Photosynthetic CO₂ assimilation: the last gap in the structural proteome is closed

Libero Gurrieri[1], Alessandra Del Giudice[2], Nicola Demitri[3], Giuseppe Falini[4], Nicolae Viorel Pavel[2], Mirko Zaffagnini[1], Maurizio Polentarutti[3], Pierre Crozet[5], Julien Henri[5], Paolo Trost[1], Stéphane D. Lemaire[5], Francesca Sparla[1]*, Simona Fermani[4]*

[1]Department of Pharmacy and Biotechnology – FaBiT, University of Bologna, Bologna, Italy

[2]Department of Chemistry, University of Rome 'Sapienza', Rome, Italy,

[3] Elettra - Sincrotone Trieste, S.S. 14 km 163.5 in Area Science Park, 34149, Basovizza - Trieste, Italy

[4]Department of Chemistry 'G. Ciamician', University of Bologna, Bologna, Italy.

[5]Institut de Biologie Physico-Chimique, UMR8226, CNRS, Sorbonne Université, 13 rue Pierre et Marie Curie, 75005 Paris, France.

*Corresponding authors: francesca.sparla@unibo.it; simona.fermani@unibo.it

Photosynthetic CO₂ fixation supports life on Earth and is a fundamental source of food, fuels and chemicals for human society. In the vast majority of photosynthetic organisms, carbon fixation is operated by the Calvin-Benson (CB) cycle, a pathway that has been extensively studied at physiological, biochemical and structural level. It consists of 13 distinct reactions catalyzed by 11 enzymes[1]. In land plants and algae, the CB cycle takes place in the chloroplast, a specialized organelle operating the photosynthetic process. Despite decades of efforts, one last enzyme, the phosphoribulokinase (PRK), remains uncharacterized at atomic-scale while the structure of the other ten enzymes have been solved or confidentially modeled by close homology from the structure of isoforms found in other pathways. PRK together with ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) enzyme, is exclusive of the CB

1

cycle. In opposition to RuBisCO, whose structure has been solved from diverse sources[2], the structure of PRK remains unknown despite its characterization is crucial to better understand and control photosynthesis.

Here, we report for the first time the crystal structures of redox-sensitive PRK from two model species: the green alga *Chlamydomonas reinhardtii* (CrPRK) and the land plant *Arabidopsis thaliana* (AtPRK). The enzyme is an elongated homodimer characterized by a large central β-sheet of eighteen strands, extending between the two catalytic sites positioned at the dimer edges. The structure was also studied in solution through the combination of size exclusion chromatography and small-angle X-ray scattering (SEC-SAXS) and no appreciable differences were found.

This study completes the description at atomic level of the 11 CB cycle enzymes providing a rational basis for catalytic improvement of carbon fixation in chloroplasts.

The challenge of support 9 billion people requires a multidisciplinary approach. Looking at food production only, several studies suggest that a 70-100% increase will be necessary in the first half of the XXI century[3]. Different approaches have been made to improve crop yield trying to fill the gap between agricultural output and agricultural demand. Among them, the increase of photosynthetic carbon assimilation through the improvement of the CB cycle is regarded as a promising strategy and preliminary positive results have been obtained[4]. By consuming ATP, PRK catalyses the regeneration of the five-carbon sugar ribulose-1,5-bisphosphate (RuBP), exclusively used by RuBisCO for CO_2 fixation (Fig. S1a).

To gain insight into the molecular mechanism of RuBP regeneration, the crystal structures of CrPRK and AtPRK have been determined at 2.6 Å and 2.5 Å, respectively.

The enzyme is a dimer of two identical monomers related by a 2-fold non-crystallographic axis (Fig. 1a, b). Each monomer contains a central nine-stranded mixed β-sheet (β1−β9) surrounded by eight α-helices (α1, α3-α9), four additional β-strands (β1'-β4') and one helix α2 (Fig. 1c, d and Fig.
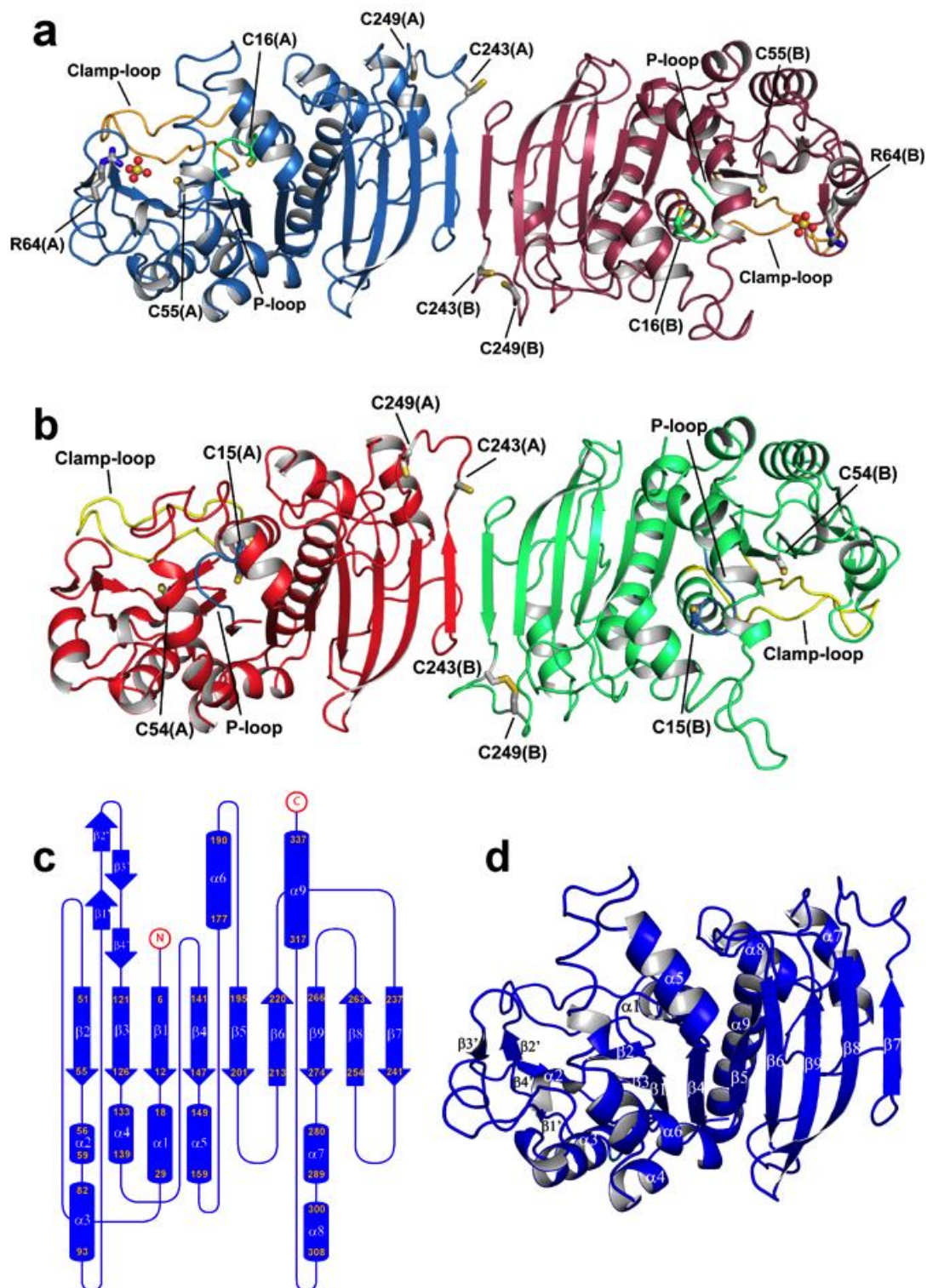
2

