

# A New Family of Intrinsically Disordered Proteins: Structural Characterization of the Major Phasin PhaF from *Pseudomonas putida* KT2440

Beatriz Maestro<sup>1</sup>, Beatriz Galán<sup>2</sup>, Carlos Alfonso<sup>2</sup>, Germán Rivas<sup>2</sup>, Maria A. Prieto<sup>2</sup>, Jesús M. Sanz<sup>1\*</sup>

<sup>1</sup> Instituto de Biología Molecular y Celular, Universidad Miguel Hernández, Elche, Spain, <sup>2</sup> Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Madrid, Spain

## Abstract

Phasins are intracellular polyhydroxyalkanoate (PHA)-associated proteins involved in the stabilization of these bacterial carbon storage granules. Despite its importance in PHA metabolism and regulation, only few reports have focused so far on the structure of these proteins. In this work we have investigated the structure and stability of the PhaF phasin from *Pseudomonas putida* KT2440, a protein that is involved in PHA granule stabilization and distribution to daughter cells upon cell division. A structural, three-dimensional model of the protein was built from homology modeling procedures and consensus secondary structure predictions. The model predicts that PhaF is an elongated protein, with a long, amphipathic N-terminal helix with PHA binding capacity, followed by a short leucine zipper involved in protein oligomerization and a superhelical C-terminal domain wrapped around the chromosomal DNA. Hydrodynamic, spectroscopical and thermodynamic experiments validated the model and confirmed both that free PhaF is a tetramer in solution and that most part of the protein is intrinsically disordered in the absence of its ligands. The results lay a molecular basis for the explanation of the biological role of PhaF and, along with an exhaustive analysis of phasin sequence databases, suggest that intrinsic disorder and oligomerization through coiled-coils may be a widespread mechanism among these proteins.

**Citation:** Maestro B, Galán B, Alfonso C, Rivas G, Prieto MA, et al. (2013) A New Family of Intrinsically Disordered Proteins: Structural Characterization of the Major Phasin PhaF from *Pseudomonas putida* KT2440. PLoS ONE 8(2): e56904. doi:10.1371/journal.pone.0056904

**Editor:** Vladimir N. Uversky, University of South Florida College of Medicine, United States of America

**Received:** November 4, 2012; **Accepted:** January 15, 2013; **Published:** February 15, 2013

**Copyright:** © 2013 Maestro et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by grants BFU2010-17824, BIO2010-21049, BIO2007-67304-C02 (Spanish Ministry of Economy and Competitiveness, [http://www.mineco.gob.es/portal/site/mineco?lang\\_chosen=en](http://www.mineco.gob.es/portal/site/mineco?lang_chosen=en)) and FPA/2010/055 (Valencian Government, Spain, <http://www.gva.es/en/inicio/presentacion>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: [jmsanz@umh.es](mailto:jmsanz@umh.es)

## Introduction

Natural polyhydroxyalkanoates (PHAs) are organic polyoxoesters composed of (R)-3-hydroxy fatty acids which constitute the carbon and energy storage material of certain bacterial species under nutrient limitation conditions [1–6]. PHA is synthesized in the cytoplasm of bacteria and accumulates as multiple granules composed by the polyester (93–97% of the cell dried weight, CDW) coated by a phospholipid monolayer (1–6% of CDW) and proteins associated to the granule (1–2% of CDW), forming a layer at the granule surface [3]. Currently four classes of proteins associated to the bioplastic granules (GAP) have been defined: *i*) the PHA synthases, involved in the PHA polymerization, *ii*) the PHA depolymerases, implicated in bioplastic degradation and monomer mobilization, *iii*) the so-called phasins, devoided of enzymatic activity, and *iv*) other proteins [3,7].

PHAs are considered environmental-friendly plastics because they are biodegradable and biocompatible materials, so they have attracted much attention in diverse fields such as biotechnology, biomedicine, and agriculture [8–11]. Therefore, it follows that the understanding of protein-PHA interactions from a biophysical point of view will undoubtedly widen the biotechnological and clinical applications of these bioplastics. In this sense, phasins are low molecular weight proteins that constitute the major GAP fraction [12] and have been identified in different bacteria such as

*Ralstonia eutropha* [13], *Bacillus megaterium* [14], *Rhodococcus ruber* [15], *Paracoccus denitrificans* [16], *Pseudomonas putida* [17–19], and *Acinetobacter sp* [20]. These proteins are only synthesized when the PHA is produced [13;21–23] and play an important role in stabilizing the PHA granules, preventing them from coalescing and affecting their size and number in the cells [3,18,24]. Moreover, phasins are involved in the structure and in the degradation of poly(3-hydroxybutyrate) granules [25]. Besides their structural roles, phasins play also a regulating role in PHA metabolism. In this sense, the phasin PhaP from *Ralstonia eutropha* increases the specific activity of type II polyhydroxyalkanoate (PHA) synthases PhaC1 and PhaC2 from *Pseudomonas aeruginosa* [26]. We have recently demonstrated how PhaF, the major phasin from *P. putida* KT2440, is involved in the control of expression of the *phaC1* synthase and *pha I* phasin genes, as well as in granule localization within the cell and granule segregation during cell division [18].

In spite of their extensive physiological characterization, little is known about the structure and folding of phasins. From the analysis of the amino acid sequence we have previously suggested that the PhaF phasin from *P. putida* is structured in two domains: *i*) the N-terminal domain, with a possible function of binding to the PHA granule, and *ii*) the C-terminal domain showing characteristic 4- and 5-aa repetitions similar to those found in DNA-binding, nucleoid-associated proteins [17,18]. Several evidences

support the structural and functional independence of the domains. First, the PhaI phasin, which shares considerable sequence similarity with the N-terminal region of PhaF (57% similarity, 38% identity) [17] is capable of acquiring a folded, stable and functional structure by itself [27]. Also, the N-terminal domain is conserved in other PHA-binding phasins with a different C-terminal part [17,18]. Moreover, very diverse fusion proteins containing the N-terminal moiety of PhaF (the BioF affinity tag) can also be adsorbed on polyhydroxyalkanoate granules without compromising their function [27,28]. Finally, the isolated C-terminal moiety is able to bind to DNA [18].

The importance in understanding the molecular basis underlying the PHA-phasin interaction and its biological consequences, together with its biotechnological application such as the BioF tag, prompted us to characterize the structure and stability of the PhaF phasin of *P. putida* KT2440 by ultracentrifugation and optical spectroscopy techniques, and to elaborate a model of its three-dimensional structure explaining the binding to both PHA granules and DNA. The results suggest a peculiar structural organization of PhaF that supply an explanation to its biological role as coupling agents between PHA and the bacterial genetic material, and establish the basis for the designing of new variants of these proteins with novel biochemical properties.

## Materials and Methods

### Chemicals

Urea, ammonium sulphate, carboxymethyl cellulose, bovine pancreas deoxyribonuclease, magnesium chloride, sodium oleate, 2,2,2-trifluoroethanol and isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Butyl-sepharose-4 Fast Flow was from General Electric Healthcare (Piscataway, NJ, USA).

### Molecular modeling

The three-dimensional structure of the N- and C-terminal domains of PhaF protein was modeled separately as follows:

i) *N-terminal domain (residues 1-142) and C-terminal tail (residues 226-261)*. Secondary structure prediction was carried out by three different methods, *i.e.* PredictProtein (PHD, <http://www.predictprotein.org>) [29], JPred (<http://www.compbio.dundee.ac.uk/www-jpred/>) [30] and multivariate linear regression combination (MLRC) [31] ([http://npsa-pbil.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_server.html](http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_server.html)) (Fig 1A). A consensus prediction was performed by assigning a particular secondary structure to each residue if that was predicted by at least two of the methods. Helical coiled-coil sequences were predicted by the method of Lupas *et al.* [32] in the COILS server ([http://www.ch.embnet.org/software/COILS\\_form.html](http://www.ch.embnet.org/software/COILS_form.html)) and MARCOIL (<http://bcf.isb-sib.ch/webmarcoil/webmarcoilC1.html>) [33], selecting in both cases the segments with coil probability above 75%. The helical wheel representation of the coiled-coils was accomplished with DrawCoil 1.0 (<http://www.gevorggrigoryan.com/drawcoil/>). To build the three dimensional model, the coiled-coil predicted stretch (residues 111-133) was modelled using Swiss PDB Viewer 3.7 [34] using tropomyosin, a known coiled-coil protein, as the template (Protein Data Bank -PDB- code 1C1G). The rest of the domain was also modelled using the Swiss PDB Viewer 3.7 utility, assigning standard  $\alpha$ -helix ( $\phi = -65^\circ$ ,  $\psi = -40^\circ$ ) or randomized Ramachandran angles taking into account the consensus prediction.

ii) *C-terminal domain (residues 143-225)*. This section was *de novo* modeled using as a guide the theoretical model of the DNA-binding region of the AlgP protein from *Pseudomonas aeruginosa* [35]. This model assigned standard helix and coil Ramachandran angles to the different 4- and 5-aa repeats found in AlgP. A sequence alignment allowed the identification of these repeats in the C-terminal domain of PhaF (Fig 1C). Therefore, we assigned the same group of standard angles to those repeats that were conserved between the two proteins.

iii) *Overall structure*. The peptidic backbones of the models for the N- and C-terminal domains were linked together using SwissPDB Viewer, and the whole structure was subjected to steepest descent energy minimization and checked for structural errors. In order to get a representation of the binding of PhaF to DNA, a three-dimensional model of the 31-bp oligonucleotide that was already shown experimentally to bind to PhaF [18] (5'-AATTCACAGTAAAACCA-GATGGCTTGATTAC-3', and its complementary strand) was generated with the *Model.it* server ([http://hydra.icgeb.trieste.it/dna/model\\_it.html](http://hydra.icgeb.trieste.it/dna/model_it.html)) and manually docked onto the PhaF model using Swiss PDB Viewer for presentation purposes (Fig. 1E). Figures were rendered with PyMol (Delano Scientific LLC).

Prediction of protein disorder was accomplished using the following methods, all contained in the Disprot Database of Protein Disorder (<http://www.disprot.org/predictors.php>): PONDR-FIT<sup>TM</sup> [36]; DISOPRED2 [37], selecting a false positive rate threshold of 2%; DisEMBL<sup>TM</sup> [38], selecting the predicted regions with nonassigned electron densities in the Protein Data Base (REM465); and DISpro [39].

### Protein purification

Full-length PhaF and its C-terminal domain (C-PhaF) were expressed, purified and quantified as previously described [18].

### Analytical ultracentrifugation

Sedimentation velocity experiments of were carried out by spinning a PhaF solution (38  $\mu$ M; equilibrated in 20 mM sodium phosphate buffer, pH 7.0, plus 150 mM NaCl) at 48000 rpm and 20°C in an XL-A analytical ultracentrifuge (Beckman-Coulter Inc.) with UV-VIS optics detection system, using an An50Ti rotor and 12 mm double-sector centrepieces. Sedimentation profiles were registered every 5 minutes at 260 and 275 nm. The sedimentation coefficient distributions were calculated by least-squares boundary modelling of sedimentation velocity data using the *c(s)* method [40], as implemented in the SEDFIT program. These *s*-values were corrected to standard conditions (water, 20 °C, and infinite dilution) [41] using the SEDNTERP program [42] to get the corresponding standard *s*-values ( $s_{20,w}$ ). A sedimentation equilibrium experiment was carried out to determine the state of association of PhaF. Short-column (80–100  $\mu$ l) equilibrium runs were carried out at 10000 and 16000 rpm and the corresponding scans were measured at 238 nm, using the same experimental conditions and instrument as in the sedimentation velocity experiments. After the equilibrium scans a high-speed centrifugation run (43000 rpm) was done to estimate the corresponding baseline offsets. Weight-average buoyant molecular weight of the protein was determined by fitting a single species model to the experimental data using a MATLAB program (kindly provided by Dr. Allen Minton, NIH) based on the conservation of signal algorithm [43]. The corresponding protein molecular weight was determined from the experimental buoyant mass using 0.744 cm<sup>3</sup>/g as the partial specific volume of PhaF (calculated



from the amino acid composition using the SEDNTERP program, [42]).

