

RESEARCH

Open Access



# Discovery of a readily heterologously expressed Rubisco from the deep sea with potential for CO<sub>2</sub> capture

Junli Zhang<sup>1,2†</sup>, Guoxia Liu<sup>1†</sup>, Alonso I. Carvajal<sup>3</sup>, Robert H. Wilson<sup>3,4\*</sup>, Zhen Cai<sup>1\*</sup> and Yin Li<sup>1\*</sup>

## Abstract

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), the key CO<sub>2</sub>-fixing enzyme in photosynthesis, is notorious for its low carboxylation. We report a highly active and assembly-competent Form II Rubisco from the endosymbiont of a deep-sea tubeworm *Riftia pachyptila* (RPE Rubisco), which shows a 50.5% higher carboxylation efficiency than that of a high functioning Rubisco from *Synechococcus* sp. PCC7002 (7002 Rubisco). It is a simpler hexamer with three pairs of large subunit homodimers around a central threefold symmetry axis. Compared with 7002 Rubisco, it showed a 3.6-fold higher carbon capture efficiency in vivo using a designed CO<sub>2</sub> capture model. The simple structure, high carboxylation efficiency, easy heterologous soluble expression/assembly make RPE Rubisco a ready-to-deploy enzyme for CO<sub>2</sub> capture that does not require complex co-expression of chaperones. The chemosynthetic CO<sub>2</sub> fixation machinery of chemolithoautotrophs, CO<sub>2</sub>-fixing endosymbionts, may be more efficient than previously realized with great potential for next-generation microbial CO<sub>2</sub> sequestration platforms.

**Keywords:** Rubisco, *Riftia pachyptila* endosymbiont, Form II, Hexamer, CO<sub>2</sub> capture in vivo

## Introduction

Photosynthesis uses light energy to convert inorganic CO<sub>2</sub> into organic carbohydrates and forms the basis of most life on earth. Most of photoautotrophs employ ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) for CO<sub>2</sub> fixation in the Calvin–Benson–Bassham (CBB) cycle. Extensive structural and biochemical studies have been performed on the Form I Rubisco enzymes from plants, algae and cyanobacteria. They share a similar hexadecameric structure composed of eight large subunits and eight small subunits, and are notorious for their low carboxylation activities, with turnover numbers toward

CO<sub>2</sub> in the range of 1–13 s<sup>-1</sup> (Galmes et al. 2014; Hanson 2016; Whitney et al. 2011). Much effort in rational design and directed evolution has been made to improve their catalytic efficiencies, with limited success (Cai et al. 2014; Durao et al. 2015). Earlier studies by our group and others have demonstrated the feasibility of the Rubisco-based pathway for carbon fixation in *E. coli* (Gleizer et al. 2019; Gong et al. 2015; Zhuang et al. 2013), using phosphoribulokinase (PRK) and Form I Rubisco from cyanobacteria *Synechococcus*. Although cyanobacterial Rubisco exhibit the highest activity among the Form I Rubiscos, the difficulty to achieve efficient heterologous soluble expression in *E. coli*, and other hosts, hampers the carbon fixation efficiency (Orr et al. 2020; Wilson et al. 2018; Zhou and Whitney 2019).

Nevertheless, nature has generated a variety of different Rubisco enzymes over more than 3.5 billion years of evolution (Fig. 1b) (Price et al. 2013). In addition to the extensively studied hexadecameric Form I Rubiscos, many Form II and Form III Rubiscos from anoxygenic

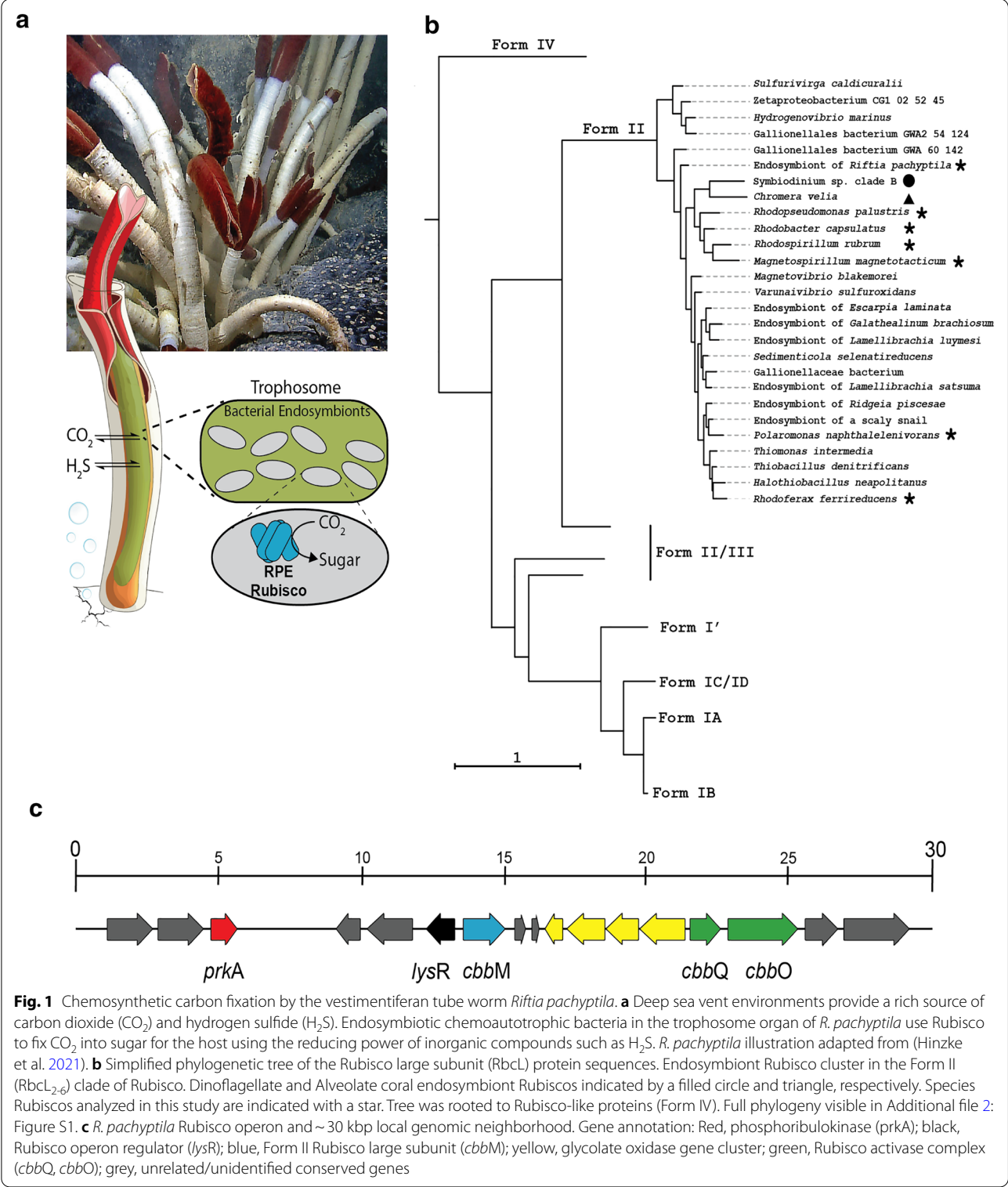
\*Correspondence: rhwilson@mit.edu; caiz@im.ac.cn; yli@im.ac.cn