S1b, c).

Figure 1 | Crystal structure of photosynthetic PRK. Overall crystal structure in cartoon representation of reduced phosphoribulokinase (PRK) from **a**, *Chlamydomonas rehinardtii* and **b**, *Arabidopsis thaliana*. The two monomers are differently colored. The P-loop is highlighted in green

3

and light-blue and the clamp loop in orange and yellow. Each monomer contains two pairs of cysteines, one at the active site (Cys16 and Cys55 for CrPRK and Cys15 and Cys54 for AtPRK) and one at the dimer interface close to the C-terminal end of the protein chain (Cys243 and Cys249). Cysteine residues are indicated and represented as sticks. In AtPRK a pair of the C-terminal cysteines forms a disulfide bond. CrPRK binds two sulfate ions (one for each monomer) represented as spheres, coming from the crystallization solution. Arg64 represented as stick, is one of the residues stabilizing the anions. **c**, Topology diagram of CrPRK; the monomer is composed by a mixed β-sheet of nine strands, by nine α-helices and four additional small β-strands indicated by β'. **d**, Cartoon representation of the monomer structure of CrPRK; the central β-sheet is sandwiched between helices α3, α4 and α6 and helices α1, α7, α8 and α9. Strand β7 is involved in dimer interface, while the four additional β-strands (β1' to β4') form the edge of the dimer.

The active sites (one per monomer) are located in an elongated cavity at the edges of the dimer (Fig. 2a, b). In the CrPRK structure, the active site cavity is marked by a sulfate ion deriving from the crystallization medium that highlights the position occupied by the phosphate group of the substrate ribulose 5-phosphate (Ru5P) (Fig. 2a, c). The anion is stabilized by the side chains of Arg64, Arg67, Tyr103 and the main chain carbonyl group of His105 (Fig. S2).
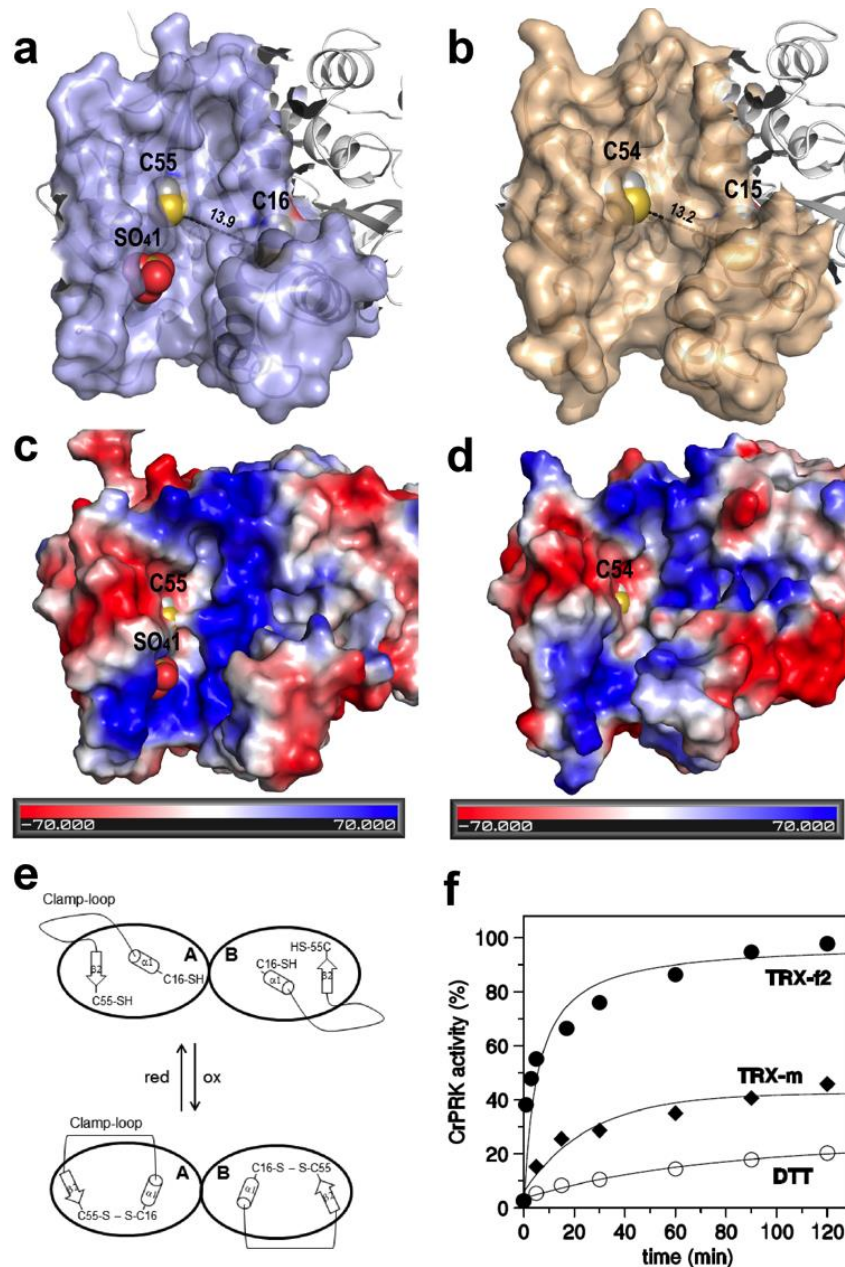
Figure 2 | Active site and TRX-dependent regulation of photosynthetic PRK. Catalytic cavity of **a**, CrPRK and **b**, AtPRK. The distance between the regulatory cysteines is higher than 13 Å, for both proteins. Catalytic cavity electrostatic surface potential of **c**, CrPRK and **d**, AtPRK. The positive potential observed at the bottom of the cavity is more intense and continuous in CrPRK than in AtPRK. A negative potential region observed in both proteins on the left side of the cavity, could be involved in the correct positioning of TRX close to regulatory cysteines. **e**, Schematic representation of the proposed clamp loop role and conformational changes occurring in chloroplast