### Circular dichroism spectroscopy

Circular dichroism (CD) experiments were carried out in a Jasco J-810 spectropolarimeter equipped with a Peltier PTC-423S system. Isothermal wavelength spectra were acquired at a scan speed of 50 nm/min with a response time of 2 seconds and averaged over at least 6 scans at 20 °C. Protein concentration was 3.8 or 38 μM and the cuvette pathlength was 0.1 cm (far-UV) or 1 cm (near-UV). Molar ellipticities ( $[\theta]$ ) are expressed in units of deg cm<sup>2</sup> (dmol of residues)<sup>-1</sup>. With urea present, spectra could not be recorded below 215 nm due to the high absorbance of the sample. Estimations of secondary structure content were calculated by deconvolution of the far-UV CD spectra using the CDNN program (Gerald Böhm, Institut für Biotechnologie, Martin-Luther-Universität Halle-Wittenberg, Germany) as well as the CONTIN [44] and CDSSTR [45] procedures contained in the Dichroweb utilities (<http://dichroweb.cryst.bbk.ac.uk/html/home.shtml>) [46,47]. Buffer was 20 mM sodium phosphate, pH 7.0, unless otherwise stated. C-PhaF-DNA binding experiments were carried out as previously described [18]. Briefly, an unspecific DNA fragment (nspDNA) was prepared from hybridization of oligonucleotides 5'-AATTCACAGTAAAACCA-GATGGCTTGATTAC-3' and its complementary strand (Invitrogen), to a final stock concentration of 0.5 mg ml<sup>-1</sup>. The C-PhaF and nspDNA concentration was 9 mM. For the pH stability experiments buffers were 50 mM sodium phosphate (pH 6.0–8.0), 50 mM sodium acetate (pH 3.5–5.5) or 50 mM glycine (pH 2.0–3.0 and 9.0–9.5). Final pH was measured *in situ* using a Crison Basic-20 pH-meter. Samples were centrifuged 5 min prior CD measuring.

### Fluorescence spectroscopy

Emission scans were performed at 20 °C on an PTI-QuantaMaster spectrofluorimeter (Birmingham, NJ, USA), model QM-62003SE, using a 5×5 mm path length cuvette and a protein concentration of 3.8 μM. Tryptophan emission spectra were obtained using an excitation wavelength of 280 nm, with excitation and emission slits of 0.5 nm and 0.7 nm respectively, and a scan rate of 60 nm min<sup>-1</sup>. Wavelength of the maximum intensity was calculated by first derivative analysis. The average emission intensity,  $\langle \lambda \rangle$ , in fluorescence spectra was calculated as depicted in Eq. 1

$$\langle \lambda \rangle = \frac{\sum_1^n \left( \frac{1}{\lambda_i} I_i \right)}{\sum_1^n I_i} \quad (1)$$

where  $I_i$  is the fluorescence intensity measured at any  $\lambda_i$  wavelength. Buffers used were the same as for CD.

### Equilibrium denaturation

CD-monitored thermal denaturation experiments were performed in a 0.1 cm (far-UV) or 1 cm (near-UV) path cell. The sample was layered with mineral oil to avoid evaporation, and the heating rate was set to 1 °C min<sup>-1</sup> unless otherwise stated. The thermodynamic analysis assumed a two-state unfolding-dissociation coupled mechanism of tetramer denaturation:



where F and U denote the folded and unfolded species, respectively. Effective free energies ( $\Delta G_{eff}^0$ ) were calculated from the CD titration traces:

$$\Delta G_{eff}^0 = -RT \ln K_{eq} = -RT \ln \frac{[\theta]_I - [\theta]_x}{[\theta]_x - [\theta]_F} \quad (3)$$

where  $K_{eq}$  is the equilibrium constant between the initial and final states,  $[\theta]_I$  and  $[\theta]_F$  are the ellipticities of the initial and final state, respectively, and  $[\theta]_x$  is the experimental ellipticity at a given urea concentration.

Thermal scans were fitted by least squares to the Gibbs-Helmholtz equation (Eq. 4):

$$\Delta G_{eff}^0 = \Delta H_m \left( 1 - \frac{T}{T_m} \right) - \Delta C_p \left[ (T_m - T) + T \ln \frac{T}{T_m} \right] \quad (4)$$

where  $\Delta G_{eff}^0(T)$  is the effective free energy of the transition,  $\Delta H_m$  is the van't Hoff enthalpy,  $T_m$  is the midpoint of denaturation (in Kelvin) and  $\Delta C_p$  is the difference in heat capacity between the native and denatured states. The intrinsic free energy of the tetramer ( $\Delta G_{int}^0$ ) is therefore calculated as follows:

$$\Delta G_{int}^0(T) = \Delta G_{eff}^0(T) - RT \ln(32C_0^3) \quad (5)$$

being  $C_0$  the total monomer concentration. Reader is referred to Backmann *et al.* [48] for the detailed description of equations 4–5.

For urea titrations, aliquots from an 8.0 M denaturant stock solution containing the same concentration of protein as in the cuvette (to keep the protein concentration constant throughout the titration, this is, 3.8 μM) were added stepwise and incubated for 5 min prior to record the wavelength spectra. Experiments were performed at 10 °C in 20 mM sodium phosphate buffer, pH 7.0. Intrinsic free energy calculations ( $\Delta G_{int}^0$ ) were accomplished using eqs. 3 and 5, but in this case  $\Delta G_{eff}^0$  was obtained using the linear extrapolation method [49]:

$$\Delta G_{eff} = \Delta G_{eff}^0 - m[\text{urea}] \quad (6)$$

where  $m$  denotes the dependence of  $\Delta G_{eff}$  with [urea], and  $\Delta G_{eff}^0$  is the free energy in the absence of denaturant, which can be calculated as:

$$\Delta G_{eff}^0 = m[\text{urea}]_{1/2} \quad (7)$$

being  $[\text{urea}]_{1/2}$  the midpoint of the transition.

## Results

### Three-dimensional structural model of PhaF

The setup of an efficient overexpression and purification system for PhaF [18] allowed the availability of high amounts of pure protein (Fig. S1) that was subsequently subjected to crystallization trials aimed to the elucidation of its structure by X-ray diffraction techniques. However, no suitable crystals have been obtained so far, either in the presence or absence of DNA, prompting us to elaborate instead a three-structural model of the protein. Homology modeling constitutes a commonly used procedure, provided that there are sufficiently close homologous sequences in the databases of protein structures. Nevertheless, only the C-terminal, DNA-binding domain displays some similarity with nucleoid-associated-like proteins [18]. Therefore, we decided to

employ different procedures separately for each of the N- and C-terminal domains and then join together the two models into a single structure. As it will be discussed below, the predicted extended structure and lack of tertiary contacts between domains supported this approach.

The secondary structure of the N-terminal domain was predicted by three different methods as predominantly  $\alpha$ -helical with a high consensus degree (Fig. 1A). A long, uninterrupted  $\alpha$ -helix would span residues 13–105, followed by a shorter helix encompassing residues 111–133. Such a long helix is likely to be unstable unless it is involved in stabilizing protein-ligand or protein-protein interactions. It should be reminded that secondary structure prediction methods usually unveil local propensities of peptide segments which might not acquire such structure when isolated from the rest of the protein or other stabilizing ligands, the PHA granule in this case. The rest of positions (residues 1–12, 106–110 and 134–142) were mostly assigned random-coil conformations by the methods. Remarkably, this second helical stretch is also compatible with a heptad-repeat pattern characteristic of a helical coiled-coil conformation, as suggested from the analyses by the Lupas and Marcoil methods (Fig. 1A). The fact that all "d" positions are occupied by leucine residues strongly suggests that amino acids 111–133 conform a leucine-zipper involved in PhaF oligomerization (Fig. 1B). Hence, we assigned standard  $\alpha$ -helical angles to the long helix, and tropomyosin-based coiled-coil angles to the short helix. Scarcity of irregular structures, that accumulate mostly at the extremes of the sequence, as well as of loop sequences, makes the N-terminal domain very likely to adopt an extended structure without loops and bends that allow significative tertiary contacts (Fig. 1E). One of the most remarkable features in the long helix is the segregation of polar residues from the hydrophobic ones on both sides of the helix, creating an amphipathic sequence in which the polar face is in turn very abundant in acidic and basic charged residues (Fig. S2).

With respect to the C-terminal domain, its highly repetitive sequence between residues 143–225 displays an appreciable similarity with that of the AlgP transcriptional regulatory protein from *P. aeruginosa* (Fig. 1C), whose structure was previously modelled [35]. This allowed the assignment of the Ramachandran angles to conserved residues, with very few gaps that were otherwise assigned random-coil values. Fig. 1E displays the final modelled structure of the C-terminal domain, within the entire protein, forming a superhelix able to be inserted into the major groove of DNA. Fig. S2 also shows the disposition of repeated lysine residues, suitably placed to interact with the negatively charged sugar-phosphate backbone in the DNA. The highly charged nature of this region suggests that electrostatic repulsions render it structurally disordered unless complexed with the nucleic acid.

Finally, residues 226–261 constitute a degenerated tail sequence derived in part from the DNA-binding moiety, but lacking any basic residues. No homologous sequences were found in the databases of proteins with known tertiary structure, and all methods predicted unequivocally a random-coil conformation, although this sequence contains some PXXP repetitions that are have been described to promote poly(proline)-II conformations (see below).

As neither domain is predicted to establish supersecondary or tertiary structure contacts, nor interdomain interactions, the separate modeled domains were simply joined by ligating the backbone of residues 142 and 143, giving rise to the structure shown in Fig. 1E. The result is an extended polypeptide composed of secondary structure segments lacking a defined hydrophobic core. As stated above, this lack of stabilizing contacts might give

rise to a partially unfolded structure in the absence of PHA and DNA. To check this, PhaF disorder was predicted by four methods as described in Materials and Methods. Results are shown in Fig. 2. All methods are coincident in pointing at the C-terminal domain as mostly disordered, whereas they also predict an appreciable flexibility in the N-terminal helical moiety of the protein. According to the PONDR-FIT procedure, stretches with more than 50% probability of being disordered in solution comprise residues 1–17, 58–91 and 134–261 (Fig. 2).

Our next step was trying to validate experimentally the most noteworthy features drawn from the model.