†Junli Zhang and Guoxia Liu contributed equally to this work

<sup>1</sup> CAS Key Laboratory of Microbial Physiological and Metabolic Engineering, State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

<sup>3</sup> Department of Cellular Biochemistry, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany

Full list of author information is available at the end of the article



photoautotrophic, chemoautotrophic or heterotrophic bacteria and archaea have also been identified (Tabita et al. 2008). These forms lack the small subunit and

assemble as different hierarchies of large subunit dimers, which share low (~30%) primary sequence similarity to Form I Rubisco, with the exception of a recently

characterized ancestral Form I' clade which also lacks RbcS, but accordingly shows much higher similarity (~50%) to bona fide Form I members (Banda et al. 2020). To date, limited enzymological data were available regarding the catalytic parameters of these Rubisco enzymes. Most of the characterized Form II and Form III Rubiscos had far lower activities than that of the conventional Form I Rubisco from *Synechococcus* (Whitney et al. 2011) until a recent study reported a few highly active Form II Rubiscos (Davidi et al. 2020).

The aim of this study was to identify and characterize novel Rubisco with potential for carbon capture. To this end, seven Form II Rubiscos from a variety of microbes were selected for characterization of their carboxylation catalytic efficiency according to the diversities of classifications, growth types and habitats of their hosts. Heterologous expression and assembly level of these Rubiscos were evaluated. The high expression yields and activity of the endosymbiont Rubisco, particularly in comparison to cyanobacterial Rubisco, led us to investigate its structural properties. To demonstrate the potential application of RPE Rubisco, a bacterial carbon capture model was designed to characterize its CO<sub>2</sub> fixation capability in vivo as a versatile tool for microbial CO<sub>2</sub> capture platforms.

## Materials and method

### Plasmids construction

Seven Rubisco enzymes-encoding genes were selected and synthesized according to their NCBI accession numbers (Additional file 1: Additional Rubisco sequences information). All genes were inserted between the *Nde* I/*Xho* I sites of pET30a-7002-PRK (Additional file 2: Table S1) to replace the original 7002 Rubisco on the plasmid. A his<sub>6</sub>-tag was fused to the C terminus of the large subunit of the selected Rubisco enzymes for purification.

### Rubisco expression and purification

A single colony of the *E. coli* BL21(DE3) strain harboring Rubisco was inoculated into LB medium containing 50 ng µL<sup>-1</sup> of kanamycin and cultured overnight at 37 °C. An aliquot comprising 200 µL of the overnight culture was inoculated into 20 mL of fresh LB medium containing 50 ng µL<sup>-1</sup> of kanamycin. The culture was shaken at 37 °C until its OD<sub>600</sub> reached 0.8–1. Then, IPTG was added to a final concentration of 0.05 mM. The temperature was reduced to 22 °C and Rubisco expression was continued for 6 h. Cells harvested from 2 mL of the culture were resuspended in 1 mL of Rubisco buffer A (50 mM HEPES, pH 8.0, 10 mM MgCl<sub>2</sub>, 5 mM KCl, 0.5 mM EDTA, 5 mM DTT). The cells were disrupted by ultrasonication and clarified by centrifugation

to obtain the supernatant, which was cellular solution protein. An aliquot of 10 µL and 20 µL of cellular solution protein were subjected to SDS-PAGE (12%, w/v), and native-PAGE (6%, w/v). The his-tagged Rubisco in the cellular solution protein was purified using a nickel-chelating His-Bind column (Novagen) and the buffer was exchanged for Rubisco buffer B (50 mM Tris-HCl, pH 8.0, 10 mM MgCl<sub>2</sub>, 1 mM EDTA) using YM-30 Microcon device (Millipore).

### Rubisco activity assay

Carboxylation activity of Rubisco in cellular solution protein was assayed using NaH<sup>13</sup>CO<sub>3</sub> (Sigma) as substrate in an anaerobic chamber. Two hundreds microliter aliquots of crude cell extracts were activated for 30 min on ice with 20 µL of 200 mM NaH<sup>13</sup>CO<sub>3</sub>. The reaction was initiated by adding 13.5 µL of 29.6 mM D-ribulose-1,5-bisphosphate (RuBP, sigma) and conducted for 5 min at 25 °C. An aliquot comprising 60 µL of the reaction mixture was taken out and stopped by adding 20 µL of 1 M HCl. The solution was then neutralized with 20 µL of 1 M NaOH, and subjected to LC/MS/MS to determine the amount of <sup>13</sup>C-3PGA produced (Gong et al. 2015). One unit of carboxylation activity was defined as the amount of enzyme that produces 1 µmol of <sup>13</sup>C-3PGA per minute under the described condition.

Purified enzymes were subjected to the standard <sup>14</sup>CO<sub>2</sub> activity assay (Kubien et al. 2011; Wilson et al. 2018, 2019b) with modifications. Briefly, carboxylation activity of Rubisco was assayed anaerobically by preparing the reaction mixture in air-tight GC glass vials (Shimadzu) in an anaerobic chamber. The NaH<sup>14</sup>CO<sub>3</sub> (PerkinElmer) and deoxygenated RuBP solution needed for initiating the reaction were added using air-tight glass syringes. Each vial consisted of 260 µL of Rubisco buffer B containing 30 mM NaHCO<sub>3</sub>, 2 µL of NaH<sup>14</sup>CO<sub>3</sub>, and 20 µL of purified enzyme. The reaction mixture was activated for 30 min at 25 °C followed by addition of 20 µL of 5 mM RuBP to start the reaction. The background control reactions were initiated by with Milli-Q ddH<sub>2</sub>O instead of RuBP. Each reaction was conducted at 25 °C for 2 min and stopped by the addition of 100 µL of 50% (v/v) formic acid. Reactions were dried in heat block at 80 °C, dissolved in 0.25 mL of H<sub>2</sub>O, mixed with 1 mL of Ultima-Gold scintillant (PerkinElmer), and measured for radioactivity in a scintillation counter (PerkinElmer).