PRK, upon TRX interaction. **f**, Activation kinetics of CrPRK by plastidial *C. reinhardtii* TRX-f2 and m, and DTT.

Except for Arg67, all others residues are strictly conserved in PRKs (Fig. S3a). The catalytic role of the invariant Arg64 was also confirmed in *C. reinhardtii*[5]. In addition, it was found that Arg64 plays a major role in the interaction of PRK with a second CB cycle enzyme, glyceraldehyde-3-phosphate dehydrogenase (GAPDH)[5]. Consistently, Arg64 faces out at the far end of the dimer (Fig. 1a and Table S1).

Despite the role of PRK in photosynthesis, putative PRK sequences are also found in heterotrophic prokaryotes[6]. Phylogenetic analyses performed on 69 PRK sequences including 31 from non-photosynthetic prokaryotes, clearly shows a first evolutionary separation between Bacteria and Archaea. Cyanobacterial and eukaryotic PRKs emerge as a third clade from the Archaea (Fig. S3b). A single structure of PRK from the anoxygenic photosynthetic bacterium *Rhodobacter sphaeroides* (RsPRK; PDB ID 1A7J)[6] and one from the non-photosynthetic Archaea *Methanospirillum hungatei* (MhPRK; PDB ID 5B3F)[7] have been reported. Both RsPRK and MhPRK sequences are shorter compared to plant PRKs and share low sequence identity making impossible to infer reliable structural models for *C. reinhardtii* and *A. thaliana* enzymes (Fig. S4). All PRKs from bacteria (type I PRKs), including phototrophic ones, are octamers of about 32 kDa subunits allosterically activated by NADH and inhibited by AMP[8,9]. Archaea, cyanobacteria and cyanobacterial-derived (*i.e.* plant-type) PRKs are dimers of about 40 kDa (type II PRKs)[10,11] with no allosteric regulation. While little is known about the regulation of archeal PRK, in oxygenic phototrophs from ancient cyanobacteria to modern flowering plants, PRK forms at low NADP(H)/NAD(H) ratios, a complex with glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and the regulatory protein CP12[12]. Moreover, cyanobacterial PRK is inhibited by AMP via the exclusive fusion protein CBS-CP12[13], while plant type PRK is directly regulated by thioredoxin (TRX)-dependent interconversion of a specific dithiol-disulfide bridge[10,11].

6

The pair of regulatory cysteines corresponding to Cys15/16 and Cys54/55 in AtPRK and CrPRK, are located within the P-loop (Walker A; responsible for ATP binding[14]) and at the C-terminal end of strand β2 (Fig. 1a-d, Fig. S1b, c, S3a and S4), respectively. Both residues stick out from the catalytic cavity showing a considerable accessibility to facilitate TRX interaction (Table S1). In addition, the similar midpoint redox potential (-312 ± 3 mV for CrPRK and -330 mV for AtPRK[15]) and the slightly acidic pK_a values (6.79 ± 0.01 for CrPRK and 6.95 ± 0.13 for AtPRK) denote similar redox properties for both PRKs (Fig. S5).

Moreover, pH and temperature affect the activity of AtPRK and CrPRK denoting a wider range of activity for algal enzyme (Fig. S6).

Interestingly, the pair of cysteines involved in TRX regulation appears already in TRX-insensitive PRK of cyanobacteria (Fig. S3a). However, cyanobacterial PRKs are deprived of an unstructured amino acid stretch[16] (here named clamp loop) separating the regulatory cysteines in plant-type PRKs whose thiol groups lie at more than 13 Å (Fig. 2a, b). The clamp loop connects α1 and β2 (Fig. 1, Fig. S1b, c and S4) and possibly allows the conformational change required for disulfide formation and consequent inhibition of enzyme activity (Fig. 2e).

The electrostatic surface potential of the catalytic cavity reveals an elongated positive region at the bottom and a negative portion on the side (Fig. 2c, d), the former suitable for binding the phosphate groups of substrates, *i.e.* ATP and Ru5P, and the latter relevant for recognition and positioning of TRX as proposed for fructose-1,6-bisphosphatase[17,18]. Compared to TRX-m, TRX-f2 is more efficient in the reductive activation of both AtPRK[19] and CrPRK (Fig. 2f). Accordingly, the structural model of AtTRX-f1 and the crystal structure of CrTRX-f2 (personal communication) show that the catalytic cysteines are surrounded by a large positive region, less extended or alternated with a negative one in TRX-m (Fig. S7), suggesting that the early stages of approaching and pairing between TRX and PRK are mainly governed by electrostatic interactions between exposed portions of the two proteins.