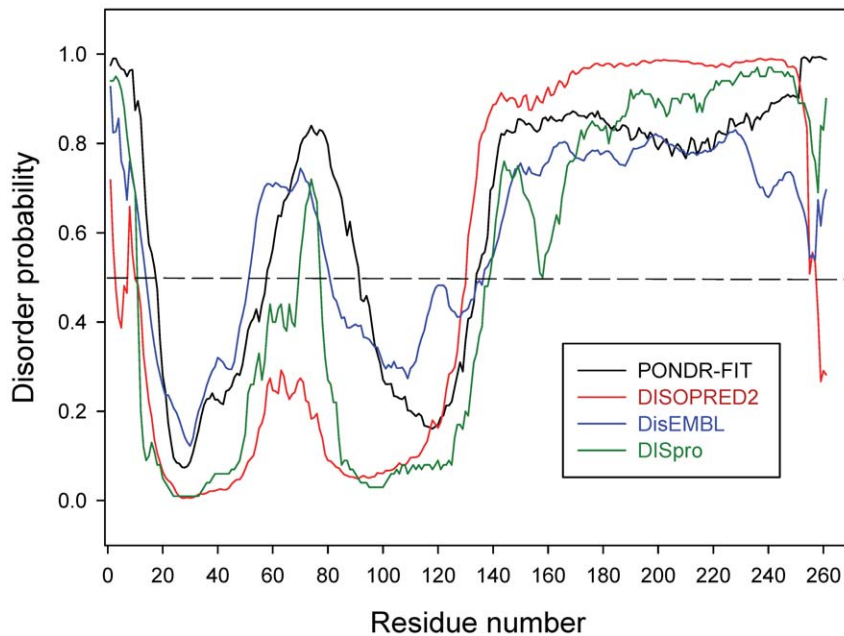
### Analytical ultracentrifugation

To check whether PhaF is an oligomer in solution, possibly *via* the predicted leucine-zipper region, a sedimentation velocity experiment was carried out (Fig. 3A) with the phasin purified by hydrophobic chromatography as previously described [18] (Fig. S1). A major species with sedimentation coefficient of  $3.4 \pm 0.1$  S and a frictional ratio ( $f/f_0$ ) of 2.3 (indicative of a highly elongated species) was detected, accounting for around 71% of the loading signal. This species is compatible with an extended tetramer. Two other species with *s*-values  $2.7 \pm 0.1$  (21%) and  $2.0 \pm 0.1$  S (7%) were also found. The sedimentation coefficient distribution and relative abundance of the species did not change with protein concentration (from 3.8 to 38  $\mu$ M; data not shown), suggesting that protein exists as a mixture of species at slow equilibrium or as an irreversible non-equilibrium mixture. These results were confirmed by low-speed sedimentation equilibrium experiments done using short solution columns to yield relatively shallow gradients (Fig. 3B). Under these conditions the apparent molar mass becomes essentially independent of the radial distance and is well described by the solution-average molar mass value. In our case, the experimental gradients were best described by an average mass value of  $85000 \pm 3000$  Da, somehow lower than the expected for a tetramer (105000 Da) and compatible with the relative abundance of species previously estimated from the sedimentation velocity data, being the tetramer the predominant one.

### Circular dichroism studies

We used circular dichroism as structural probe to determine the thermal stability of PhaF. The far-UV CD spectrum registered at 20 °C displays an appreciable contribution of  $\alpha$ -helical structure, with a minimum centered at 206 nm and a shoulder around 222 nm [50] (Fig. 4A). Three different mathematical procedures were employed to deconvolute the far-UV CD spectrum into secondary structure contributions (Table 1). Overall, the methods are coincident in the quantification of  $\alpha$ -helix as the predominant structure (around 30%), although with somehow lower values than those predicted from the structural model (44%, Fig. 1). One possible explanation is that, in the absence of other stabilizing contacts with its PHA substrate, the long, N-terminal helix may be partially unfolded in solution, as it is predicted by several disorder prediction methods (Fig. 2), so that it would only acquire its full structure upon binding to the bioplastic granule. In agreement with this hypothesis, the consideration of sequences 1–17 and 58–91 as disordered instead of  $\alpha$ -helical, according to the PONDR-FIT program (Fig. 2) reduces the helix content of unligated PhaF down to 31%, in full concordance now with the experimental CD data (Table 1). It would be desirable to analyze the conformational changes induced in the protein by the PHA polymer. However, PHA forms insoluble granules that, at best, may only be prepared as latex emulsions that are not suitable for fluorescence or CD spectroscopy techniques. Instead, we checked the effect of sodium





**Figure 2. Prediction of PhaF disordered regions.** Programs used were PONDR-FIT, DISOPRED2, DisEMBL AND DISpro (see Materials and Methods).

doi:10.1371/journal.pone.0056904.g002

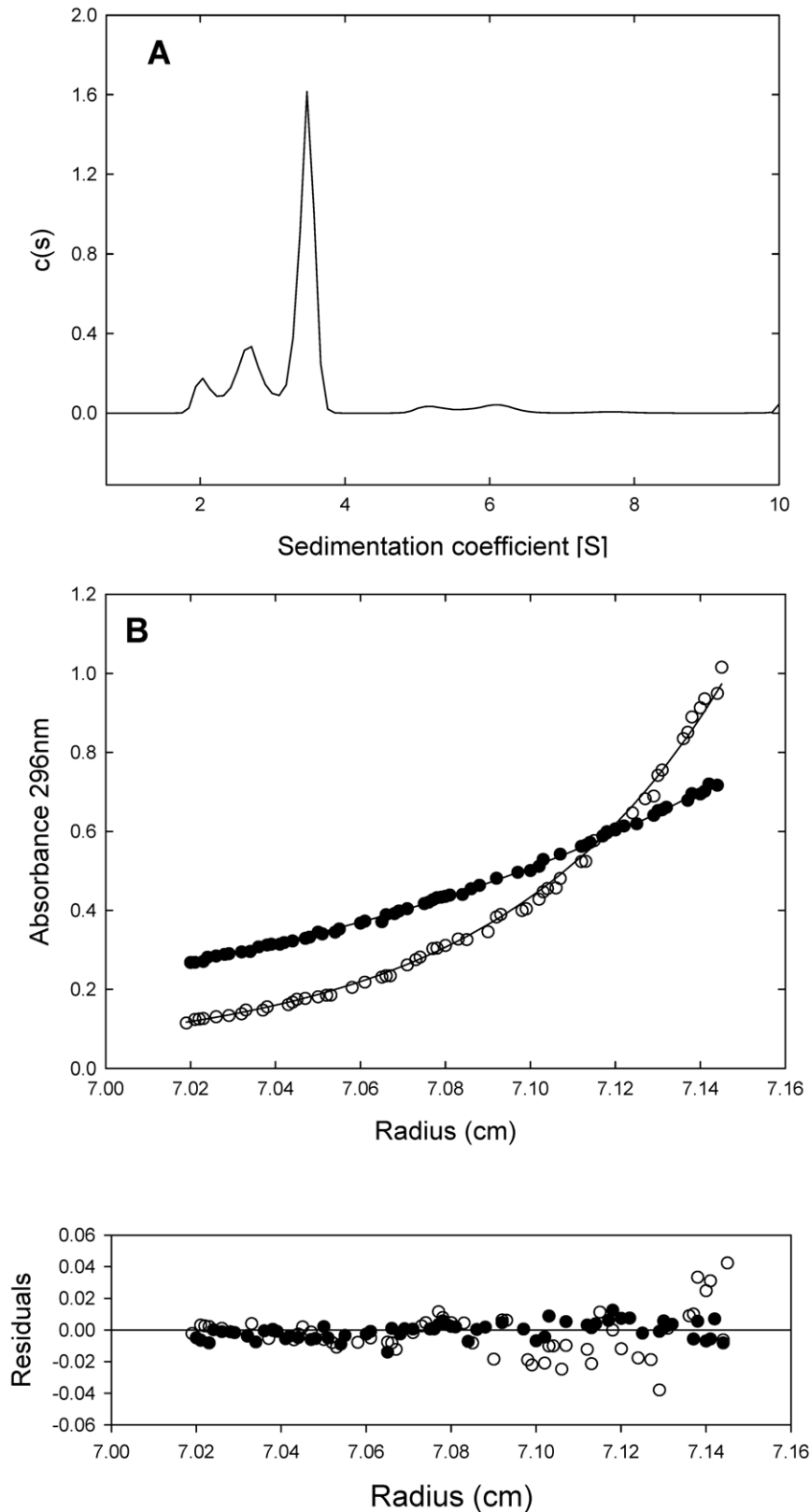
oleate, a common coating component for PHA preparations *in vitro*, as a hydrophobic mimic of PHA [51]. Fig. S3 shows that addition of 1 mM oleate effectively enhances the negative ellipticity of PhaF, with more prominent minima at 208 nm and 222 nm as a consequence of the increase in  $\alpha$ -helical content at the expenses of remainder conformations, according to a CDSTTR deconvolution of the spectrum (39%  $\alpha$ -helix, 12%  $\beta$ -structures, 49% remainder structures). This suggests that a hydrophobic environment such as that provided by oleate or PHA may interact with the apolar face of the amphipathic N-terminal helix, producing a stabilization effect.

With the aim of studying the structure of the N- and C-terminal domains separately, we overexpressed both moieties independently. However, while the C-terminal domain (C-PhaF protein) could be efficiently purified in high amounts [18] (Fig. S1), attempts to obtain the isolated N-terminal domain were unsuccessful due to the high insolubility of the polypeptide. In any case, the far-UV CD spectrum of the C-PhaF domain displays as the most remarkable feature the presence of a single minimum centered at 198 nm (Fig. 4B), which is indicative of random-coil conformation as the major structural component [50]. This is apparently at odds with the predicted superhelical structure of this part of the protein in the model when bound to DNA (Fig. 1E, Fig. S2). Nevertheless, the high number of positively charged lysines in the C-terminal domain (Fig. 1C) is likely to promote strongly repulsive interactions that, unless compensated with its polyanionic DNA ligand, would promote extensive unfolding of the polypeptide in this region. In this sense, this lack of structure does not prevent the isolated C-PhaF polypeptide to recognize and bind its DNA target [18]. Fig. S3 also shows that binding of a DNA oligonucleotide to C-PhaF induces a conformational change in the protein that is reflected in a slight change in its far-UV CD spectrum. The difference spectrum between the free and bound conformations (inset) displays a maximum at 219 nm and a sharp minimum at 194 nm that is compatible with a decreased content in poly(proline)-II structure upon binding [52].