The <sup>14</sup>C-labeled 2-carboxyarabinitol-1,5-diphosphate (CABP) was synthesized by RuBP and <sup>14</sup>C-KCN (American Radiolabeled Chemicals, Inc.) as before (Andersson et al. 1983; Pierce et al. 1980). Active sites of Rubisco in purified enzymes were determined as reported (Kubien et al. 2011). The <sup>14</sup>CO<sub>2</sub> carboxylation activity was defined as the molar amount of fixed <sup>14</sup>CO<sub>2</sub> divided

by the molar amount of active sites determined by the  $^{14}\text{C}$ -labeled CABP binding per second. Rubisco soluble cellular concentration (%CSP) was obtained by dividing the Rubisco-soluble protein expressed by total cellular solution protein (Wilson et al. 2018, 2019b). Among them, Rubisco-soluble protein expressed was calculated by multiplying the nmol active sites according to the  $^{14}\text{C}$ -CABP-binding by Rubisco molecular weight, the total cellular soluble protein was assayed by  $A_{280}$  using NanoDrop ND-1000. Furthermore, the enzyme activity (nmol  $\text{CO}_2$  fixed  $\text{min}^{-1} \text{mg}^{-1}$ ) was specified as the molar amount of fixed  $^{14}\text{CO}_2$  divided by total cellular solution protein, which  $\text{CO}_2$  fixation was assayed by  $^{14}\text{C}$ -method as mentioned above.

### Measurement of kinetic parameters

Kinetic parameters for  $\text{CO}_2$  of the purified enzymes were measured anaerobically using  $\text{NaH}^{14}\text{CO}_3$  (Wilson et al. 2018). A reaction mixture consisted of 210  $\mu\text{L}$  of Rubisco buffer B containing 10  $\mu\text{g/mL}$  carbonic anhydrase and 30–1800  $\mu\text{M}$   $\text{CO}_2$  and 20  $\mu\text{L}$  of 10 mM RuBP was prepared in air-tight GC glass vials in an anaerobic chamber.  $\text{NaH}^{14}\text{CO}_3$  was added by air-tight glass syringes. The assays were initiated by the addition of 20  $\mu\text{L}$  of preactivated Rubisco by gas-tight syringes, incubated at 25  $^\circ\text{C}$  for 2 min, and stopped by 100  $\mu\text{L}$  of 50% (v/v) formic acid. Determination of radioactivity and active sites were the same as above. Kinetic parameters were calculated by non-linear fitting to the Michaelis–Menten equation by software OriginPro 8.5.

### Crystallization and structure determination

Purified RPE Rubisco was buffer-exchanged into a solution containing 10 mM Tris–HCl (pH 8.0) and 100 mM NaCl. The protein concentration was determined to be 10  $\text{mg mL}^{-1}$ . Seven 96-well crystal screening kits were set up for crystallization using the sitting drop vapor diffusion method by mixing equal volumes of protein and a buffer solution containing 1.26 M sodium phosphate monobasic monohydrate and 0.14 M dibasic potassium phosphate at 18  $^\circ\text{C}$ . The best diffracting apo crystals were grown using reservoir solution comprising 0.15 M magnesium formate and 12% (w/v) polyethylene glycol 3350. Crystals were rapidly soaked in reservoir solution supplemented with 20% glycerol as cryoprotectant, mounted on loops, and flash-cooled at 100 K in a nitrogen gas cryostream. Crystals diffraction data were collected from a single crystal at the BL18U beamline of the Shanghai Synchrotron Radiation Facility (SSRF, China), with a wavelength of 0.9793  $\text{\AA}$  at 100 K. The diffraction data were processed and scaled using HKL-3000 (Otwinowski and Minor 1997). The relevant statistics are summarized in Additional file 2: Table S2. The structure was solved

by the molecular replacement method using *Gallionella* Rubisco (PDB code 5C2C as the starting model. The initial model was built using PHENIX autobuild (Adams et al. 2002). Manual adjustment of the model was carried out using the program COOT (Emsley and Cowtan 2004) and the models were refined using PHENIX refinement (Adams et al. 2002) and REFMAC5 (Murshudov et al. 1997). The stereochemical quality of the structures was checked using PROCHECK (Laskowski et al. 1993). All residues were found to be located in the favored and allowed regions and none in the disallowed regions. Refinement resulted in a model with excellent refinement statistics and geometry (Additional file 2: Table S2). The coordinates of RPE Rubisco have been deposited in Protein Data Bank under the PDB code 6IUS.

### Cell growth and D-lactate production using Rubisco $\text{CO}_2$ fixation

For comparing cell colonies growth on the plate, 1 mL overnight cell culture in LB broth with 50  $\text{ng mL}^{-1}$  of kanamycin was harvested by centrifugation (10,000 g, 1 min), washed and resuspended in 0.9% NaCl twice, followed by gradient dilution, 10  $\mu\text{L}$  of a  $10^{-3}$ ,  $10^{-4}$ ,  $10^{-5}$ ,  $10^{-6}$  dilution was individually spotted on M9 minimum medium (5 g  $\text{L}^{-1}$  xylose, 6.13 g  $\text{L}^{-1}$  glycerol, 0.5 g  $\text{L}^{-1}$  casamino acid, 0.05 mM IPTG, 50  $\text{ng mL}^{-1}$  of kanamycin) and cultured in 10%  $\text{CO}_2$ , 90% air, 25  $^\circ\text{C}$ . For comparing cell growth in liquid medium, 1:100 (1%, v/v) overnight culture overnight cell culture in LB broth with 50  $\text{ng mL}^{-1}$  of kanamycin was inoculated into 50 ml fresh M9 minimum medium and cultured under the same condition as above.

After Rubisco expression in LB broth, 10 mL cell culture was harvested and washed by 0.9% NaCl, cell supernatants were collected at an  $\text{OD}_{600}$  of  $0.20 \pm 0.02$ , resuspended in 10 ml M9 minimal medium in addition with 100 mM HEPES and 100 mM  $\text{NaH}^{13}\text{CO}_3$ , pH 7.0. Following incubation for 3 days in anaerobic condition at 30  $^\circ\text{C}$  and regular sampling, cell growth was monitored using a 96-plate reader (Infinite 200pro, Tecan). The consumption of xylose and glycerol in the supernatant was determined using Agilent 1200 Infinity HPLC system with an Aminex HPX-87H column (300\*7.8 mm, Bio-Rad) and refractive index detector. Samples were run at 55  $^\circ\text{C}$  and eluted at 0.6 ml/min with 5 mM sulfuric acid. The D-lactate production including the  $^{13}\text{C}$ -labeled D-lactate concentration was carried out by QTRAP 6500 LC–MS/MS with Multiple Reaction Monitoring (MRM) mode. The temperature of electron spray ionization (ESI) was 550  $^\circ\text{C}$ , and the detection voltage was -4500 V. Samples were analyzed with a 10- $\mu\text{L}$  injection onto a HyperREZ XP column (7.7\*100 mm) heated to 40  $^\circ\text{C}$ . The deionized water eluted for 10 min at a flow rate of 0.4  $\text{mL min}^{-1}$ .



