1965.07	1965.04	−0.03	DAQE[LI][LI]A[LI]QAAAVQPVAEK
T9	79–84	724.39	724.27	−0.12	–
T10	85–99	1782.86	1782.83	−0.03	H[LI]YE[LI]ASETQSEFTK
T11	100–110	1102.58	1102.47	−0.11	VAEAQ[LI]AEGSK
T12	111–121	1198.68	1198.63	−0.05	NVQA[LI]VEN[LI]AK
T13	122–135	1343.72	1343.64	−0.08	NAPAGSESTVA[LI]VK
T14	136–150	1552.76	1552.73	−0.03	SA[LI]SAANNAYESVQK
T15	151–153	319.20	–	–	–
T16	154–172	1934.98	1935.01	+0.03	QAVE[LI]AETNFQAAATAATK
T17	173–182	974.50	974.47	−0.03	–
T18	183–187	491.28	–	–	–
T19	188–188	147.11	–	–	–
T20	189–192	363.19	–	–	–
T3,4	26–42	1915.09	1915.12	+0.03	–
T5,6	43–54	1266.63	1266.60	−0.03	–

Tryptic peptides were numbered in sequence order (T1–T20). Peptide protonated masses ($[M+H]^{+1}$) measured by MALDI TOF MS are shown in column 4. Corresponding sequences were obtained by low-energy CAD following derivatization with *N*-succinimidyl-2-morpholine acetate. Leucine and isoleucine [LI] cannot be distinguished as they are isomers. Masses were also observed for the partial tryptic fragments T3+T4 and T5+T6.

```
y-ions      (2044.9) 1933.7 1862.7 1763.7 1634.6 1521.6 1450.4 1321.4 1220.3 1106.5 959.4
              Gln  -  Ala - Val -  Glu -  Xle -  Ala -  Glu -  Thr -  Asn -  Phe -  Gln -
b-ions      (256.1)  327.3  426.2  555.5  668.2  739.2  868.2  969.2  1083.1 1230.6 1358.3

y-ions      831.3  760.4  689.3  618.2  517.0  446.3  375.1  (274.1)
            Ala -  Ala -  Ala -  Thr -  Ala -  Ala -  Thr -  Lys
b-ions      1429.4 1500.4 1571.4 1672.3 1743.5 1814.4 1915.6 (2170.8)
```

Fig. 4. CAD MS data for peptide T16. Observed masses for b- and y-type fragment ions are shown for peptide T16 (SMA derivative). The doubly protonated ion at m/z 1095.0 was fragmented with a collision offset of −33 eV. Figures in parenthesis were not observed but could be calculated from the mass of the intact peptide, giving complete assignment of all 19 residues. Residue 5 is given the designation Xle as Leu and Ile cannot be distinguished.

4. Discussion

The model of phasin-PHA interaction described in the literature was deduced from a comparison between the deduced amino acid sequences of the phasins from *R. eutropha* and *R. ruber*. In the *R. ruber* phasin the PHA granule-binding regions HD1 and HD2 consist of two stretches of ten and nine consecutive hydrophobic or amphiphilic residues uninterrupted by hydrophilic residues near the carboxyl terminus. Deletion of these proposed membrane-spanning domains prevented the protein from adhering to PHA granules in heterologous expression studies [9]. The putative membrane-binding regions of the *R. eutropha* phasin were identified by searching the deduced amino acid sequence for similar regions of non-hydrophilicity based upon an assumption of structural similarity between the two phasins. Our work shows that the *R. eutropha* phasin does not contain the regions so identified and therefore the model proposed cannot be correct. The possibilities that the same stretch of DNA encodes the 24 kDa and 20 kDa phasins through a tRNA mutation (permitting the frameshift) or that there are two genes, are unlikely due to a single band observed on Western hybridizations previously