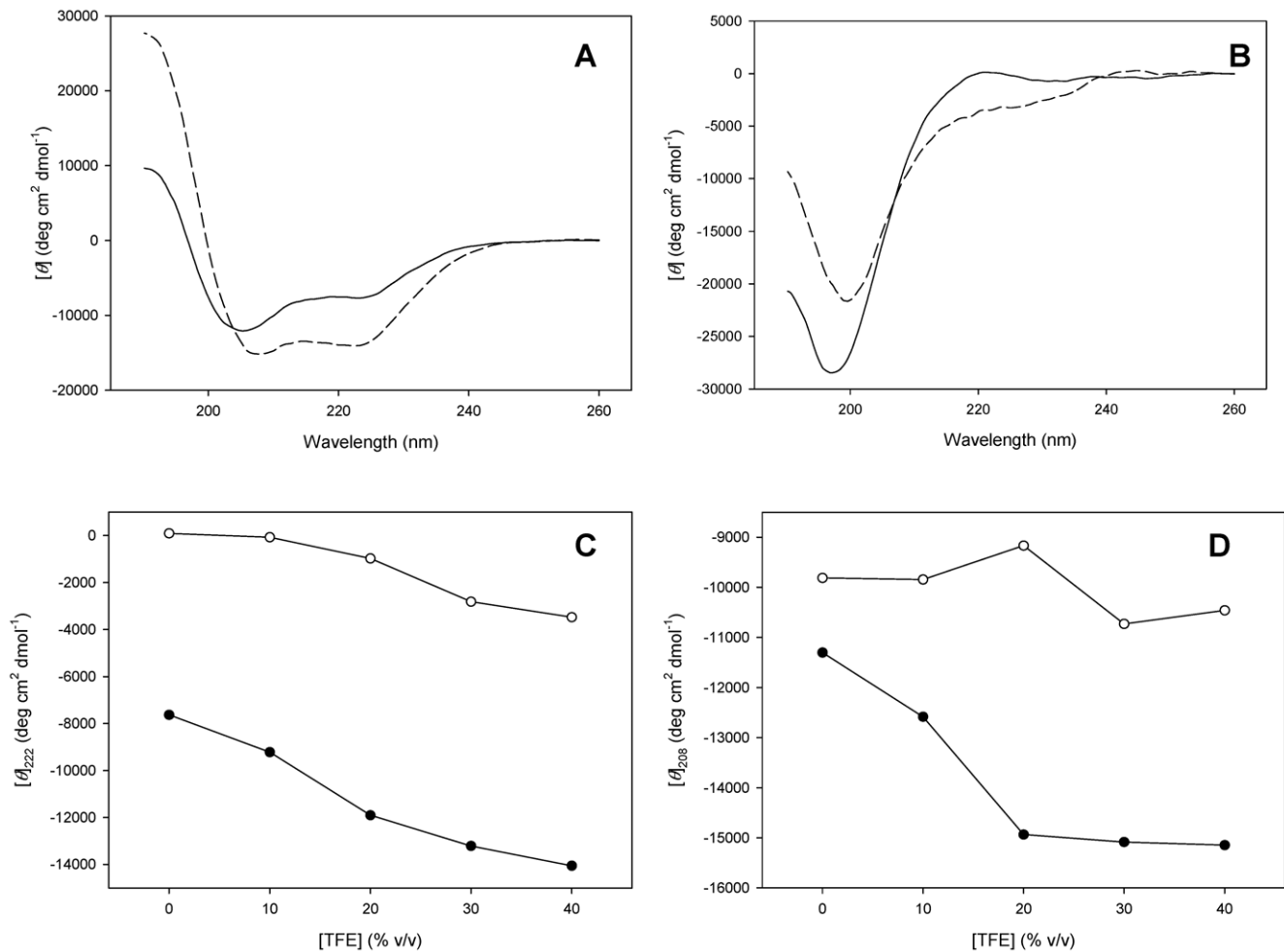
To overcome, insofar as it is possible, the tendency to disorder and therefore to evaluate the secondary structure propensity of the protein, we subjected both the PhaF and C-PhaF proteins to 2,2,2-trifluoroethanol (TFE) titrations, monitoring the CD signal. TFE is a commonly used cosolvent that usually unveils the intrinsic conformational preferences of peptides and proteins [53]. Fig. 4 shows that addition of TFE induces a cooperative conformational change that substantially increases the negative ellipticity leading to minima at 208 nm and 222 nm. Deconvolution of the spectrum recorded at the end of the titration (40% TFE) reveals an increase in  $\alpha$ -helical content up to approximately 47%, which is very close to the values predicted from the structural model (44%) (Table 1). On the other hand, titration of C-PhaF induces a much lower change in ellipticity, which is translated to a minor increase in helicity (around 7%, Table 1). This result shows that the tendency of the C-terminal domain of the protein to acquire a regular secondary structure is low even under strong helix-inducing conditions

### Thermal stability

To get an estimation of the stability of the PhaF protein, we performed CD-monitored thermal scans on the full-length PhaF. Incubation up to 90 °C caused extensive unfolding of the protein, as deduced from the decrease in the far-UV CD signal together with the emergence of a single minimum at 202 nm (Fig. 5A). On the other hand, the near-UV CD spectrum indicates the existence of rigid surroundings around aromatic residues (Fig. 5B), probably Trp-88 and Phe-95, which are very close to the leucine-zipper sequence (111–133) and are likely to be affected by the tetramerization process. Heating also caused the disappearance of this near-UV CD signal (Fig. 5B). Both far- and near-UV CD signals underwent a cooperative change with a transition midpoint centered at about 51 °C using a protein concentration of 38  $\mu$ M (Fig. 5D, Table 2), suggesting that the protein contains a degree of packing in the N-terminal domain (since the C-terminal moiety does not display any cooperative transition – see Fig. 5D, inset).



**Figure 3. Analytical centrifugation experiments.** (A) Sedimentation coefficient distributions  $c(s)$  corresponding to the sedimentation velocity of PhaF. (B) Sedimentation equilibrium analysis of the association state of PhaF. Sedimentation equilibrium absorbance gradients were carried out at 10,000 r.p.m. (open circles) and 16,000 r.p.m. (closed circles). The solid lines show the corresponding best-fit gradients for a single sedimenting species at equilibrium with a solution average molar mass of 85000 Da. The residuals (difference between the experimental data and the fitted data for each point) are shown at the bottom of this panel.  
doi:10.1371/journal.pone.0056904.g003



**Figure 4. TFE titrations of PhaF and C-PhaF.** (A) Far-UV CD spectra of PhaF in the absence (solid line) and presence (dashed line) of 40% TFE. (B) Far-UV CD spectra of C-PhaF in the absence (solid line) and presence (dashed line) of 40% TFE. (C) CD-monitored TFE titration of PhaF (closed circles) and C-PhaF (open circles) following the ellipticity signal at 222 nm. (D), same as in (C) but following the signal at 208 nm. doi:10.1371/journal.pone.0056904.g004

Moreover, Fig. 5D and Table 2 also show that the transition is shifted to lower temperatures as the protein concentration is decreased 10-fold, indicating that the thermal denaturation involves a change in molarity. This, together with the coincidence of the far-UV and near-UV CD scans using the same protein concentration (Fig. 5D), suggests a two-state unfolding-dissociation coupled transition in which the predominant protein tetramer leads to unfolded monomers upon heating, without the accumulation of intermediate species, as indicated in Eq. 2. The hydrophobic core would therefore be conformed by, at least, the packed leucine zipper sequences. On the other hand, the transitions were not fully reversible after sample incubation at 90 °C (Figs. 5A and 5B), but the recovery of the signal substantially improved when the samples were cooled down right after the thermal transition was completed (60 °C) (Fig 5A). This indicates that any irreversible step takes place only once the protein is fully unfolded, so that the equilibrium shown in Eq. 2 is not affected by these post-transitional events and therefore can be analyzed by equilibrium thermodynamics. This is a particular case of the Lumry-Eyring model [54] and has been extensively applied to the analysis of the irreversible unfolding of proteins similar to our case [55,56]. To ensure this, we carried out a temperature scan at half the heating rate (0.5 K min<sup>-1</sup>) and found very little difference with

that acquired at 1 K min<sup>-1</sup> (Fig S4) and yielding similar thermodynamic parameters (Table 2, see below). This demonstrates that any kinetic effect on the thermal denaturation of the protein is negligible. It is noteworthy that, at the slower heating rate, there occurs an appreciable loss of ellipticity signal starting at about 72 °C, which was not detected at the faster heating rate (Fig S4). This reinforces our idea that incubation of the protein at high temperatures causes aggregation and/or other irreversible changes in a relatively slow time scale.

According to the analytical ultracentrifugation experiments shown above, the minor species that accompany the predominant tetramer are not in fast equilibrium among them, so they should contribute very little to the equilibrium thermodynamic analysis. Table 2 displays the thermodynamic quantities of the thermal transitions as analyzed by the Gibbs-Helmholtz equation (Eqs 3–5). Since  $\Delta C_p$  is unknown for PhaF, we tried to obtain its value from the slope of the plot of  $\Delta H_m$  versus  $T_m$  from thermograms recorded at different pH's using the procedure described by Swint and Robertson [57]. However, the protein was greatly destabilized on decreasing pH (see below), with very short pre-transitional baselines, so that curve fittings yielded poor results (data not shown). Alternatively, we employed two estimates of  $\Delta C_p$ . The extended nature of the predicted structure of PhaF points to a



**Table 1.** Secondary structure quantitation of PhaF based in CD data.

Sample	Method	$\alpha$ -Helix (%)	$\beta$ -Structures (%)	Remainder (%)
PhaF	CDNN	32	18	51
	CONTIN	29	7	64
	CDSSTR	32	11	57
PhaF (40% TFE)	CDNN	42	12	46
	CONTIN	47	7	46
	CDSSTR	55	18	27
C-PhaF	CDNN	N.D. <sup>a</sup>	N.D.	N.D.
	CONTIN	1	4	95
	CDSSTR	1	13	86
C-PhaF (40% TFE)	CDNN	N.D.	N.D.	N.D.
	CONTIN	9	7	84
	CDSSTR	5	8	87
PhaF (whole protein)	Model	44	0	57
	Model (accounting for disorder) <sup>b</sup>	31	0	69

Estimations using the CDNN, CONTIN and CDSSTR procedures, and comparison with that predicted from the three-dimensional model.

<sup>a</sup>N.D.: Not determined (poor fitting)

<sup>b</sup>Calculations assuming that the helical-predicted stretches 13–17 and 58–91 are disordered in solution in the unligated protein (according to PONDR-FIT procedure). doi:10.1371/journal.pone.0056904.t001

predictably low  $\Delta C_p$  value, directly related to the increment in accessible surface area upon unfolding [58]. We firstly estimated for our calculations a value of  $0.7 \text{ kcal mol}^{-1} \text{ K}^{-1}$ , that has been determined experimentally for the designed 31-aa AB<sub>SS</sub> leucine zipper [59] (the predicted leucine zipper sequence of PhaF spans 22 residues – see Fig. 1A). Nevertheless, even lower values have been reported for other coiled-coils [60]. In any case, even estimating  $\Delta C_p = 0 \text{ kcal mol}^{-1} \text{ K}^{-1}$ , the calculated free energies of unfolding, extrapolated to  $10^\circ\text{C}$ , barely differ in 10%, that lies within the experimental error (Table 2).

A thermal stability analysis of the isolated C-PhaF was carried out by monitoring the effect of the temperature on its far-UV CD spectrum. Fig. 5C shows that at high temperature ( $90^\circ\text{C}$ ) the intensity of the spectrum substantially decreases, a change that is reversible upon cooling down the sample. While some negative ellipticity also appears in the 215–235 nm range at  $90^\circ\text{C}$ , this should not be necessarily taken as indicative of residual structure, since Privalov *et al.* [61] noticed similar changes in completely unfolded peptides and ascribed this phenomenon to non-specific effects of temperature on the CD spectrum. An alternative/additional explanation to this phenomenon arises from the analysis of the difference spectrum of C-PhaF between  $20^\circ\text{C}$  and  $90^\circ\text{C}$  (a strong minimum at 193 nm and a maximum at 220 nm, Fig. 5C inset), which suggests a loss of poly(proline)-II structure upon denaturation [52]. In any case, Fig 5D (inset) displays a non-cooperative loss of CD signal with increasing temperature, which is typical of peptides devoid of fixed tertiary structure, thus demonstrating the absence of a significant hydrophobic core and suggesting that the C-terminal domain of PhaF is essentially unfolded in the absence of its DNA ligand.