### Metabolic flux analysis

The constrained metabolic flux was analyzed according to the Rubisco-based CO<sub>2</sub> capture model. The consumption of xylose (Xyl<sub>total</sub>) was separated into pentose phosphate pathway (X<sub>1</sub>) and Rubisco bypass pathway (X<sub>2</sub>), which was calculated by Eq. 1. The mole of F6P was calculated 5/6 X<sub>1</sub> by carbon rearrangement through pentose phosphate pathway. The generated G3P (X<sub>3</sub>) was derived from F6P and glycerol metabolism (Gly<sub>total</sub>), as shown in Eq. 2:

$$\text{Xyl}_{\text{total}} = X_1 + X_2, \quad (1)$$

$$X_3 = 2 \times \frac{5}{6} X_1 + \text{Gly}_{\text{total}}. \quad (2)$$

The important intermediate metabolite 3PGA was produced by central carbon metabolism and Rubisco-based CO<sub>2</sub> capture bypass pathway, and the mole of 3PGA was the summation of the mole of G3P and the mole of uptake CO<sub>2</sub> (X<sub>2</sub>). Under this premise, that NADH production was greater than or equal to NADH consumption, it was assumed that <sup>13</sup>C-labeled ratio of 3PGA (R<sub>3PGA</sub>) was equal to the ratio of the detected labeled lactate, so the detected ratio of <sup>13</sup>C-labeled lactate (R<sub>lactate</sub>, after deducting the background interference) was calculated by X<sub>2</sub> and X<sub>3</sub> in Eq. 3.

As NADH production ≥ NADH consumption,

$$R_{3\text{PGA}} \% = \frac{X_2}{X_3 + 2X_2} = R_{\text{lactate}} \%. \quad (3)$$

The <sup>13</sup>C-labeled D-lactate was detected at the end of 72 h, the consumption rates and the production rates, and the uptake CO<sub>2</sub> flux were calculated in the last 12 h.

### Rubisco large-subunit protein phylogenetic analysis

Eighty six Rubiscos were manually selected to have representatives of the different groups of Rubisco, focusing specially on Type II Rubiscos and Rubiscos reported from endosymbionts. This set of Rubisco sequences was then aligned and trimmed using MAFFT (Katoh et al. 2002) in Guidance2 server (Sela et al. 2015). The columns with scores lower than 0.93 were removed and subsequently, a maximum-likelihood phylogenetic was built using PhyML 3.0 (Guindon and Gascuel 2003) with 1000 Bootstraps and a LG model for amino acid substitution.

### Statistical analysis

The statistical analysis of data and plots was performed using an unpaired 2-tailed Student's t test in GraphPad Prism software version 7.0. *P* values of <0.05 were

considered to indicate statistical significance. Data are presented as means ± standard errors of the mean.

## Results

### Discovery of a highly active Form II RPE Rubisco

Seven Form II Rubiscos from different microbes were selected for characterization (Additional file 2: Table S1). These 7 microbes belong to 3 classes, 4 orders, 4 families, and 7 genera, and are capable of photoautotrophic, chemolithoautotrophic, and heterotrophic growth. They live in diverse environments including extreme habitats such as deep-sea hydrothermal vents and cold lakes. The multiple sequence alignment was provided and the similarity of each two Rubiscos amino acid sequences was about 79%~91%. Their Rubisco genetic evolutionary relationship was analyzed by the simplified phylogenetic tree (Fig. 1b).

The large subunit gene of each selected Rubisco (RbcL) was placed under the control of a T7 promoter in a pET30a plasmid and expressed in *E. coli* BL21(DE3). A representative cyanobacterial Rubisco from *Synechococcus* sp. PCC 7002 (7002 Rubisco) was used as a reference for Form I Rubisco expression yield and activity. The large and small subunit genes of 7002 Rubisco, together with *rbcX* which encodes a Rubisco specific assembly chaperone that improves enzyme yield in *Escherichia coli* (Emlyn-Jones et al. 2006), were cloned into pET30a from our previous study (Cai et al. 2014). All the selected Form II Rubiscos were expressed and assembled in *E. coli* BL21(DE3) without addition of any foreign chaperones (Additional file 2: Figure S2a, b). For most of these Form II Rubiscos, the expression/assembly levels and the carboxylation activities assessed using cellular solution protein were higher than those of 7002 Rubisco expressed with the assistance of RbcX. Among them, the Rubisco from endosymbiotic bacteria which lived in the trophosome of the *Riftia pachyptila* (RPE Rubisco) showed the highest carboxylation activity in cellular solution protein (Additional file 2: Figure S2c). RPE Rubisco could be the key enzyme of ubiquitous chemosynthetic symbioses between invertebrate and bacteria to fix CO<sub>2</sub> using H<sub>2</sub>S as energy to support the rapid growth of the host (Fig. 1a) (Hinzke et al. 2021; Li et al. 2018). Interestingly, RPE Rubisco resides in a separate clade away from other endosymbionts (Fig. 1b). Its operon and ~30-kbp local genomic neighborhood contains coding genes of Rubisco (*cbbM*), phosphoribulokinase (*prkA*), and Rubisco activase complex (*cbbQ*, *cbbO*) (Fig. 1c).

### Kinetic parameters and solubility of RPE Rubisco

Both RPE and 7002 Rubiscos were purified by nickel affinity chromatography after fusion with a his<sub>6</sub>-tag at the C terminus of the large subunit. Standard <sup>14</sup>CO<sub>2</sub>

activity assay revealed a  $k_{\text{cat}}^{\text{C}}$  value for RPE Rubisco of  $16.4 \text{ s}^{-1}$  at  $25^\circ\text{C}$ , whereas  $12.3 \text{ s}^{-1}$  for 7002 Rubisco. The  $K_{\text{M}}^{\text{C}}$  value of RPE Rubisco was 11% lower compared to 7002 Rubisco. The resulting carboxylation efficiency ( $k_{\text{cat}}^{\text{C}}/K_{\text{M}}^{\text{C}}$ ) of RPE Rubisco was thus 50.5% higher than that of 7002 Rubisco (Table 1).