In addition, it has been demonstrated that the oxidized-state of the target protein is an essential

feature for TRX-target interaction, since disulfide formation in TRXs does not imply any significant conformational variations[20]. Indeed, the disulfide bond between the two distant regulatory cysteines introduces a significant conformational restriction that decreases the overall PRK entropy. Therefore, a favourable entropic contribution is proposed to be the main driving force for the reduction of PRK by TRXs.

By comparing the dimer interfaces, it emerges a large contact area in octameric RsPRK (1667 Å^2)[6] and dimeric MhPRK (1695 Å^2)[7] (Fig. S8) but a small one in eukaryotic phototrophs PRK. The dimer interface in CrPRK (545.6 Å^2) and AtPRKs (560.3Å^2) is exclusively formed by strand β7 (Fig. 1 and Fig. S1b, c) suggesting a less rigid overall structure. In cyanobacteria and cyanobacterial-derived, PRK is found in organisms that contain CP12 and therefore the greater flexibility of PRK could be more apt to form the regulatory GAPDH/CP12/PRK complexes[21].

By considering dimeric PRKs, AtPRK and CrPRK are characterized by a higher amount of exposed random coiled regions (Fig. 1a, b; Fig. S8b and Table S2). Some of that show a poor or absent electron density, thus being highly flexible and disordered. Two regions are common to algal and plant enzyme, while a third one is exclusive to AtPRK (Fig. S9).

The long loop lasting from helices α5 and α6 (Fig. S9) contains strictly conserved residues as the catalytic Asp160 or Arg159 and Arg164 involved in Ru5P binding[22,23]. A second flexible region, corresponding to the loop between strands β7 and β8, contains Cys243 and Cys249 (Fig. 1a, b and Fig. S9). In both PRK structures this portion is so flexible that the two thiol groups are far in one monomer and closer in the other, but differently from AtPRK they were not oxidized in CrPRK (Fig. 1a, b and Fig. S10), in agreement with the observation that in CrPRK the disulfide bond between Cys243 and Cys249 occurs uniquely through a dithiol/disulfide exchange reaction with the small unstructured protein CP12[24]. This interaction is an essential step of the assembly between PRK and GAPDH leading to the regulatory complex GAPDH/CP12/PRK[25-26].

The AtPRK crystal structure confirmed the elongated shape of the oxidized protein previously obtained by SAXS analysis[21], but it appears less bent and screwed. Structural information of

CrPRK in solution obtained by SEC-SAXS analysis provides a single recognizable species having a R_g value of 35 ± 1 Å and an estimated MW around 70 kDa (Fig. S11 and Table S3). The SAXS scattering profile well superimpose to the theoretical one calculated from CrPRK crystal structure (Fig. 3a) allowing us to conclude that no relevant conformational change of the protein occurs in solution (Fig. 3b).
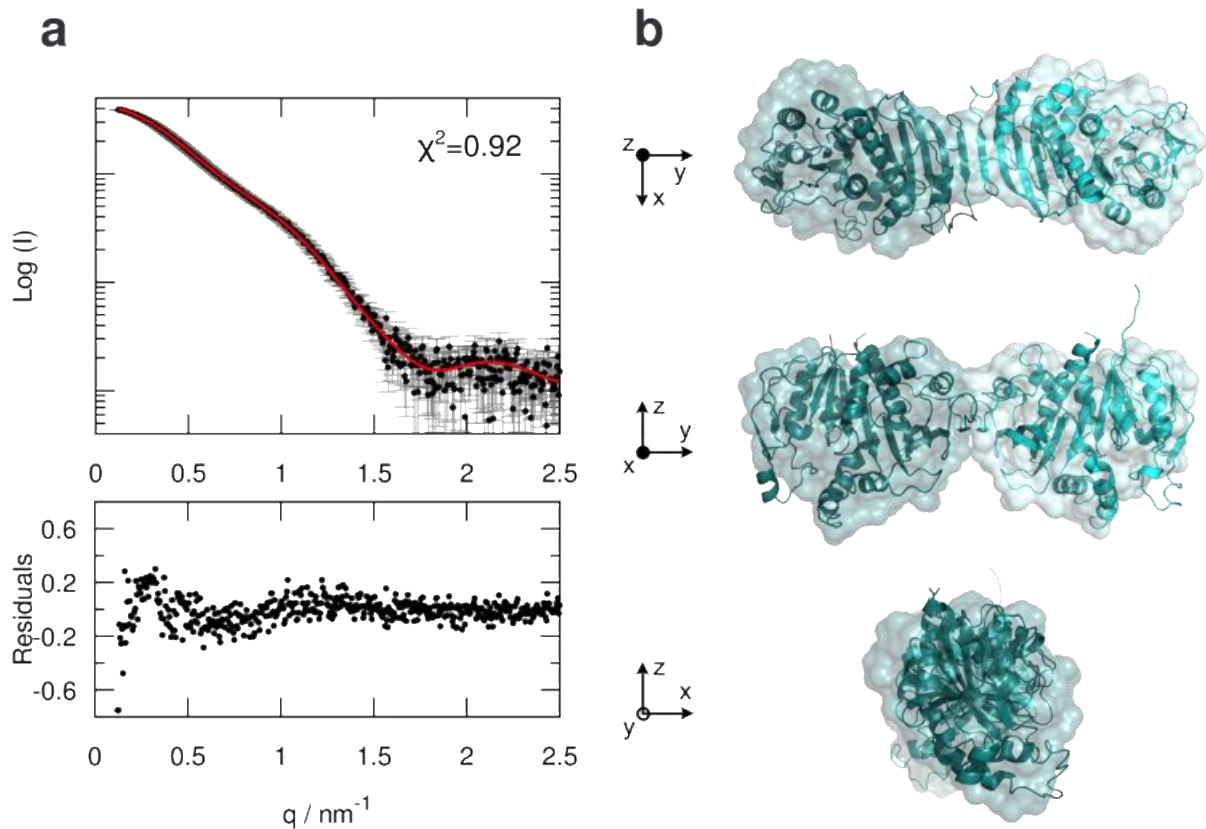

Figure 3 | SAXS analysis of CrPRK. **a**, Superimposition between the experimental SAXS scattering profile (black) and the theoretical SAXS curve (red) calculated from CrPRK crystal structure (upper graph). The lower graph shows the residuals obtained by the difference of experimental and calculated intensities. **b**, Comparison between the ab-initio model (surface representation) and the crystal structure (ribbon representation) of CrPRK, showing an excellent agreement between the two models.