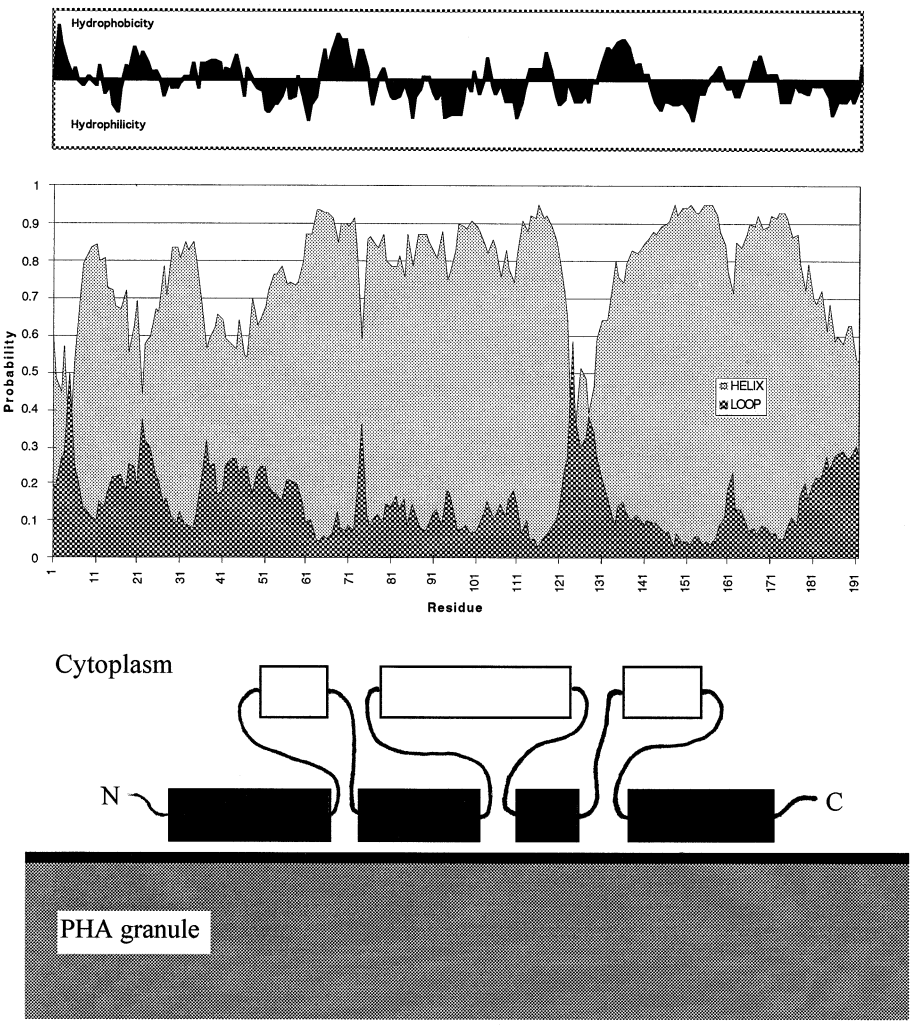

Fig. 5. Secondary structure predictions for the new phasin sequence. A plot of hydropathy is shown (positive values represent regions of hydrophobicity) above data for the probability per residue of being part of a helix or loop. A schematic of the proposed arrangement of hydrophobic and hydrophilic regions in relation to the surface of the PHA granule is also shown.

reported [6], since in both cases proteins of different sizes would be produced and resolved with the electrophoresis conditions used.

We have analyzed the sequence of the *R. eutropha* phasin generated by our work and have determined that it contains no consecutive runs of seven or more non-hydrophilic residues. The disparity in size between the *R. eutropha* and *R. ruber* phasins (20 kDa and 14 kDa, respectively) and the scores for identity (16.2%) and similarity (39.8%) between the two suggest that these two phasins are not phylogenetically related. The granule-associated protein encoded by the *phaP* gene (formerly designated ORF5 [19]) of *Chromatium vinosum* (Genbank accession number A27012) also has no significant homology to other phasin sequences. Two granule-bound proteins of small apparent molecular mass from *Acinetobacter* RA3849 (Genbank accession number L37761; termed ORF1) and *Aeromonas caviae* (Genbank accession number D88825, termed ORF) have been studied by separate groups [20,21] and are related (61% sequence similarity to each other but no significant similarity to other available phasin sequences). None of the above putative phasins contain regions of purely hydrophobic or amphiphilic character as seen in the *Rhodococcus* protein, and only the *Aeromonas* and *Acinetobacter* sequences share any significant similarity. The amino-terminal sequences of a number of putative phasins from methylotrophic bacteria [22] also show little homology to other phasins. This indicates that phasin proteins are functionally but not phylogenetically related, unlike the analogous oleosins which share a common ancestor. Extrapolation of structure-function relationships from one phasin to another must therefore be undertaken with caution.