In addition to the higher carboxylation efficiency, the solubility of RPE Rubisco in *E. coli* was also noteworthy. Based on the active sites of Rubisco that were determined by  $^{14}\text{C}$ -labeled CABP binding (Kubien et al. 2011), the cellular concentration of RPE Rubisco was calculated to be 12% of the cellular solution protein (%CSP), which is about 12-fold higher than 7002 Rubisco (Fig. 2a). Such a high cellular concentration of RPE Rubisco coincides well with its high-level soluble expression and assembly. Consequently, the in vitro  $\text{CO}_2$  fixation rate of RPE Rubisco in bacterial crude extracts reached up to  $820 \pm 182 \text{ nmol CO}_2/\text{min}^{-1}/\text{mg CSP}$ , which was 19-fold higher than 7002 Rubisco (Fig. 2b), and eightfold higher in comparison to abundantly expressing higher plant (*Arabidopsis thaliana*) Rubisco in *E. coli* (Wilson et al. 2019b). In fact, most Form I Rubisco, with the exception of cyanobacterial isoforms like 7002 Rubisco, display carboxylation rate values ( $k_{\text{cat}}^{\text{C}}$ ) that fundamentally limit host  $\text{CO}_2$  fixation rates at normal recombinant expression yields (Fig. 2c). As shown in this study and others (Lin et al. 2014; Orr et al. 2020; Wilson et al. 2018), the poor heterologous expression of cyanobacterial Rubisco provides an incredibly large barrier to effective  $\text{CO}_2$  capture utility outside of their normal hosts and in the absence of the  $\text{CO}_2$  concentrating apparatus and chaperones that support their function. In these terms, RPE Rubisco stands in a class of its own with the potential to capture carbon at a significant rate under elevated atmospheric  $\text{CO}_2$  concentrations and straightforward heterologous expression (Fig. 2c).

### Structure determination and analysis

Analytical ultracentrifugation revealed that RPE Rubisco has a molecular weight of 322 kDa (Additional file 2: Figure S3), indicating that it is a hexamer. Structure resolution confirmed this and showed that RPE Rubisco (PDB code 6IUS) consisted of three pairs of homodimers of the large subunit arranged around a central threefold symmetry axis (Fig. 3a). Structural comparison of RPE

Rubisco and two previously reported hexameric RPA and GAL Rubisco (PDB codes 4LF1 and 5C2G, respectively) (Satagopan et al. 2014; Varaljay et al. 2016) showed a strong similarity, which can be ascribed to their high identity of amino acid sequences (73–78%). Superimposition of their monomeric large subunits revealed that the C $\alpha$  backbones were virtually identical (Fig. 3b). Residues within 4 Å of CABP were conserved among the three Rubiscos. Fifteen active-site amino acids predicted based on the protein sequence of Rubisco are labeled in Fig. 3c, d. The active-site geometries of RPE and GAL Rubiscos without ligands are highly similar, showing the “apo” forms (Fig. 3c). The “activated” forms of RPA and GAL Rubiscos with CABP binding were also highly similar (Fig. 3d). The main conformational changes between the “apo” and “activated” forms occurred in residues Lys 166, Met 330, and Lys 329 (Fig. 3c). These three active-site amino acids were potentially related with the major conformational changes accompanying the reaction of Rubisco with substrate RuBP. Lys 166 and Lys329 participate in catalysis (Cleland et al. 1998), thus cannot be substituted. Met 330 residue described above in RPE Rubisco is similar to that of Met 331 in RPA Rubisco loop 6, which closes an active site in conjunction with residues from a neighboring subunit of homodimer. The  $K_{\text{cat}}$  of mutant enzyme M331L was decreased about 95% (Satagopan et al. 2014), highlighting the importance of Met 331 in RPA Rubisco. Likewise, Met 330 might also play an indispensable role in RPE Rubisco. Slight differences were observed when comparing their oligomeric structures. The interaction surfaces within one dimer and between two neighboring dimers of RPE Rubisco were  $3223 \text{ Å}^2$  and  $2426 \text{ Å}^2$ , respectively, while those for RPA Rubisco were  $4232 \text{ Å}^2$  and  $2131 \text{ Å}^2$ , and  $4209 \text{ Å}^2$  and  $2612 \text{ Å}^2$  for GAL Rubisco (Fig. 3e).

### Creation of a Rubisco-based $\text{CO}_2$ capture model

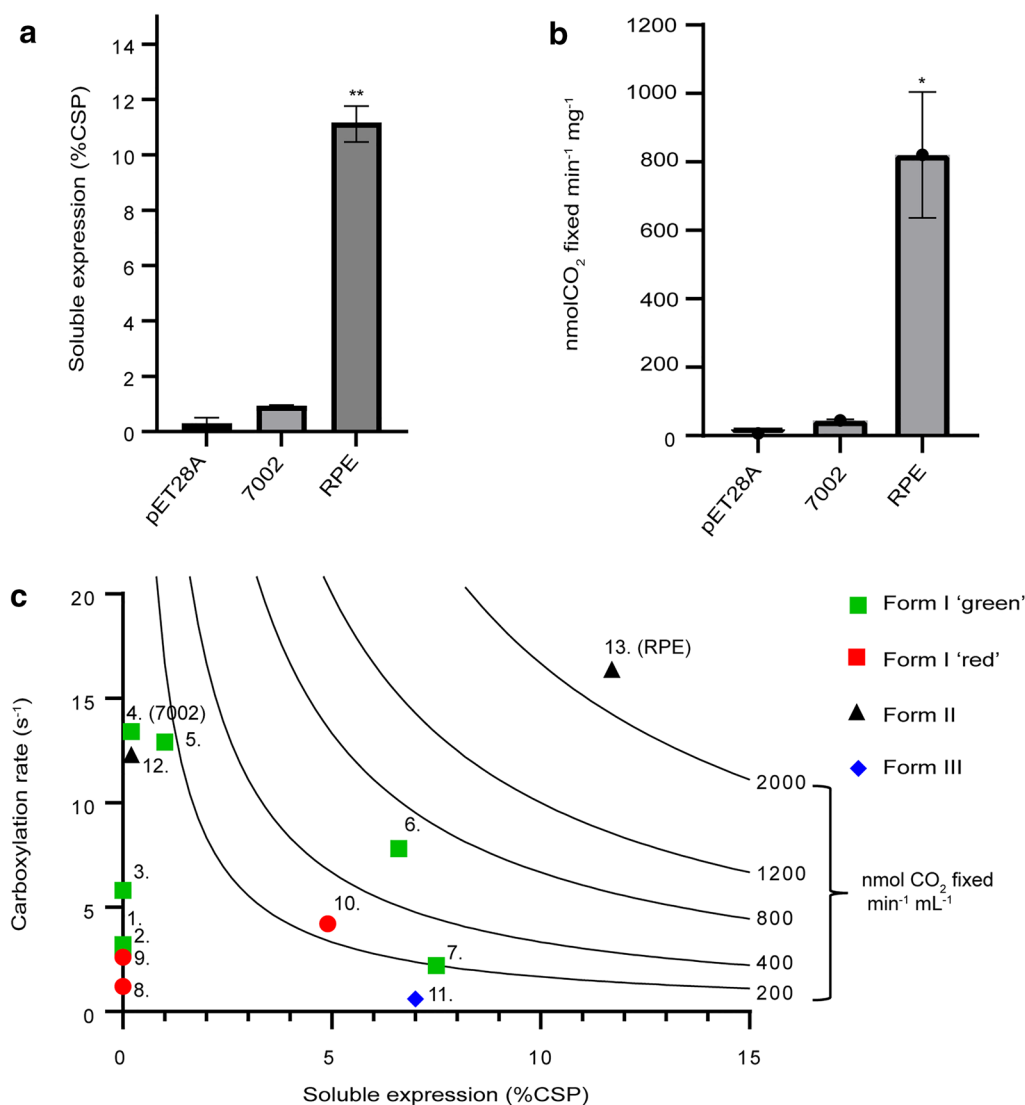
To investigate whether the higher carboxylation catalytic efficiency or heterologous expression of RPE Rubisco could improve heterotrophic  $\text{CO}_2$  fixation in *E. coli*, a  $\text{CO}_2$  capture model was designed. The model should be able to evaluate the activity of different Rubiscos and the efficiency of  $\text{CO}_2$  capture. To this end, the basic idea was to choose xylose as the starting substrate to capture  $\text{CO}_2$  through Rubisco. Moreover, the xylose and the captured  $\text{CO}_2$  are directed towards the production of D-lactate so that  $\text{CO}_2$  is deposited in the form of D-lactate (Fig. 4). Three principles were considered when developing this model.