In conclusion, our results fill the last gap of the structural proteome of the photosynthetic carbon

9

assimilation process, the biochemical pathway that more than the others contributes to the movement of inorganic carbon dioxide in the organic world, sustaining life on our planet.

In recent years, an increasing number of studies conducted through reverse genetic approaches have led to brilliant results. However, this approach has also highlighted many weaknesses, thus the idea that deleting or overexpressing a gene could lead to the desired results turned out to be simplistic.

A deep structural knowledge along with the advantages brought by the new technologies such as the genome editing, pave the way for a future improvement of the CB cycle.

References

1. Michelet, L. *et al.* Redox regulation of the Calvin-Benson cycle: Something old, something new. *Frontiers in Plant Science* **4,** 470 10.3389/fpls.2013.00470 (2013).

2. Tabita, F. R. *et al.* Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *J. Exp. Bot.* **59,** 1515–1524 (2008).

3. Godfray, H. C. *et al.* Food security: the challenge of feeding 9 billion people. *Science* **327,** 812–818 (2010).

4. Long, S. P. *et al.* Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. *Cell* **161,** 56–66 (2015).

5. Avilan, L. *et al.* Information transfer in multienzyme complexes 2. The role of Arg64 of *Chlamydomonas reinhardtii* phosphoribulokinase in the information transfer between glyceraldehyde-3-phosphate dehydrogenase and phosphoribulokinase. *Eur. J. Biochem.* **250,** 296–302 (1997).

6. Harrison, D. H. *et al.* The crystal structure of phosphoribulokinase from *Rhodobacter sphaeroides* reveals a fold similar to that of adenylate kinase. *Biochemistry* **37,** 5074–5085 (1998).

7. Kono, T. *et al.* A RuBisCO-mediated carbon metabolic pathway in methanogenic archaea. *Nat.*

Commun. **8,** 14007 (2017).

8. Kung, G. *et al.* Identification of the allosteric regulatory site in bacterial phosphoribulokinase. *Biochemistry* **38,** 15157–15165 (1999).

9. Martin, W. & Schnarrenberger, C. The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. *Curr. Genet.* **32,** 1–18 (1997).

10. Porter, M. A. & Hartman, F. C. Exploration of the function of a regulatory sulfhydryl of phosphoribulokinase from spinach. *Arch. Biochem. Biophys.* **281,** 330–334 (1990).

11. Brandes, H. K. *et al.* Efficient expression of the gene for spinach phosphoribulokinase in *Pichia pastoris* and utilization of the recombinant enzyme to explore the role of regulatory cysteinyl residues by site-directed mutagenesis. *J. Biol. Chem.* **271,** 6490-6496 (1996).

12. Tamoi, M. *et al.* The Calvin cycle in cyanobacteria is regulated by CP12 via the NAD(H)/NADP(H) ratio under light/dark conditions. *Plant J.* **42,** 504–513 (2005).

13. Hackenberg, C. *et al.* Structural and functional insights into the unique CBS-CP12 fusion protein family in cyanobacteria. *Proc. Natl. Acad. Sci. U S A* doi: 10.1073/pnas.1806668115 (2018).

14. Walker, J. E. *et al.* Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1,** 945–951 (1982).

15. Marri, L. *et al.* Reconstitution and properties of the recombinant glyceraldehyde-3-phosphate dehydrogenase/CP12/phosphoribulokinase supramolecular complex of Arabidopsis. *Plant Physiol.* **139,** 1433–1443 (2005).

16. Kobayashi, D. *et al.* Molecular characterization and redox regulation of phosphoribulokinase from the cyanobacterium Synechococcus sp. PCC 7942. *Plant Cell Physiol.* **44,** 269–276 (2003).

17. Mora-Garcia, S. *et al.* Role of electrostatic interactions on the affinity of thioredoxin for target proteins. Recognition of chloroplast fructose-1,6-bisphosphatase by mutant Escherichia coli thioredoxins. *J. Biol. Chem.* **273,** 16273–16280 (1997).

18. Gütle, D. D. *et al.* Chloroplast FBPase and SBPase are thioredoxin-linked enzymes with similar architecture but different evolutionary histories. *Proc. Natl. Acad. Sci. U S A* **113,** 6779–6784 (2016).

19. Marri, L. *et al.* Prompt and easy activation by specific thioredoxins of Calvin cycle enzymes of *Arabidopsis thaliana* associated in the GAPDH/CP12/PRK supramolecular complex. *Mol. Plant.* **2,** 259–269 (2009).

20. Palde, P. B. & Carroll, K. S. A universal entropy-driven mechanism for thioredoxin-target recognition. *Proc. Natl. Acad. Sci. U S A* **112,** 7960–7965 (2015).

21. Del Giudice, A. *et al.* Unravelling the shape and structural assembly of the photosynthetic GAPDH-CP12-PRK complex from *Arabidopsis thaliana* by small-angle X-ray scattering analysis. *Acta Cryst. D Biol. Cryst.* **71,** 2372–2385 (2015).

22. Charlier, H. A. *et al.* Evidence Supporting Catalytic Roles for Aspartate Residues in Phosphoribulokinase. *Biochemistry* **33,** 9343–9350 (1994).

23. Runquist, J. A. *et al. Rhodobacter sphaeroides* phosphoribulokinase: identification of lysine-165 as a catalytic residue and evaluation of the contributions of invariant basic amino acids to ribulose 5-phosphate binding. *Biochemistry* **38,** 13999–14005 (1999).

24. Thieulin-Pardo, G. *et al.* Phosphoribulokinase from *Chlamydomonas reinhardtii*: a Benson-Calvin cycle enzyme enslaved to its cysteine residues. *Mol. Biosyst.* **11,** 1134–1145 (2015).

25. Graciet, E. *et al.* The small protein CP12: a protein linker for supramolecular complex assembly *Biochemistry* **42,** 8163–8170 (2003).