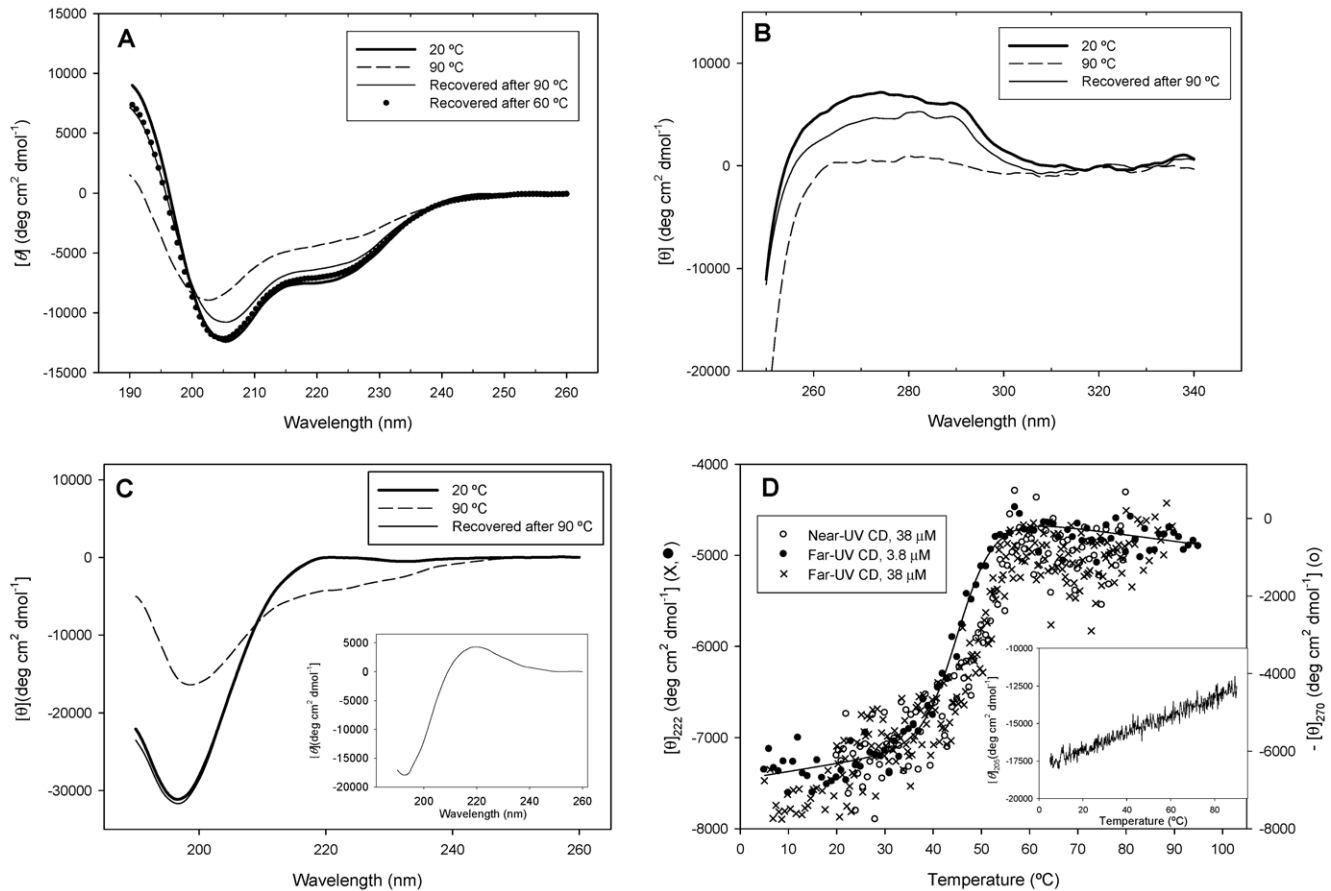
### Chemical unfolding

PhaF ( $3.8 \mu\text{M}$ , total monomer concentration) was subjected to equilibrium denaturation by urea monitored by far-UV CD. Experiments were carried out at low temperature ( $10^\circ\text{C}$ ) in order to acquire reliable baselines since the stability of the protein is only marginal at room temperature (Fig. 5D). As shown in Fig. 6, the

protein was fully unfolded in a cooperative way using relatively low concentrations of denaturant even at  $10^\circ\text{C}$ . The thermodynamics of the transition was analyzed by the linear extrapolation method (Eqs. 5 and 6), yielding the following fitting parameters:  $m = 2.1 \pm 0.1 \text{ kcal mol}^{-1} \text{ M}^{-1}$ ;  $[\text{urea}]_{1/2} = 1.5 \pm 0.1 \text{ M}$ ;  $\Delta G_{\text{eff}}^\circ = 3.2 \pm 0.2 \text{ kcal mol}^{-1}$ ; and  $\Delta G_{\text{int}}^\circ = 22.2 \pm 0.2 \text{ kcal mol}^{-1}$ . This value is in good agreement with those calculated from the thermal unfolding experiments (Table 2). When the protein concentration increased 3-fold ( $11.4 \mu\text{M}$ , total monomer concentration), the urea midpoint also shifted to higher concentrations (Fig. 6), showing again a dependence on protein concentration in agreement with the change in molarity upon unfolding. In this case, the thermodynamic parameters were:  $m = 1.4 \pm 0.1 \text{ kcal mol}^{-1} \text{ M}^{-1}$ ;  $[\text{urea}]_{1/2} = 2.7 \pm 0.1 \text{ M}$ ;  $\Delta G_{\text{eff}}^\circ = 3.8 \pm 0.2 \text{ kcal mol}^{-1}$ ; and  $\Delta G_{\text{int}}^\circ = 21.0 \pm 0.2 \text{ kcal mol}^{-1}$ . Finally, lack of structure in the C-terminal region of the protein was confirmed by the linear, non-cooperative change in CD signal of the C-PhaF polypeptide (Fig. 6B).

### pH stability

The structural integrity of PhaF was assessed at different pH's by far-UV CD and intrinsic fluorescence spectroscopies (Figs. 7 A–C). Fig. 7A shows the reduction of the 206 nm and 222 nm minima and the appearance of a band at 200 nm upon decreasing the pH, reflecting the loss of  $\alpha$ -helical structure in favour of random-coil conformations. The existence of an isodichroic point at 204 nm points to a two-state transition, without detectable intermediates. Plotting the far-UV CD signal *vs* pH shows a low-intensity, linear change in CD signal between pH 9.5 and 4.0, and an incomplete transition starting below pH 4.0 (Fig. 7B). This indicates that the protonation of one or more aspartate or glutamate side chains in the N-terminal domain of PhaF is responsible for the pH stability of the phasin. On the other hand, the tryptophan fluorescence spectrum (Fig. 7C) displays a maximum at 336 nm that is shifted at lower pH's to higher wavelengths (343 nm), concomitantly with a decrease in fluorescence intensity. This is in accordance with tryptophan exposure to



**Figure 5. Thermal stability of PhaF.** (A) Far-UV CD spectra of PhaF at different temperatures. Recovered spectra were registered upon immediate cooling down the scanned sample and after a 30 min waiting period at 20 °C. (B) Near-UV CD spectra of PhaF. (C) Far-UV CD spectra of the isolated C-terminal domain of PhaF (C-PhaF protein). *Inset*, difference CD spectrum of C-PhaF (20°C–90 °C). (D) CD-monitored temperature scans of PhaF. Solid line indicates fitting of the far-UV CD-monitored transition (protein concentration: 3.8 μM) to the Gibbs-Helmholtz equation (see Materials and Methods). *Inset*, temperature scan of C-PhaF monitored at 205 nm. doi:10.1371/journal.pone.0056904.g005

the solvent arising from protein unfolding. Upon examining the change on the average emission intensity ( $\langle\lambda\rangle$ ) (Fig. 7B), the low-pH CD-monitored transition was reproduced, but additionally a second transition appeared centered at pH 5.2. Nevertheless, since no significant changes were observed in this pH range in terms of secondary structure (Fig. 7A) and overall fluorescence intensity (Fig. 7B and 7C), we believe that this additional transition arises from the protonation of some acidic group(s) that locally affect the

polarity of the environment around the tryptophan residues, but without inducing any appreciable conformational change in the protein. In fact, plotting the raw fluorescence intensity only yields a single transition similar to the CD-monitored one (Fig. 7B). Finally, as all acidic residues of PhaF except one (Glu-190), are located in the N-terminal domain, it should be expected that the pH-induced conformational changes were also restricted to this

**Table 2. Thermodynamic quantities of the thermal denaturation of PhaF<sup>d</sup>.**

Probe	Monomer concentration (μM)	$t_m$ (°C)	$\Delta H_m$ (kcal/mol)	$\Delta G_{int}^0$ (kcal/mol) <sup>b,c</sup>	$\Delta G_{int}^0$ (kcal/mol) <sup>b,d</sup>
Far-UV CD	3.8	45.5±0.2	53.8±2.3	25.0±0.3	23.5±0.3
Far-UV CD	3.8 <sup>e</sup>	46.1±0.4	44.2±2.8	24.0±0.6	22.5±0.6
Far-UV CD	38.0	50.8±0.3	103.6±13.5	28.1±1.8	26.2±1.8
Near-UV CD	38.0	51.5±0.3	95.4±12.4	27.3±1.6	25.4±1.6

<sup>a</sup>Heating rate: 1 K min<sup>-1</sup> unless otherwise stated

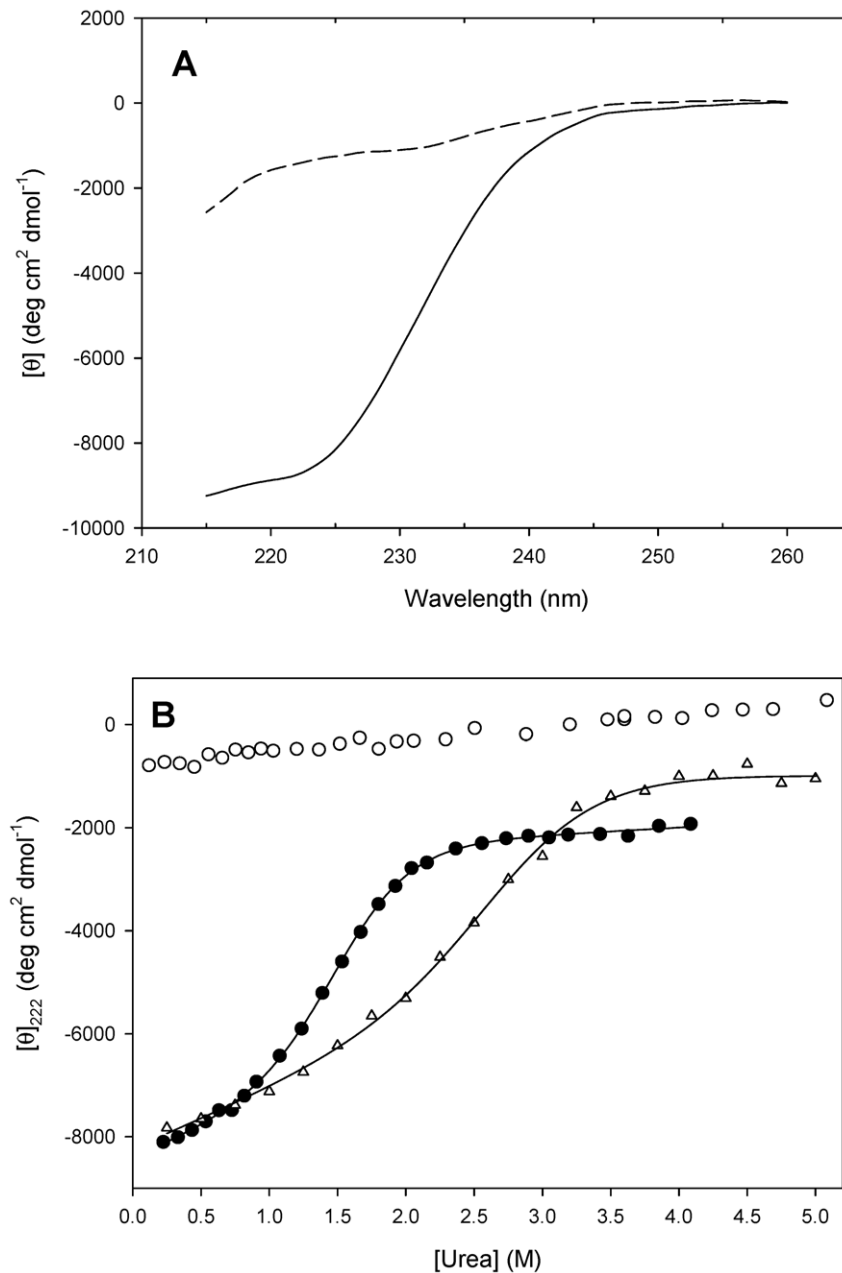
<sup>b</sup>Data calculated for a temperature of 10 °C