First, since xylose can be metabolized by *E. coli* through the pentose phosphate pathway, selective pressure must be introduced to force xylose to be used to capture  $\text{CO}_2$ . PRK and Rubisco were thus cloned under an inducible

**Table 1** Kinetic parameters of 7002 and RPE Rubiscos

| Enzyme       | $k_{\text{cat}}^{\text{C}} (\text{s}^{-1})$ | $K_{\text{M}}^{\text{C}} (\mu\text{M})$ | $k_{\text{cat}}^{\text{C}}/K_{\text{M}}^{\text{C}} (\text{s}^{-1} \text{ mM}^{-1})$ |
|--------------|---|---|---|
| 7002 Rubisco | $12.3 \pm 0.7$                              | $194.6 \pm 35.7$                        | $63.2 \pm 8.0$  |
| RPE Rubisco  | $16.4 \pm 0.7$                              | $172.4 \pm 23.6$                        | $95.1 \pm 9.0^*$  |

Data represent mean  $\pm$  standard deviation of at least three independent measurements. *P*-values relative to 7002 Rubisco indicated significance at  $^*P < 0.05$ ,  $^{**}P < 0.01$

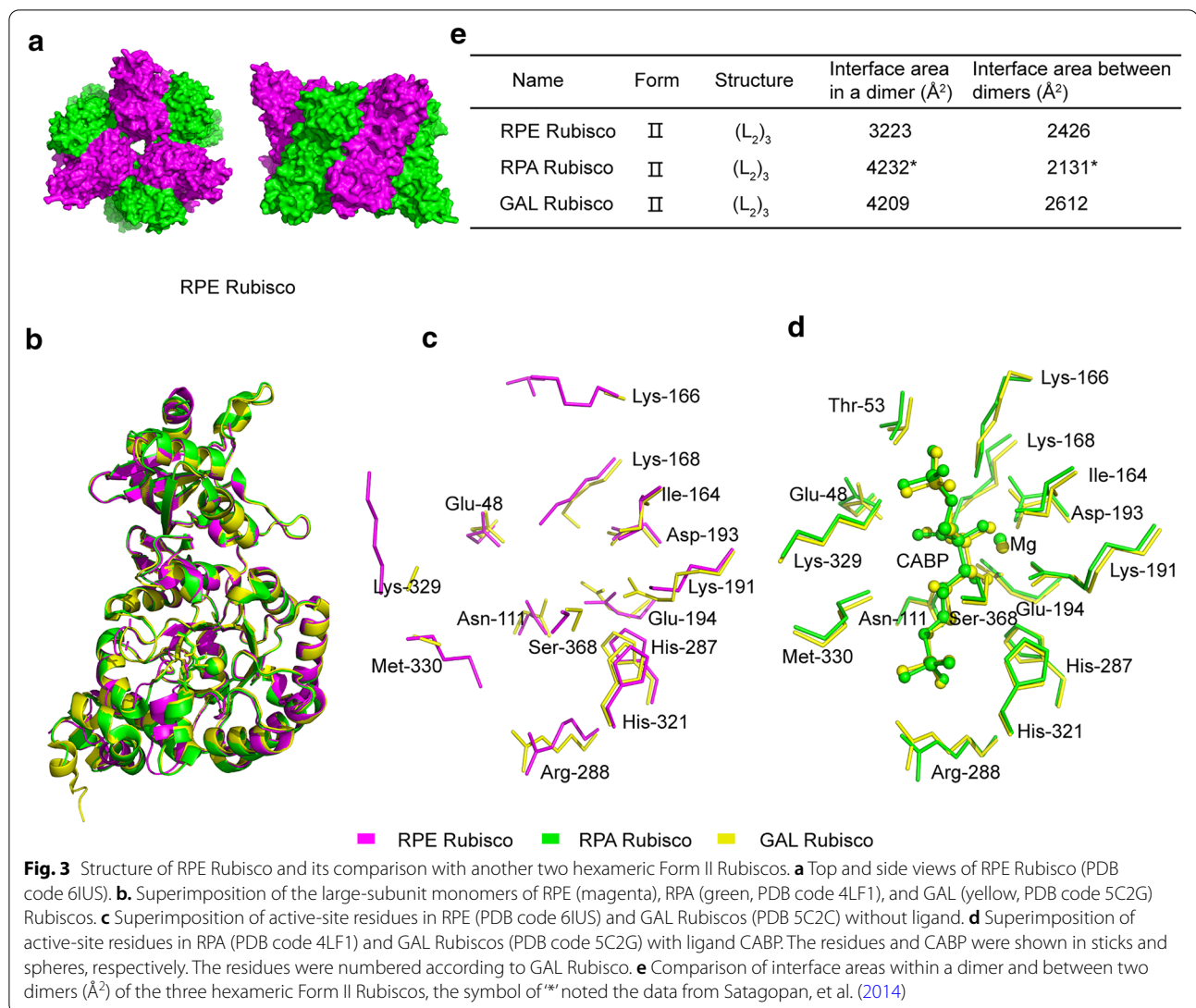


**Fig. 2** Comparative analysis of carboxylation activity and solubility between 7002 and RPE Rubisco by <sup>14</sup>C-labeled method. **a** Rubisco soluble cellular concentration as a percentage of total cellular solution protein (%CSP), **b** The enzyme activity measured as the molar amount of fixed <sup>14</sup>CO<sub>2</sub> divided by total cellular solution protein, **c** Rubisco maximal carboxylation rates ( $V_{max}$ ) and soluble expression yields in *E. coli* as reported %CSP and according to labeling in Additional file 2: Table S3. The %CSP values are represented as the observed or reported yield when no additional chaperones are co-expressed alongside the Rubisco genes. Contour lines represent the %CSP yield and Rubisco carboxylation rate ( $V_{max}$ ) required to achieve a maximum theoretical CO<sub>2</sub> fixation rate in vivo. This rate is represented in nmol CO<sub>2</sub> fixed min<sup>-1</sup> mL<sup>-1</sup> for a bioreactor culture that has an assumed average protein concentration of 1 mg mL<sup>-1</sup>. *P*-values relative to 7002 strain indicated significance at \**P* < 0.05, \*\**P* < 0.01

promoter and transformed into the host cell. It is known that D-ribulose-1,5-bisphosphate (RuBP), the product of PRK-catalyzed reaction, is toxic to *E. coli*. The growth of cells containing active PRK will thus be retarded or even repressed (Cai et al. 2014; Hudson et al. 1992; Mueller-Cajar et al. 2007). Rubisco converts RuBP into glycerate-3-phosphate (3PGA), an intermediate of glycolysis. The higher the activity of Rubisco, the quicker the growth inhibition can be relieved via catalytic processing of the