26. Marri, L. *et al.* Spontaneous assembly of photosynthetic supramolecular complexes as mediated by the intrinsically unstructured protein CP12. *J. Biol. Chem.* **283,** 1831–1838 (2008).

METHODS

Protein expression and purification. Recombinant AtPRK and CrPRK were expressed in *E. coli* BL21(DE3) strain (New England Biolabs) harboring the pET-28 expression vector (Novagen) containing the coding sequence of the mature form of both enzymes (Fig. S4). AtPRK was expressed and purified as previously described[26]. The sequence coding for CrPRK was cloned in frame with a 6xHis-Tag and a thrombin cleavage site at the 5' end of the nucleotide sequence in order to purify the recombinant enzyme through a single chromatographic step performed loading the supernatant of the cellular lysate onto a Chelating Sepharose Fast Flow (GE Healthcare) column, following the manufacturer instruction. Immediately after elution, CrPRK was desalted in 30 mM Tris-HCl, 1 mM EDTA, pH 7.9 by PD-10 Desalting Columns (GE Healthcare). The 6xHis-Tag was removed by overnight digestion with thrombin protease. Samples purity was checked by 12.5% SDS-PAGE and purified proteins were quantified by absorbance at 280 nm (Nanodrop; Thermo Scientific) using the following extinction coefficients: εAtPRK 31985 M^{-1} cm^{-1}; εCrPRK 35995 M^{-1} cm^{-1} and molecular weights: AtPRK 38784.23 g mol^{-1} and CrPRK 40790.7 g mol^{-1}. Proteins were stored at -80 °C.

When required, oxidized and reduced AtPRK and CrPRK were prepared by incubation at 25°C for 2-3 h in the presence of 40 mM trans-4,5-dihydroxy-1,2-dithiane (oxidized DTT) or 1,4-dithiothreitol (reduced DTT). Following incubation, samples were desalted in 30 mM Tris-HCl, 1 mM EDTA, pH 7.9 through NAP-5 column (GE Healthcare) and brought to the desired concentration either by dilution or by concentration through Amicon-Ultra device (Millipore; cut-off 10 kDa).

Activity assay and determination of pH- and temperature-dependence. Phosphoribulokinase activity was measured spectrophotometrically (Cary 60 Uv-Vis; Agilent) as previously described[27]. The pH dependence of purified enzymes was determined in Britton-Robinson buffer for pH values ranging from 4.0 to 8.5. For pH ranging from 9 to 10, 100 mM Glycine buffer was used. Activities were measured on two independent protein batches and are expressed as percentage of the highest

13

measured activity. Temperature-dependence was evaluated on aliquots of AtPRK and CrPRK incubated for 20 min at temperatures ranging from 20° to 60°C. Following incubation the activity was assayed at 25°C. Activities were measured on two independent protein purifications and are expressed as percentage of the highest measured activity.

Thioredoxin specificity and redox titration. Thioredoxin specificity for activation of oxidized CrPRK (2 μM) was measured in the presence of 0.2 mM reduced DTT and *C. reinhardtii* recombinant TRX-f2 and TRX-m (5 μM each), as previously described[19]. At different incubation times, PRK activity was measured. To determine the midpoint redox potential of CrPRK, three independent redox titrations were performed. Briefly, a mixture containing pure recombinant CrPRK (12μM), commercial *E. coli* thioredoxin (1 mg ml^{-1}) and 20 mM DTT, at different reduced to oxidized ratios, were incubated for 3 h at 25°C before measuring PRK activity. The obtained curves were fitted by non-linear regression (CoStat, Co-Hort Software) to the Nernst equation.

Determination of the pK_a values of the catalytic cysteines. The determination of pK_a values for both CrPRK and AtPRK was performed as previously described[28]. Briefly, PRK activity was measured on reduced PRK samples (2 μM) following 20 min of incubation at 25°C in presence or in absence of 0.2 mM iodoacetamide (IAM). Incubations were performed in different buffers within a 4.5-10 pH range. Ten-20 μl aliquots were used to measure PRK activity in 1 ml of assay mixture. The residual activity was expressed as percentage of inhibition between IAM-treated and untreated samples, and expressed as a function of pH. The obtained curves were fitted by non-linear regression (CoStat, Co-Hort Software) to modified Henderson-Hasselbalch equation.

Crystallization and data collection. Reduced CrPRK was concentrated at 10 mg ml^{-1} in 30 mM Tris-HCl, 1 mM EDTA, pH 7.9. Reduced AtPRK was concentrated at 10 mg ml^{-1} in 25 mM potassium phosphate buffer solution, pH 7.5. They were crystallized by the vapor diffusion method at 293 K with sitting drop for CrPRK and hanging drop for AtPRK. A protein solution aliquot of 2 μl was mixed to an equal volume of reservoir, and the prepared drop was equilibrated against 900 (CrPRK) and 750 (AtPRK) μl of reservoir.

14

CrPRK crystals were obtained in different conditions of the Extension Kit from Hampton Research (solutions 22: 12 % w/v PEG 20K, 0.1 M MES, pH 6.5; 26: 30 % w/v PEG MME 5K, 0.1 M MES, pH 6.5, 0.2 M ammonium sulfate; 30: 10% w/v PEG 6K, 5% v/v MPD, 0.1 M HEPES, pH 7.5) and Structure screen 1 from Molecular Dimension (MD1-01-CF, solutions 35: 30 % w/v PEG 4K, 0.1 M Tris-HCl, pH 8.5, 0.2 M lithium sulfate; and 50: 15 % w/v PEG 8K, 0.5 M lithium sulfate). The conditions were optimized and the best diffracting crystal grew in about 10 days, from 22 % w/v PEG MME 5K, 0.1 M MES, pH 6.5, 0.2 M ammonium sulfate. CrPRK crystal appeared as needle-like, in the majority of the cases forming a cluster. Aggregates were manually separated and the thicker individuals fished.

AtPRK crystals showed a bipiramidal morphology and grew in three to five weeks from a reservoir solution containing $1.4 - 1.7$ M sodium malonate, pH 5.0. The best diffracting crystal of AtPRK was obtained in 1.5 M sodium malonate, pH 5.0.