<sup>c</sup>Assuming  $\Delta C_p = 0$  kcal mol<sup>-1</sup> K<sup>-1</sup>

<sup>d</sup>Assuming  $\Delta C_p = 0.7$  kcal mol<sup>-1</sup> K<sup>-1</sup>

<sup>e</sup>Heating rate: 0.5 K min<sup>-1</sup>

doi:10.1371/journal.pone.0056904.t002



**Figure 6. Chemical stability of PhaF.** (A) Far-UV CD spectra of PhaF in the absence (solid line) and the presence (dashed line) of 4.1 M urea. (B). CD-monitored urea titration of 3.8  $\mu\text{M}$  PhaF (closed circles), 11.4  $\mu\text{M}$  PhaF (open triangles) and C-PhaF (open circles). The solid lines represents the fitting to a two-state equilibrium denaturation transition (see text). doi:10.1371/journal.pone.0056904.g006

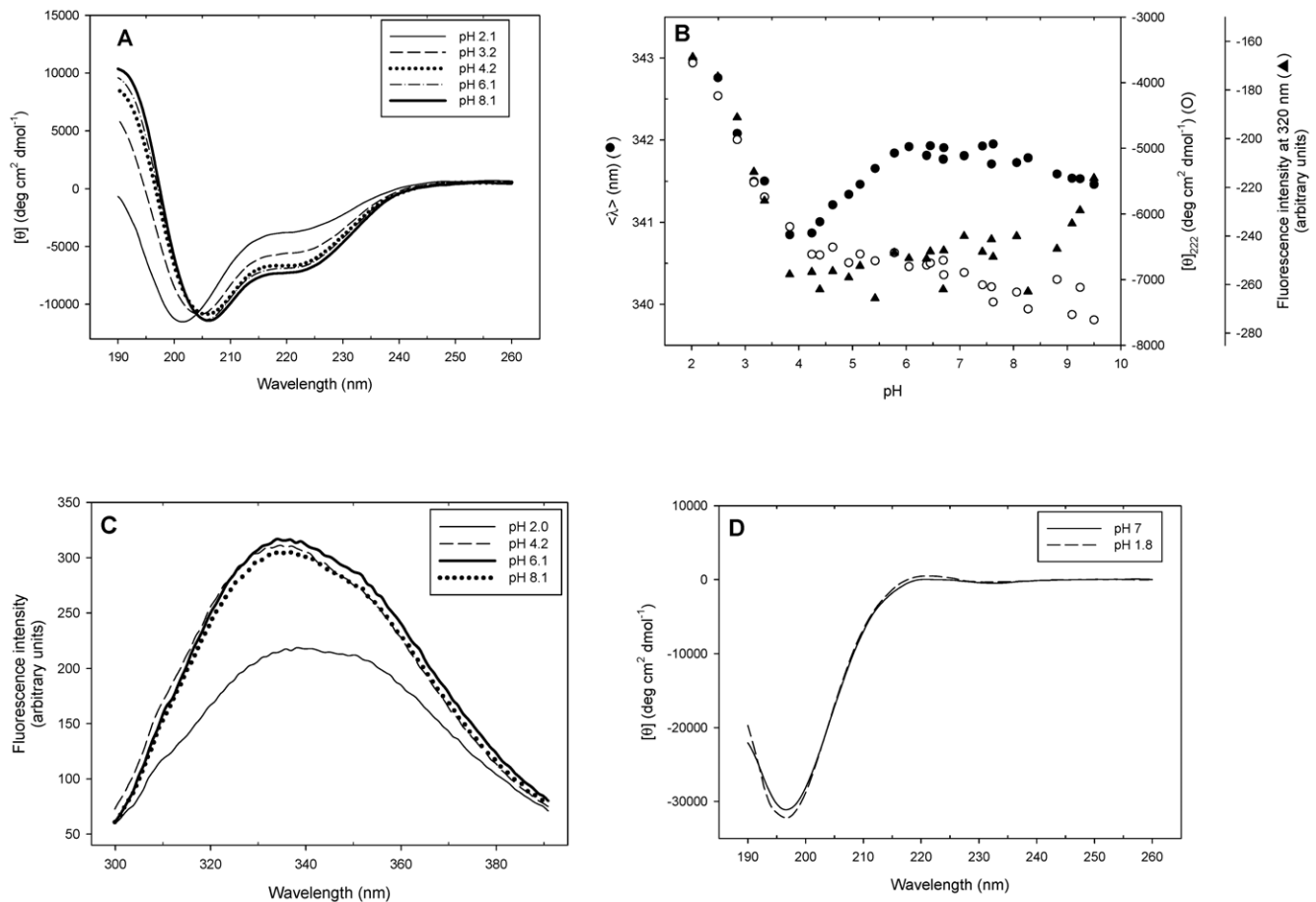
region of the protein: lack of CD spectral changes in the C-PhaF protein at different pH's (Fig. 7D) supports this hypothesis.

## Discussion

The traditional concept of proteins as rigid, compact entities, the flexibility of which is limited to moderate conformational rearrangements necessary to carry out any molecular recognition events, is being currently challenged with the increasing description of polypeptides lacking defined secondary and tertiary structures, the so-called intrinsically disordered proteins [62]. Far from being curious exceptions, intrinsic disorder is a widespread mechanism allowing a polypeptide to suitably adapt itself to its

biological role, and to carry out very important physiological functions including regulation, signalling, chaperoning and nucleic acid binding [63,64].

Phasins interact with a hydrophobic, lipid-coated insoluble substrate (the polyhydroxyalkanoate intracellular granule), providing it with a surface layer that avoids granule coalescence. Other functions proposed for phasins are: controlling the number and surface of granules, protecting the host cell by contributing to coverage of the hydrophobic surface of the polymer, preventing protein misfolding on the hydrophobic granule [3] or serving as a storage source of nitrogen [14]. In many occasions, stabilization of the PHA granule must simultaneously coincide with other biological functions. In the case of PhaF, it has been demonstrated



**Figure 7. pH stability of PhaF.** (A) Far-UV CD spectra of PhaF registered at different pH's. (B) pH titration of PhaF monitored by far-UV CD (open circles), average tryptophan fluorescence emission intensity (closed circles) and fluorescence intensity at 320 nm (triangles). (C) Intrinsic tryptophan fluorescence spectra of PhaF at different pH's. (D) Far-UV CD spectra of C-PhaF.  
doi:10.1371/journal.pone.0056904.g007

that this phasin binds DNA non-specifically and plays an essential role in cytoplasmic granule localization and segregation into daughter cells upon cell division [18]. Such different functions can only be supported by a peculiar PhaF structure distributed in separated domains that recognize their different macromolecular substrates (PHA and DNA). According to their structural characteristics, phasins are good candidates for becoming part of the intrinsically disordered protein family [62], as their mean hydrophobicity is low (reducing the possibility of an appreciable hydrophobic core) and there is a high net charge repulsion, especially in the C-terminal domain (Fig. 1C; Fig. S2). Moreover, their regulatory biological role demands a high degree of plasticity to simultaneously adjust to the structure of such macromolecules as chromosomal DNA and PHA granules.

Given the specific bifunctionality of PhaF not previously found in other proteins, it is hardly surprising that no homologs have been found yet in the databases of proteins with known 3-D structure, from which a structural model of PhaF could be derived. However, the neat separation into domains that can be discerned from the inspection of its sequence [17] and experimental evidence [18,27,28] (see Introduction) prompted us to model such domains separately, and subsequently link the resulting structures. Then, the model was subjected to experimental validation using hydrodynamic and spectroscopical techniques. Fig. 1E displays the final structural model of a PhaF monomer, based in secondary

structure predictions (N-terminal domain and C-terminal tail, Figs. 1A and D) and sequence similarity with the AlgP regulator (C-terminal domain, Fig. 1C). According to our model, the protein has an extended shape, in which the N-terminal, PHA-binding domain maintains little, if any, interaction with the C-terminal, DNA-binding moiety. This separation of domains ensures that the bifunctionality of the protein is accomplished without any spatial hindrance.

The N-terminal domain is predicted to contain a long, amphipathic  $\alpha$ -helix (Fig. S2) which is partially disordered in the absence of PHA (Fig. 2), although this partial unfolding may be reverted in the presence of hydrophobic PHA mimics such as sodium oleate (Fig. S3) or structure-stabilizing cosolvents such as TFE (Fig. 4A). A disorder-to-order transition upon ligand binding is a common characteristic of intrinsically disordered proteins [62]. The long N-terminal helical stretch is followed by a short coiled-coil sequence, possibly a leucine zipper (Figs. 1A and B), that is believed to be involved in the tetramerization of the protein as unveiled by the ultracentrifugation experiments (Fig. 3). Circular dichroism spectroscopy provides experimental support for the helical nature of the N-terminal domain (Fig. 4A, Table 1). Moreover, the dependence of the thermal and chemical stability on protein concentration (Figs. 5D and 6B, Table 2) is in agreement with the oligomeric nature of PhaF. We have not detected in the literature any references so far to coiled-coil sequences in polyhydroxyalk-

anoate-associated proteins, so that, to the best of our knowledge, this might be the first documented case of such an oligomerization motif in phasins. In order to check the distribution of coiled-coil sequences among other polyhydroxyalkanoate-associated proteins, we analyzed a total of 973 sequences belonging to the UniProtKB database containing the term "phasin". The COILS utility identified 499 sequences (51.3% of the database) with a probability higher than 75% of containing coiled-coil stretches (Table S1), plus an additional 188 sequences with a probability between 25% and 75% (19.4% of the database, Table S2). The best hits encompass a wide variety of organisms and gene lengths. The fact that approximately 70% of the analyzed phasin sequences show hints of coiled-coils to an appreciable degree is remarkable and points to this motif as a widespread mechanism for the oligomerization of PHA-associated proteins. It is noteworthy that the accompanying PhaI phasin in *P. putida* KT2440 holds a predicted coiled-coil sequence within its primary structure as well (Table S1).