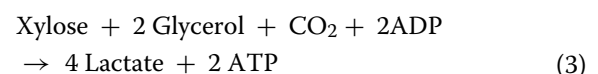
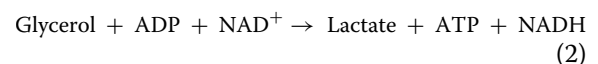
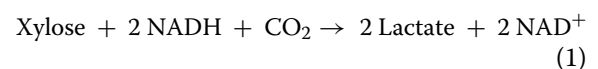
dead-end metabolite RuBP in these cells. By comparing the resulting growth profile, the relative carboxylation activity of different Rubiscos in vivo can be assessed.

Second, pyruvate is the end product of glycolysis, but since pyruvate can be easily channeled into different metabolism pathways, D-lactate was chosen as the end product for CO<sub>2</sub> capture. To maximize the carbon flux towards D-lactate, a gene *ldh* encoding D-lactate dehydrogenase from *Lactobacillus delbrueckii* was expressed,



while the gene *pflB* encoding pyruvate formate-lyase and the genes *frdABCD* encoding succinate dehydrogenase were inactivated to block the conversion of pyruvate to acetyl-CoA and succinate, respectively.

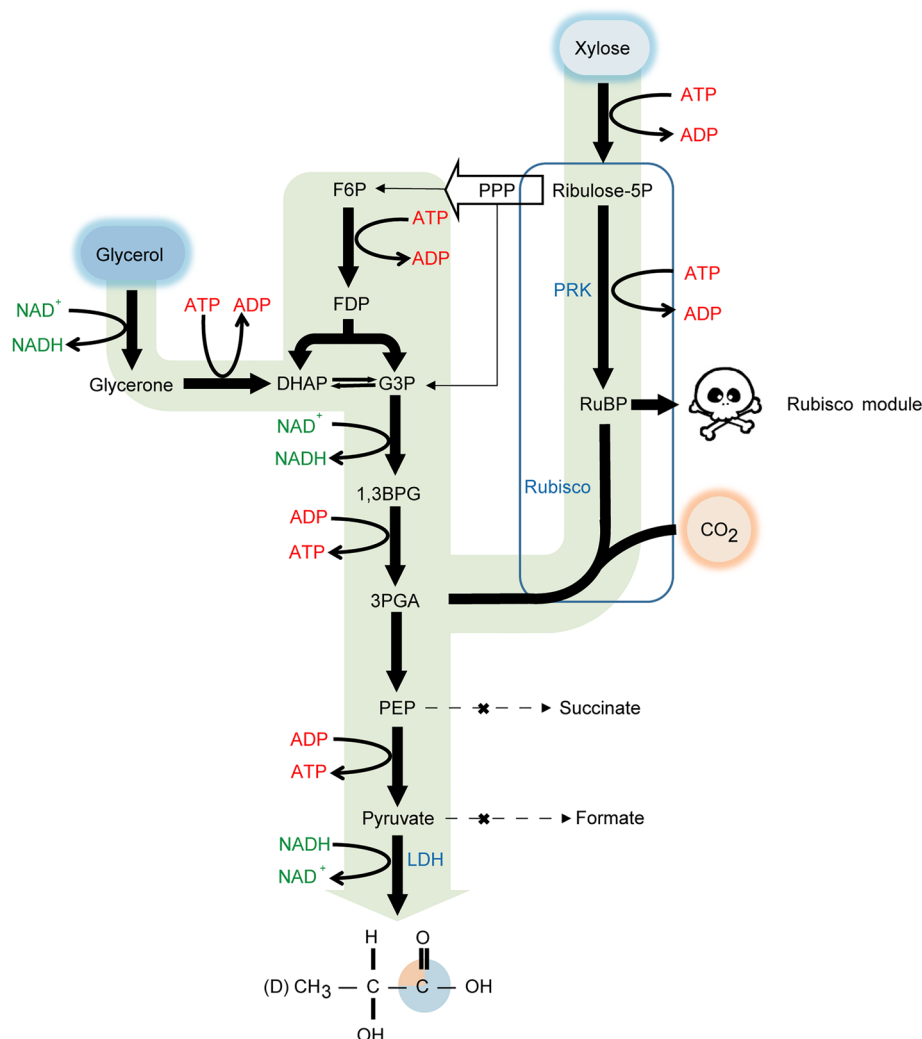
Third, it is expected that under ideal anaerobic conditions no CO<sub>2</sub> will be released once captured. Under such a condition, the NADH generated in association with xylose catabolism from xylose is insufficient if xylose is used as the sole substrate. Thus, additional NADH must be provided as shown in formula (1). The shortage of NADH can be provided by glycerol metabolism as shown in formula (2). Consequently, by using glycerol as a co-substrate, a general formula (3) could be deduced. In this conversion process, captured CO<sub>2</sub> will accumulate in the form of D-lactate. The amount of labeled fraction of CO<sub>2</sub> in D-lactate can be used to assess the efficiency of labeled CO<sub>2</sub> capture.