Crystals were mounted from the crystallization drop into cryo-loops, briefly soaked in a cryo-protectant solution containing 30 % w/v PEG MME and 20 % v/v PEG 200 for CrPRK and 1.7 M sodium malonate, pH 5.0, and 30% v/v glycerol for AtPRK, then frozen in liquid nitrogen. Diffraction images were recorded at 100 K at the Elettra synchrotron radiation source (Trieste, beam line XRD1) for CrPRK and at the European Synchrotron Radiation Facility (Grenoble, beam line ID14-4) for AtPRK. Data collection parameters are reported in Table S4.

The data at a resolution of 2.6 Å for CrPRK and 2.5 Å for AtPRK were processed using XDS[29] and scaled with SCALA[30]. The correct space group was determined with POINTLESS[30] and confirmed in the structure solution stage. Data collection statistics are reported in Table S5.

Structure solution and refinement. CrPRK structure was solved by molecular replacement using the program PHASER[31] from PHENIX[32] starting from the coordinates of PRK from *Methanospirillum hungatei* (PDB code 5BF3)[7] deprived of sulfate ions and water molecules. The protein chain was traced by Autobuilt from PHENIX[32] and Buccaneer[33] from CCP4 package. The refinement was performed with REFMAC 5.8.0135[34] selecting 5% of reflections for R_{free}. The

manual rebuilding was performed with Coot[35]. The residual electron density map showed the position of a sulfate ion for each monomer coming from the crystallization solution, which was added to the model. Water molecules were automatically added and, after a visual inspection, confirmed in the model if the relative electron density value in the $(2F_o - F_c)$ maps exceeded 0.19 e$^-$ Å^{-3} (1.0 σ) and if they fell into an appropriate hydrogen bonding environment. The last refinement cycle was performed with PHENIX[32]. The structure of CrPRK without sulfate ions and waters, was used as initial model to solve the AtPRK structure by molecular replacement using MOLREP[36]. The refinement was performed as described for CrPRK. The refinement statistics are reported in Table S5. All structure figures were prepared using PyMOL (The PyMOL Molecular Graphics System, Schrödinger, LLC).

Small angle X-ray scattering data collection. Small angle X-ray scattering (SAXS) data were collected at the BioSAXS beamline BM29[37] at European Synchrotron Radiation Facility (ESRF, Grenoble) and the data collection parameters are reported in Table S4. A Size Exclusion Chromatography SEC-SAXS experiment was performed using a HPLC system (Shimadzu) directly connected to the measurement capillary. A volume of 100 μl of reduced CrPRK (6.1 mg ml^{-1}) was loaded onto a Superdex 200 10/300 GL column (GE Healthcare) pre-equilibrated in 50 mM Tris-HCl, 150 mM KCl, pH 7.5. The sample was eluted at a flow rate of 0.5 ml min^{-1} and SAXS frames obtained by 1 s exposure were collected continuously. The automatic pipeline for SEC-SAXS data analysis implemented at BM29 was used to assess the quality of the collected data[38].

Small angle X-ray scattering data analysis. Afterwards, a classification of the collected frames as buffer (0-14 ml) or protein frames (14-17.75 ml) was performed on the basis of the SAXS intensity trace (Fig. S12). Statistical test implemented in CorrMap[39] aided by visual inspection, was used to choose the superimposable buffer intensity profiles. The averaging of the buffer profiles, the subtraction of the averaged buffer intensity from the protein data and an automatic analysis of the subtracted protein profiles was performed with a Matlab script. The script used the tools of the ATSAS package[40] to automatically obtain from the subtracted intensity I(q): (i) the I(0) and the

16

gyration radius (R_g) via the Guinier approximation[41] $I(q) = I(0) \cdot \exp[-(qR_g)^2/3]$; (ii) the pair-distance [p(r)] function, from which the maximum particle dimension (D_{max}) was estimated, in addition to an independent calculation of I(0) and R_g. Estimates of the MW were also determined both from the Porod invariant[42] as 0.6 times the Porod volume (V_p) for roughly globular particles[40] and by the invariant volume-of-correlation length (V_c), through a power-law relationship between V_c, R_g and MW that has been parametrized[43]. The protein frames giving constant R_g values were scaled to the intensity of the elution maximum and averaged in order to obtain a single representative scattering profile with good signal to noise ratio, presented in the results and used for modelling.

Modelling from SAXS data. The experimental SAXS profile of the CrPRK dimer was fitted with the theoretical profile calculated from the atomic coordinates using CRYSOL3[44] with default parameters. In addition, low-resolution models of the CrPRK dimer based on the SAXS profile were built using the ab-initio program GASBOR[45]. The sequence and the homodimeric state of CrPRK were given as inputs in GASBOR and a 2-fold symmetry was imposed in the calculations. A series of 10 models was generated. The similarity of the structures obtained by repeated calculations was checked by DAMAVER[46] in which the superposition is performed by the SUPCOMB code[47].

All programs used for SAXS data analysis and reconstruction, belong to the ATSAS package 2.7[40]. The graphical representations of the obtained three-dimensional models were built by using PyMOL (The PyMOL Molecular Graphics System, Schrödinger, LLC).

Data availability. Atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) under accession codes 6H7G and 6H7H for CrPRK and AtPRK, respectively. The SEC-SAXS data of CrPRK have been deposited in SASBDB with accession code SASDDH9.

References

28. Roesler, K.R. & Ogren, W.L. *Chlamydomonas reinhardtii* phosphoribulokinase: sequence, purification and kinetics. *Plant Physiol.* **93,** 188–193 (1990).

29. Bedhomme, M. *et al.* Glutathionylation of cytosolic glyceraldehyde-3-phosphate dehydrogenase from the model plant *Arabidopsis thaliana* is reversed by both glutaredoxins and thioredoxins in vitro. *Biochem. J.* **445,** 337–347 (2012).