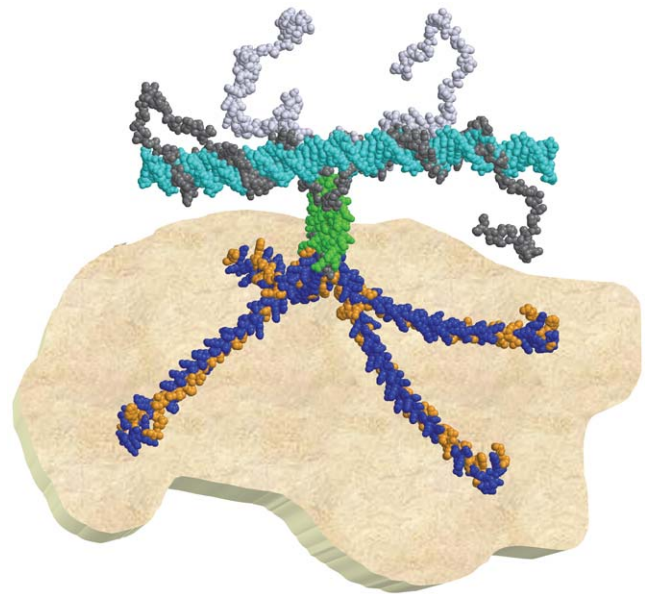
With respect to the C-terminal moiety of the protein, the model assumes a superhelical structure based on that of the AlgP regulatory protein (Figs. 1C and 1E; Fig. S2) with basic residues adapted to electrostatically interact with the phosphate backbone of DNA. However, this structure should be only stable upon binding to the nucleic acid, thus overcoming the otherwise strong ionic repulsion between the lysine residues. A coincident prediction by several procedures (Fig. 2) supported by experimental data such as a mostly featureless CD spectra (Fig. 4B), a moderate response to the addition of high concentrations of helix-inducing TFE (Fig. 4B) and the absence of cooperative unfolding transitions induced by heat (Fig. 5D, inset), urea (Fig. 6B) or pH (Fig. 7D) suggest that the C-terminal domain is indeed natively unfolded in solution in the absence of DNA, a fact that nevertheless does not prevent the domain to recognize its polyanionic ligand [18] (Fig. S3). This is a similar case than that of histones, an intrinsically disordered protein family for which its disorder is absolutely essential to carry out their function [65]. With respect to the 35-aa C-terminal tail that follows the DNA-binding moiety of PhaF, it lacks any basic residues and the structure is unequivocally predicted as random coil by all methods tested (Fig. 1D). However, the difference CD spectrum of C-PhaF at 20°C and 90°C (Fig. 5C, inset) suggests that some poly(Pro)-II conformation is present in the protein, a secondary structure not taken into account in most predictors. The 35-aa tail intriguingly contains repeated sequences with the PXXP signature typical of SH3-binding motifs [66] and that has been found within poly(proline)-II helices [67]. Moreover, a FASTA search with this sequence revealed that similar stretches are found in many other polypeptides (data not shown), among them other phasins and related proteins such as the above mentioned AlgP [68], and even the H1 histone. This suggests that, despite this apparent lack of structure, this C-terminal tail must fulfil a relevant biological role that is unknown at the moment.

The intrinsic disorder detected in the C-terminal part of PhaF and in a part of the N-terminal moiety prompted us to check, in the above mentioned UniProtKB database, whether this could be also a characteristic shared by other phasins. When the database was subjected to the PONDR-FIT analysis, we found that a remarkable content of disorder propensity was unveiled in most of the cases (data not shown). In order to set a more formal definition, we arbitrarily selected those sequences displaying more than 50% of the residues with a disorder disposition higher than 0.5 according to PONDR-FIT, and the results are shown in Table S3. Out of 973 sequences, a remarkable 24% (234 sequences) complied with our selective disorder threshold. We therefore

hypothesize that an important subset of phasins may constitute a new family of intrinsically disordered proteins not described so far.

Upon inspection of the model, it follows that most of the unligated PhaF stability should be acquired through oligomerization and that the protein mainly acquires its stability from quaternary-, rather than tertiary-structure contacts. In this sense, thermal and chemical denaturations can be treated as tetramer-monomer coupled unfolding-dissociation equilibria, yielding a folding energy of around 25 kcal mol<sup>-1</sup> (Figs. 5 and 6; Table 2). This value lies within the range of other proteins that oligomerize through coiled coils of the same length, such as the 21-aa Lac21E/K heterotetramer (22.4 kcal mol<sup>-1</sup>) [69].

All our results, together with previous biochemical data [18] allow us to delineate a structural model of the PHA/phasin/DNA biological entity (Fig. 8). In this scheme, the hydrophobic face of the amphipathic N-terminal helices are associated to the PHA granule, either through its lipid coating or even to the naked polyester itself (unpublished results). The positive interaction unveiled from the conformational change induced in PhaF by PHA mimics such as oleate (Fig. S3) are in accordance with this hypothesis. In turn, the hydrophilic side faces the solvent, thus providing the granule surface a polar character that prevents granule coalescence or non-specific protein association through hydrophobic interactions. On the other hand, the C-terminal domain binds to the chromosomal DNA (and perhaps to cytoskeletal-like proteins) to facilitate the needle-type array granule structure at cellular centre, and linking DNA replication and granule uniform distribution to the daughter cells upon cellular division [18]. Given the presumably low, if any, interdomain interactions, the oligomeric state of PhaF in solution is likely to be conserved when bound to PHA and DNA, providing a multipoint attachment system whose affinity for the macromolecular ligands



**Figure 8. A model of the interaction between PHA granules, phasins and chromosomal DNA in *Pseudomonas putida* KT2440.** A PhaF tetramer is depicted interacting with the PHA granule while also attached to a fragment of nucleoid DNA. The hydrophobic residues in the N-terminal domain are shown in orange, while the polar ones are colored blue. The leucine zipper is colored green, and the DNA is depicted in cyan. Figure was rendered with RASTOP 2.2 (<http://www.geneinfinity.org/rastop/>) only for visualization purposes, so that no atomic detail is intended.

doi:10.1371/journal.pone.0056904.g008

would be greatly enhanced compared to the monomeric situation thanks to multivalent effects. This could be one of the reasons why the coiled-coil (oligomerization) motif is so widely spread among phasins (Tables S1 and S2). Furthermore, the fact that the PhaI phasin, a companion of PhaF associated to the PHA granule, might also contain a coiled-coil motif, suggests that heterodimers/heterotetramers are also likely to happen, although experimental confirmation of this point must await until suitable amounts of PhaI protein can be purified.

To sum up, we have developed a structural model of the PhaF phasin from *Pseudomonas putida* KT2440 that has been experimentally tested and that may provide a molecular basis for its biological role in PHA granule stabilization, solvation, intracellular arrangement and equal distribution to daughter cells. Our hypothesis contemplates a tetrameric, intrinsically disordered protein in the absence of its ligands PHA and DNA, that acquires a higher degree of secondary structure when interacting with its binding partners. This disorder-order transition is typical of most of the intrinsic disordered proteins described so far [62]. Our results may help to understand the function of these important proteins involved in the metabolism of the PHA bioplastics, and could also be of utility to obtain improved variants of the BioF affinity tag [27,28]. Moreover, some structural features of PhaF are shared among a high number of other phasins, even from unrelated organisms, namely the conservation of coiled-coil oligomerization sequences and an appreciable level of intrinsic disorder, evidences that deserve further investigations.

## Supporting Information

**Figure S1 Purification of proteins.** (A) Purification of PhaF by butyl sepharose. Lanes: M, molecular weight markers (General Electric Healthcare); 1, total extract of *Escherichia coli* BL21(DE3) [pETPhaF]; 2 soluble fraction of the extract; 3, inclusion bodies solubilized and refolded in 20 mM sodium phosphate; 4, flowthrough upon application on a buthyl-sepharose column; 5-7, fractions of protein purified upon elution with 20 mM sodium phosphate pH 7.0. (B) Purification of the carboxy-terminal domain of PhaF (C-PhaF) by ionic exchange chromatography. Lanes: M, molecular weight markers; 1, total extract of *E. coli* BL21(DE3) [pETCterm]; 2, total extract from *E. coli* BL21(DE3) harbouring the pET-29a(+) plasmid without insert as a control; 3-5, eluted fractions of purified C-PhaF protein upon elution with 0.7 M Tris-HCl buffer, pH 8.8, plus 4 M NaCl. (PDF)

**Figure S2 Structural features of the modeled N- and C-terminal domains of PhaF.** (A) two spacefill views of the N-

terminal moiety, showing polar (blue) and hydrophobic (yellow) residues. (B) a detail of the N-terminal domain showing acidic (red) and basic (cyan) side chains. (C) two spacefill views of the C-terminal moiety, highlighting the basic side chains in blue. (PDF)

**Figure S3 Effect of compounds on PhaF structure.** (A) Far-UV CD spectra of PhaF in 20 mM sodium phosphate buffer in the absence (solid line) and the presence (dashed line) of 1 mM sodium oleate. (B), far-UV CD spectra of C-PhaF in 20 mM sodium phosphate buffer in the absence (solid line) and the presence (dashed line) of 9 mM nspDNA. *Inset*, difference spectrum calculated from those of free and DNA-bound samples. (PDF)

**Figure S4 Effect of heating rate on PhaF thermal denaturation monitored by far-UV CD.** Heating rate was set to 0.5 K min<sup>-1</sup> (open circles) and 1 K min<sup>-1</sup> (closed circles). Solid line indicates fitting of the 0.5 K min<sup>-1</sup> transition (protein concentration: 3.8 μM) to the Gibbs-Helmholtz equation (see Materials and Methods). Ellipticities above 72 °C were not taken into account for the fitting as they may reflect late aggregation events. (PDF)

**Table S1 Phasins from the UniProtKB database with >75% probability of containing coiled-coil sequences according to the COILS procedure.** (XLS)

**Table S2 Phasins from the UniProtKB database with probability of containing coiled-coil sequences between 25%-75% according to the COILS procedure.** (XLS)

**Table S3 Phasins from the UniProtKB database with intrinsic disorder probability according to the PONDR-FIT procedure.** (XLS)

## Acknowledgments

We thank M. Garzón, M. Gutierrez, C. Fuster and J. Casanova for excellent technical assistance.

## Author Contributions

Contributed with ideas: CA GR BG MAP. Conceived and designed the experiments: BM JMS. Performed the experiments: BM BG CA. Analyzed the data: BM BG CA GR MAP JMS. Contributed reagents/materials/analysis tools: GR MAP JMS. Wrote the paper: BM JMS.

## References

- Findlay RH, White DC (1983) Polymeric beta-hydroxyalkanoates from environmental samples and *Bacillus megaterium*. *Appl Environ Microbiol* 45: 71–78.
- de Smet MJ, Eggink G, Witholt B, Kingma J, Wynberg H (1983) Characterization of intracellular inclusion formed by *Pseudomonas oleovorans* during growth on octane. *J Bacteriol* 154: 870–878.
- Steinbüchel A, Aerts K, Babel W, Föllner C, Liebergesell M et al. (1995) Considerations on the structure and biochemistry of bacterial polyhydroxyalkanoic acid inclusions. *Can J Microbiol* 41: 94–105.
- García B, Olivera ER, Miñambres B, Fernández-Valverde M, Cañedo LM et al. (1999) Novel biodegradable aromatic plastics from a bacterial source. Genetic and biochemical studies on a route of the phenylacetyl-coA catabolon. *J Biol Chem* 274: 29228–29241.
- Madison LL, Huisman GW (1999) Metabolic engineering of poly(3-hydroxyalkanoates): from DNA to plastic. *Microbiol Mol Biol Rev* 63: 21–53.
- Prieto MA, de Eugenio LI, Galán B, Luengo JM, Witholt B (2007) Synthesis and degradation of polyhydroxyalkanoates. In: Ramos JL, Filloux A, editors. *Pseudomonas: a model system in biology*. New York: Springer. pp. 397–428.
- Pötter M, Steinbüchel A (2005) Poly(3-hydroxybutyrate) granule-associated proteins: impacts on poly(3-hydroxybutyrate) synthesis and degradation. *Biomacromolecules* 6: 552–560.
- Steinbüchel A, Fuchstenbusch B (1998) Bacterial and other biological systems for polyester production. *Trends Biotechnol* 16: 419–427.
- Zinn M, Witholt B, Egli T (2001) Occurrence, synthesis and medical application of bacterial polyhydroxyalkanoate. *Adv Drug Rev* 55: 5–21.
- Williams SF, Martin D (2002) Application of PHAs in medicine and pharmacy. In: Doi Y, Steinbüchel A, editors. *Biopolymers, vol 4. Polyesters III: Applications and Commercial Products*. Germany: Wiley-VCH. pp 91–128.
- Gao X, Chen J-C, Wu Q, Chen G-Q (2010) Polyhydroxyalkanoates as a source of chemicals, polymers, and biofuels. *Curr Op Biotechnol* 22: 768–774.
- Mayer F, Madkour MH, Pieper-Furst U, Wiczeorek R, Liebergesell M et al. (1996) Electron microscopic observation on the macromolecular organization of the boundary layer of bacterial PHA inclusion bodies. *J Gen Appl Microbiol* 42: 445–455.
- Wiczeorek R, Pries A, Steinbüchel A, Mayer F (1995) Analysis of a 24-kilodalton protein associated with the polyhydroxyalkanoic acid granules in *Alcaligenes eutrophus*. *J Bacteriol* 177: 2425–2435.