Public domain sequence databases were searched using the complete sequence of the *R. eutropha* phasin protein (Fig. 3). One interesting match was to Gas Vesicle Protein C from *Anabaena flos-aquae* (accession number P09413) [23], which contained several short protein motifs similar to those found in the phasin sequence (e.g. the Gln-Glu-Leu-Leu-Ala motif beginning at residue 62). A search of the databases using the carboxy-terminal 25 amino acid residues (the novel sequence of this phasin generated by this work) revealed homology with the alanine-rich region of the mouse clathrin coat assembly protein AP180 (Genbank accession number Q61548), which is believed to interact with the cytoplasmic tails of membrane proteins on vesicle surfaces during synaptogenesis [24]. This region of the phasin protein may therefore possibly be involved in protein-protein interactions with other granule-associated proteins. It may be significant that both the *A. flos-aquae* and mouse proteins are located at intracellular membrane interfaces or boundaries, as are phasin proteins.

The Protein Structure Analysis Server (http://bmerc-www.bu.edu/psa/) [25–27], which uses large data sets derived from the Brookhaven protein database to assign probabilities of amino acid structural roles, was used to analyse the *R. eutropha* phasin protein sequence (Fig. 5). It is highly probable that the phasin consists predominantly of alpha-helical arrangements of residues. No region of the protein had a high probability of forming part of a beta-sheet or a turn. When these data were compared with calculations of hydrophobicity (Fig. 5 also), several features emerged. The region from Val-75 to Ala-123 is predominantly hydrophilic and of probably helical arrangement. This region is the longest uninterrupted stretch of both hydrophilicity and probably helicity; it may represent the cytoplasmic face of the protein. It is flanked by the two most hydrophobic regions of the protein; from Leu-64 to Gln-73 (a probable helix), and from Pro-124 to Ala-141 (a possible non-ordered loop region). This central region is preceded by the predominantly hydrophobic amino-terminal part of the protein containing an inverted repeat motif Gln-Val-Ala-Ala-Ala-Gln...(55)...Gln-Ala-Ala-Ala-Val-Gln beginning at residue 7 which is unique to the *R. eutropha* protein.

The carboxyl terminus of the *R. eutropha* phasin (following Ala-141) is predominantly hydrophilic, with two short (<9 residues) regions of amphipathicity. This is similar to the *R. ruber* and *C. vinosum* phasins, both of which possess a large predominantly hydrophilic and probably helical region flanked by slightly more hydrophobic regions. Secondary structure prediction using the *R. ruber* protein sequence suggests that the HD1 granule-binding region has an extremely low probability of being helical; this hydrophobic disordered region is similar to that seen in the *R. eutropha* structure prediction (Pro-124 to Ala-141). The related *Aeromonas* and *Acinetobacter* phasins are different, with hydrophilic residues being concentrated away from the central third of the primary sequence in each case, and the carboxyl-terminal hydrophilic region showing extremely high probability of helicity. The large hydrophilic domains may represent the cytoplasmic faces of the phasins. Our schematic model shows the *R. eutropha* phasin forming several potential granule-binding and cytoplasm-oriented domains.

Recent studies of the PHA granule boundary show that the interface may be as little as 4 nm in thickness, similar to that seen at the surface of gas vesicles [3]. If the granule boundary layer consists predominantly of protein, probably phasin, then hydrophobic/amphiphilic membrane-spanning domains would be superfluous and regions which bind directly to PHA will be responsible for the anchoring of the proteins to the granule. All known phasins have at least one markedly hydrophobic region (e.g. Pro-124 to Ala-141 of the *R. eutropha* phasin). Hydrophilic domains may form the cytoplasmic face of the protein (as depicted in Fig. 5), or possibly interact with other granule-bound proteins. Protein-protein interactions have not been extensively studied; one or more domains of the phasin may interact with the PHA synthase enzyme (which has been reported to have higher turnover in the presence of phasin in in vitro experiments [7]), the depolymerases or other unidentified granule-associated proteins. Experiments to determine the validity of our model using limited proteolysis of *R. eutropha* phasin in conjunction with site-directed mutagenesis experiments using recombinant phasin are currently underway.

Acknowledgements: This work was generously supported by the BBSRC (grant number IMI 31/5371). We would like to thank Professor Alexander Steinbüchel (Universität Münster, Germany) for the strain *R. eutropha* H2275 and the plasmid pVK5000, Henry E. Valentin and Minhtien Tran (Monsanto Corporation, USA) for providing crude extracts of *R. eutropha* PHA, Alistair J. Anderson (University of Hull, UK) for the strain *R. eutropha* H16 and John Gilroy (University of Durham, UK) for help with amino-terminal sequencing.