30. Kabsch, W. XDS. *Acta Cryst. D Biol. Cryst.* **66,** 125–132 (2010).

31. Evans, P. Scaling and assessment of data quality. *Acta Cryst. D Biol. Cryst.* **62,** 72–82 (2006).

32. McCoy, A.J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).

33. Adams, P.D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. D Biol. Cryst.* **66,** 213–221 (2010).

34. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Cryst. D Biol. Cryst.* **62,** 1002–1011 (2006).

35. Murshudov, G.N. *et al.* Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst. D Biol. Cryst.* **53,** 240–255 (1997).

36. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Cryst. D Biol. Cryst.* **60,** 2126–2132 (2004).

37. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Cryst. D Biol. Cryst.* **66,** 22–25 (2010).

38. Pernot, P. *et al.* Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J. Synchrotron Radiat.* **20,** 660–664 (2013).

39. Brennich, M.E. *et al.* Online data analysis at the ESRF bioSAXS beamline, BM29. *J. Appl. Crystallogr.* **49,** 203–212 (2016).

40. Franke, D. *et al.* Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. *Nat. Methods.* **12,** 419–422 (2015).

41. Petoukhov, M.V. *et al.* New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45,** 342–350 (2012).

42. Guinier, A. La Diffraction des rayons x aux très petits angles: application à l'étude des phénomènes ultramicroscopiques. *Ann. Phys.* **12,** 161–237 (1939).

43. Porod, G. General theory, in: Glatter O. & Kratky O. (Eds.), Small-Angle X-Ray Scatt., Academic Press, London, UK., 514 (1982).

44. Rambo, R.P. & Tainer, J.A. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature.* **496,** 477–481 (2013).

45. Franke, D. *et al.* ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50,** 1212–1225 (2017).

46. Svergun, D.I. *et al.* Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80,** 2946–2953 (2001).

47. Volkov, V.V. & Svergun, D.I. Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36,** 860–864 (2003).

48. Kozin, M.B. & Svergun, D.I. Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr*. **34,** 33–41 (2001).

49. Taylor, T.C. & Andersson, I. Structure of a product complex of spinach Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase. *Biochemistry* **36**, 4041- 4046 (1997).

50. Fermani, S. *et al*. The crystal structure of photosynthetic glyceraldehyde-3-phosphate dehydrogenase (isoform A4) from Arabidopsis thaliana in complex with NAD. *Acta Crystallog. F* **66**, 621 – 626 (2010).

51. Zaffagnini, M. *et al*. High resolution crystal structure and redox properties of chloroplastic triosephosphate isomerase from *Chlamydomonas reinhardtii*. *Mol. Plant* **7**, 101-120 (2014).

52. Chiadmi, M. *et al*., Redox signalling in the chloroplast: structure of oxidized pea fructose-1,6-bisphosphate phosphatase. *EMBO J*. **18**, 6809-6815 (1999)

53. Pasquini, M., *et al*. Structural basis for the magnesium-dependent activation of transketolase from *Chlamydomonas reinhardtii*. *Biochim. Biophys. Acta* **1861**, 2132-2145 (2017).

54. Kopp, J. *et al*. Structure and mechanism of the amphibolic enzyme D-ribulose-5-phosphate 3-epimerase from potato chloroplasts. *J. Mol. Biol*. **287**: 761-771 (1999).

55. Davies, G.J. *et al*. Structure of the ADP complex of the 3-phosphoglycerate kinase from *Bacillus stearothermophilus* at 1.65 A. *Acta Cryst. D Biol. Cryst.* **50**, 202-209 (1994).

56. Heron, P.W. & Sygusch, J. Isomer activation controls stereospecificity of class I fructose-1,6-bisphosphate aldolases. *J. Biol. Chem.* **292**, 19849-19860.

57. Dereeper, A. *et al*. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res*. **36**, 465-469 (2008).

58. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res*. **44**, 242-245 (2016).

59. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucl. Acids Res*. **42**, W320-W324 (2014).

60. Sievers, F. *et al*. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 539-545 (2011).

61. Capitani, G. *et al*. Crystal structures of two functionally different thioredoxins in spinach chloroplasts. *J. Mol. Biol.* **302**, 135-154 (2000).

62. Bienert, S. *et al*. The SWISS-MODEL Repository - new features and functionality. *Nucleic Acids Res*. **45**, D313-D319 (2017).

63. Chen, V.B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst. D Biol. Cryst.* **66**, 12-21 (2010).

64. Hallenbeck. P.L. & Kaplan, S. Cloning of the gene for phosphoribulokinase activity from *Rhodobacter sphaeroides* and its expression in Escherichia coli. *J. Bacteriol*. **169**, 3669-3678 (1987).

65. Meijer, W.G. *et al*. Identification and organization of carbon dioxide fixation genes in *Xanthobacter flavu*s H4-14. *Mol. Gen. Genet*. **225**, 320-330 (1991).

66. Kossmann, J. *et al.* Sequence analysis of the chromosomal and plasmid genes encoding phosphoribulokinase from *Alcaligenes eutrophus. Gene* **85**, 247-252 (1989).

67. Michalowski, C.B. *et al*. Phosphoribulokinase from ice plant: Transcription, transcripts and protein expression during environmental stress. *Photosynth Res*. **31**, 27-38 (1992).

68. Burki, F. *et al*. Evolution of Rhizaria: new insights from phylogenomic analysis of uncultivated protists. *BMC Evol. Biol.* **10**, 377-395 (2010).

69. Okamoto, K. *et al*. A landmark protein essential for mitophagy. *Autophagy* **5**, 1203-1205 (2009).

Acknowledgements

Author contributions

S.D.L and J.H. provided the cDNA of CrPRK; F.S. and L.G. purified and biochemical analysed the recombinant proteins; F.S., L.G and M.Z. analysed biochemical data; G.F., S.F., L.G. crystallized the proteins; S.F., M.P. and N.D. performed data collection, data processing, structure determination and refinement; F.S., S.F, A.D.G and N.V.P. performed SEC-SAXS experiments; A.D.G and N.V.P. performed SEC-SAXS analysis; P.C. performed phylogenetic analyses; F.S. and S.F. coordinated the experiments and supervised the whole research. F.S., S.F., S.D.L. and P.T. wrote the paper; all the authors discussed the data and the contributed to the editing manuscript.