14. McCool GJ, Cannon MC (1999) Polyhydroxyalkanoate inclusion body-associated proteins and coding region in *Bacillus megaterium*. *J Bacteriol* 181: 585–592.
15. Pieper-Fürst U, Hussein Madkour M, Mayer F, Steinbüchel A (1994) Purification and characterization of a 14-kilodalton protein that is bound to the surface of polyhydroxyalkanoic acid granules in *Rhodococcus ruber*. *J Bacteriol* 176: 4328–4337.
16. Maehara A, Ueda S, Nakano H, Yamane T (1999) Analyses of a polyhydroxyalkanoic acid granule-associated 16-kilodalton protein and its putative regulator in the *pha* locus of *Paracoccus denitrificans*. *J Bacteriol* 181: 2914–2921.
17. Prieto MA, Bühler B, Jung K, Witholt B, Kessler B (1999) PhaF, a polyhydroxyalkanoate-granule-associated protein of *Pseudomonas oleovorans* GPo1 involved in the regulatory expression system for *pha* genes. *J Bacteriol* 181: 858–868.
18. Galán B, Dijanki N, Maestro B, de Eugenio LI, Escapa IF et al. (2011) Nucleoid-associated PhaF phasin drives intracellular location and segregation of polyhydroxyalkanoate granules in *Pseudomonas putida* KT2442. *Mol Microbiol* 79: 402–418.
19. Sandoval A, Arias-Barrau E, Arcos M, Naharro G, Elías R et al. (2007) Genetic and ultrastructural analysis of different mutants of *Pseudomonas putida* affected in the poly-3-hydroxy-n-alkanoate gene cluster. *Environ Microbiol* 9: 737–751.
20. Schembri MA, Woods AA, Bayly RC, Davies JK (1995) Identification of a 13-kDa protein associated with the polyhydroxyalkanoic acid granules from *Acinetobacter* spp. *FEMS Microbiol Lett* 133: 277–283.
21. Jendrossek D (2009) Polyhydroxyalkanoate granules are complex subcellular organelles (carbonosomes) *J Bacteriol* 191: 3195–3202.
22. York GM, Junker BH, Stubbe J, Sinskey AJ (2001) Accumulation of the PhaP phasin of *Ralstonia eutropha* is dependent on production of polyhydroxybutyrate in cells. *J Bacteriol* 183: 4217–4226.
23. York GM, Stubbe J, Sinskey AJ (2002) The *Ralstonia eutropha* PhaR protein couples synthesis of the PhaP phasin to the presence of polyhydroxybutyrate in cells and promotes polyhydroxybutyrate production. *J Bacteriol* 184: 59–66.
24. Luengo JM, García B, Sandoval A, Naharro G, Olivera ER (2003) Bioplastics from microorganisms. *Curr Op Microbiol* 6: 251–260.
25. Kuchta K, Chi L, Fuchs H, Pötter M, Steinbüchel A (2007) Studies on the influence of phasins on accumulation and degradation of PHB and nanostructure of PHB granules in *Ralstonia eutropha* H16. *Biomacromolecules* 8: 657–662.
26. Qi Q, Steinbüchel A, Rehm BH (2000) *In vitro* synthesis of poly(3-hydroxydecanoate): purification and enzymatic characterization of type II polyhydroxyalkanoate synthases PhaC1 and PhaC2 from *Pseudomonas aeruginosa*. *Appl Microbiol Biotechnol* 54: 37–43.
27. Moldes C, García P, García JL, Prieto MA (2004) *In vivo* immobilization of fusion proteins on bioplastics by the novel tag BioF. *Appl Environ Microbiol* 70: 3205–3512.
28. Moldes C, Farinós GP, de Eugenio LI, García P, García JL et al. (2006) New tool for spreading proteins to the environment: Cry1Ab toxin immobilized to bioplastics. *Appl Microbiol Biotechnol* 72: 88–93.
29. Rost B, Yachdav G, Liu J (2004) The PredictProtein Server. *Nucleic Acids Res* 32 (Web Server issue), W321–W326.
30. Cole C, Barber JD, Barton GJ (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* 35 (Web server issue), W197–W201.
31. Guermeur Y, Geourjon C, Gallinari P, Deleage G (1999) Improved performance in protein secondary structure prediction by inhomogeneous score combination. *Bioinformatics* 15: 413–421.
32. Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. *Science* 252: 1162–1164.
33. Delorenzi M, Speed T (2002) An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* 18: 617–625.
34. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18: 2714–2723.
35. Medvedkin VN., Permyakov EA, Klimenko LV, Mitin YV, Matsushima N et al. (1995) Interactions of (Ala\*Ala\**Lys*\*Pro)<sub>n</sub> and (Lys\**Lys*\*Ser\*Pro)<sub>n</sub> with DNA. Proposed coiled-coil structure of AlgR3 and AlgP from *Pseudomonas aeruginosa*. *Protein Eng* 8: 63–70.
36. Xue B, DunBrack RL, Williams RW, Dunker AK, Uversky VN (2010) PONDR-Fit: A meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta* 1804: 996–1010.
37. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* 337: 635–645.
38. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ et al. (2003) Protein disorder prediction: implications for structural proteomics. *Structure* 11: 1453–1459.
39. Cheng J, Sweredoski M, Baldi P (2005) Accurate prediction of protein disordered regions by mining protein structure data. *Data mining and knowledge discovery* 11: 213–222.
40. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys J* 78: 1606–1619.
41. van Holde KE (1986) Sedimentation in physical biochemistry. In: *Physical Biochemistry*. Englewood Cliffs, NJ: Prentice Hall, Inc. pp.3–23.
42. Laue TM., Shah BD, Ridgeway TM, Pelletier SL (1992) Computer-aided interpretation of analytical sedimentation data for proteins. In: Harding S, Rowe A, Horton J, editors. *Analytical Ultracentrifugation in Biochemistry and Polymer Science*. Cambridge, UK: Royal Society of Chemistry. pp. 90–125.
43. Minton AP (1994) Conservation of signal: a new algorithm for the elimination of the reference concentration as an independently variable parameter in the analysis of sedimentation equilibrium. In: Schuster T, Laue T, editors. *Modern Analytical Ultracentrifugation*. Cambridge, MA: Birkhauser Boston, Inc. pp 81–93.
44. Provencher SW, Glöckner J (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* 20: 33–37.
45. Johnson WC Jr (1999) Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins: Struct Funct Genet* 35: 307–312.
46. Whitmore L, Wallace BA (2004) DICHROWEB: an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research* 32 (Web server issue): W668–W673.
47. Whitmore L, Wallace BA (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* 89: 392–400.
48. Backmann J, Schäfer G, Wyns L, Bönišch H (1998) Thermodynamics and kinetics of unfolding of the thermostable trimeric adenylate kinase from the archaeon *Sulfolobus acidocaldarius*. *J Mol Biol* 284: 817–833.
49. Greene RF Jr, Pace CN (1974) Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lactoglobulin. *J Biol Chem* 249: 5388–5393.
50. Woody RW (1995) Circular Dichroism. *Methods Enzymol* 246: 34–71.
51. Wang Z, Wu H, Chen J, Zhang J, Yao Y et al. (2008) A novel self-cleaving phasin tag for purification of recombinant proteins based on hydrophobic polyhydroxyalkanoate nanoparticles. *Lab Chip* 8: 1957–1962.
52. Sreerama N, Woody RW (1994) Poly(Pro)-II helices in globular proteins: identification and circular dichroic analysis. *Biochemistry* 33: 10022–10025.
53. Buck M (1998) Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins. *Q Rev Biophys* 31:297–355.
54. Lumry R, Eyring H (1954) Conformation changes of proteins. *J Phys Chem*. 58: 110–120.
55. Sánchez-Ruiz JM, López-Lacomba JL, Cortijo M, Mateo PL (1988) Differential scanning calorimetry of the irreversible thermal denaturation of thermolysin. *Biochemistry* 27:1648–1652.
56. Freire E, van Osdol WW, Mayorga OL, Sánchez-Ruiz JM (1990) Calorimetrically determined dynamics of complex unfolding transitions in proteins. *Annu Rev Biophys Biophys Chem* 19:159–188.
57. Swint L, Robertson AD (1993) Thermodynamics of unfolding for turkey ovomucoid third domain: Thermal and chemical denaturation. *Protein Sci*. 2: 2037–2049.
58. Myers JK, Pace CN, Scholtz JM (1995) Denaturant *m* values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci* 4: 2138–2148.
59. Phelan P, Gorfe AA, Jelezarov I, Marti DN, Warwicker J et al. (2002) Salt bridges destabilize a leucine zipper designed for maximized ion pairing between helices. *Biochemistry* 41: 2998–3008.
60. Dürr E., Jelezarov I, Bosshard HR (1999) Extremely fast folding of a very stable leucine zipper with a strengthened hydrophobic core and lacking electrostatic interactions between helices. *Biochemistry* 38: 870–880.
61. Privalov PL, Tiktopulo EI, Venyaminov SYu, Griko YuV, Makhatazde GI et al. (1989) Heat capacity and conformation of proteins in the denatured state. *J Mol Biol* 205: 737–750.
62. Uversky VN (2011) Intrinsically disordered proteins from A to Z. *Int J Biochem Cell Biol* 43:1090–1103.
63. Dunker AK, Brown CJ, Obradovic Z (2002) Identification and functions of usefully disordered proteins. *Adv Protein Chem* 62: 25–49.
64. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z (2002) Intrinsic disorder and protein function. *Biochemistry* 41: 6573–6582.
65. Peng Z, Mizianty MJ, Xue B, Kurgana L, Uversky VN (2012) More than just tails: intrinsic disorder in histone proteins. *Mol Biosyst* 8: 1886–1901.
66. Feng S, Chen JK, Yu H, Simon JA, Schreiber SL (1994) Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. *Science* 266: 1241–1247.
67. Lewitzky M, Harkiolaki M, Domart M-C, Jones EY, Feller SM (2004). Mona/Gads SH3C binding to hematopoietic progenitor kinase 1 (HPK1) combines an atypical SH3 binding motif, R/KXXK, with a classical PXXP motif embedded in a polyproline type II (PPII) helix. *J Biol Chem* 279: 28724–28732.
68. Venturi V, Otten M, Korse V, Brouwer B, Leong J et al. (1995) Alginate regulatory and biosynthetic gene homologs in *Pseudomonas putida* WCS358: correlation with the siderophore regulatory gene *phxA*. *Gene* 155: 83–88.
69. Fairman R, Chao H-G, Lavoie TB, Villafranca JJ, Matsueda GR et al. (1996) Design of heterotetrameric coiled coils: evidence for increased stabilization by Glu<sup>-</sup>Lys<sup>+</sup> ion pair interactions. *Biochemistry* 35: 2824–2829.