References

[1] Lee, S.Y. (1996) Biotechnol. Bioeng. 49, 1–14.
[2] Yabuuchi, E., Kosako, Y., Yano, I., Hotta, H. and Nishiuchi, Y. (1995) Microbiol. Immunol. 39, 897–904.

[3] Mayer, F. and Hoppert, M. (1997) J. Basic Microbiol. 37, 45–52.
[4] Steinbüchel, A., Aerts, K., Babel, W., Föllner, C., Liebergesell, M., Madkour, M.H., Mayer, F., Pieper-Fürst, U., Pries, A., Valentin, H.E. and Weiczorek, R.A. (1995) Can. J. Microbiol. 41, (Suppl. 1) 94–105.
[5] Mayer, F., Madkour, M.H., Pieper-Fürst, U., Weiczorek, R.A., Liebergesell, M. and Steinbüchel, A. (1996) J. Gen. Appl. Microbiol. 42, 445–455.
[6] Weiczorek, R., Pries, A., Steinbüchel, A. and Mayer, F. (1995) J. Bacteriol. 177, 2425–2435.
[7] Jossek, R., Reichelt, R. and Steinbüchel, A. (1998) Appl. Microbiol. Biotechnol. 49, 258–266.
[8] Weiczorek, R., Steinbüchel, A. and Schmidt, B. (1996) FEMS Microbiol. Lett. 135, 23–30.
[9] Pieper-Fürst, U., Madkour, M.H., Mayer, F. and Steinbüchel, A. (1995) J. Bacteriol. 177, 2513–2523.
[10] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[11] Matsudaira, P. (Ed.) (1993) A Practical Guide to Protein and Peptide Purification for Microsequencing, Academic Press Ltd, London.
[12] Wang, M.-P., Boulter, D. and Gatehouse, J.A. (1994) Plant Mol. Biol. 24, 159–170.
[13] Shevchenko, A., Wilm, M., Vorm, O. and Mann, M. (1996) Anal. Chem. 68, 850–858.
[14] Sherman, N.E., Yates, N.A., Shabanowitz, J., Hunt, D.F., Jeffery, W.A., Bartlet-Jones, M. and Pappin, D.J.C. (1995) in: Proceedings of the 43rd ASMS Conference on Mass Spectrometry and Allied Topics, Atlanta, GA, May 21–26, pp. 626–627.
[15] Hunt, D.F., Yates, J.R., Shabanowitz, J., Winston, S. and Hauer, C.R. (1986) Proc. Natl. Acad. Sci. USA 84, 6233–6237.
[16] Wilm, M. and Mann, M. (1996) Anal. Chem. 68, 1–8.
[17] Pieper-Fürst, U., Madkour, M.H., Mayer, F. and Steinbüchel, A. (1994) J. Bacteriol. 176, 4328–4337.
[18] Hill, R.R., MacKenzie, K.R. and Flanagan, J.M. (1995) Protein Expr. Purif. 6, 394–400.
[19] Leibergesell, M. and Steinbüchel, A. (1996) Biotechnol. Lett. 18, 719–724.
[20] Schembri, M.A., Bayly, R.C. and Davies, J.K. (1995) J. Bacteriol. 177, 4501–4507.
[21] Fukui, T. and Doi, Y. (1997) J. Bacteriol. 179, 4821–4830.
[22] Föllner, C.G., Babel, W. and Steinbuchel, A. (1995) Can. J. Microbiol. 41, (Suppl. 1) 124–130.
[23] Hayes, P.K., Lazarus, C.M., Bees, A., Walker, J.E. and Walsby, A.E. (1988) Mol. Microbiol. 2, 545–552.
[24] Zhou, S., Sousa, R., Tannery, N.H. and Lafer, E.M. (1992) J. Neurosci. 12, 2144–2155.
[25] Stultz, C.M., White, J.V. and Smith, T.F. (1993) Protein Sci. 2, 305–314.
[26] White, J.V., Stultz, C.M. and Smith, T.F. (1994) Math. Biosci. 119, 35–75.
[27] Stultz, C.M., Nambudripad, R., Lathrop, R.H. and White, J.V. (1997) in: N. Allewell and C. Woodward (Eds.), Protein Structural Biology in Bio-Medical Research, Advances in Molecular and Cell Biology Vol. 22B, JAI Press, Greenwich.