#### Characterization of Rubisco-based CO<sub>2</sub> capture in vivo

The model strain constructed following the above principles was designated as BWLac (BW25113Δ*frdABCD*Δ*pflB*::*ldhA*). BWLac strains with plasmids pET-RBC197-PRK, pET-RBC197-PRK2021, pET-7002-PRK and pET-RPE-PRK (Additional file 2: Table S1) were designated as strains 197, 197-2021,

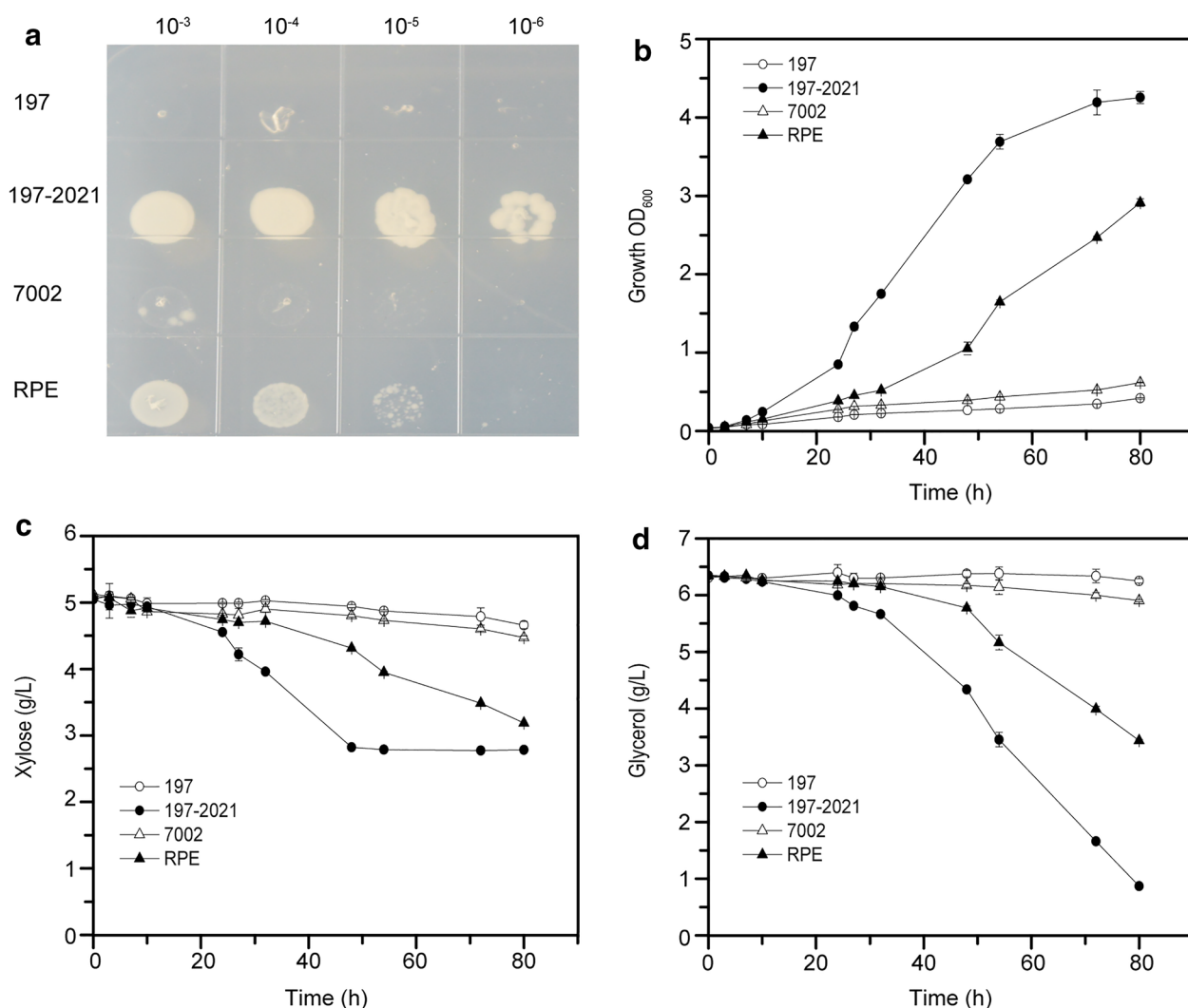




**Fig. 4** Schematic diagram of Rubisco carbon capture to produce D-lactate in *E. coli*. PRK, phosphoribulokinase, Rubisco, ribulose-1,5-bisphosphate carboxylase/ oxygenase, LDH, Lactate dehydrogenase, RuBP, ribulose 1,5-bisphosphate, 3PGA, 3-phosphoglycerate, PEP, phosphoenolpyruvate. "X" on the dotted arrows indicates that metabolic pathway was blocked. The skull and crossbones on the right side of RuBP indicates the toxicity induced by accumulation of RuBP. Orange quadrant of carboxyl carbon of the D-lactate indicates the exact position where the carbon of CO<sub>2</sub> is captured, and one out of four D-lactate produced, theoretically, would have captured carbon in that position

7002 and RPE, respectively. Strain 197 contains an inactive 7002 Rubisco as a K197M mutation was introduced into the conserved catalytic site of Rubisco large subunit (Cai et al. 2014). Strain 197-2021 contains an inactive 7002 Rubisco (K197M) and an inactive PRK where K20M/S21A mutations were introduced into the conserved nucleotide binding site of ATP-binding proteins (Cai et al. 2014; Higgins et al. 1986; Wilson et al. 2019a). According to the Rubisco-based CO<sub>2</sub> capture model, strain 197 containing an active PRK but inactive Rubisco was crippled by RuBP toxicity leading to drastic growth defects and therefore is used as a negative control for relief by Rubisco. Strain 197-2021 contains both inactive

PRK and inactive Rubisco leading to a near wild-type growth rate, which is used as control (Fig. 5a, b). Strains expressing RPE Rubisco grew much faster than strains utilizing 7002 Rubisco, in either plate or liquid culture (Fig. 5a, b). Compared with strain 7002 and the negative control strain 197, strain RPE was able to rapidly grow and metabolize both xylose and glycerol. The growth profiles of the four tested strains on agar plates (Fig. 5a) and in shake flasks (Fig. 5b) were consistent. This suggests that the RPE Rubisco quite effectively released the toxicity caused by the accumulation of RuBP. Interestingly, the positive control strain 197-2021, which showed rapid growth and achieved the highest biomass, did not



**Fig. 5** Growth profile of strains expressed by different Rubiscos. 197, negative control strain with active PRK but inactivated Rubisco; 197-2021, positive control strain with inactivated PRK and Rubisco; 7002, strain with PRK and 7002 Rubisco; RPE, strain with PRK and RPE Rubisco (details can be found in Additional file 2: Table S1). **a** Cell cultures were spotted on M9 minimal medium plate and incubated in 10% CO<sub>2</sub>, 90% air. **b** Cell growth of strains in M9 minimal medium under the same conditions as above. **c, d** Xylose and glycerol consumption of strains under the same conditions as above

consume xylose once the residual xylose concentration reached less than 2.8 g/L after 48 h, while glycerol could be continuously utilized (Fig. 5c, d). This suggests the overall activity of RPE Rubisco in strain RPE is very high, conferring the cell the capability to rapidly detoxify the RuBP generated by PRK and is consistent with the high activity of RPE Rubisco observed from cellular solution protein (Fig. 2b).

To quantitatively evaluate the CO<sub>2</sub> capture rate, BWLac strains harboring different plasmids were incubated aerobically to first induce protein expression, then subjected to catalyze CO<sub>2</sub> capture. In principle, D-lactate can also

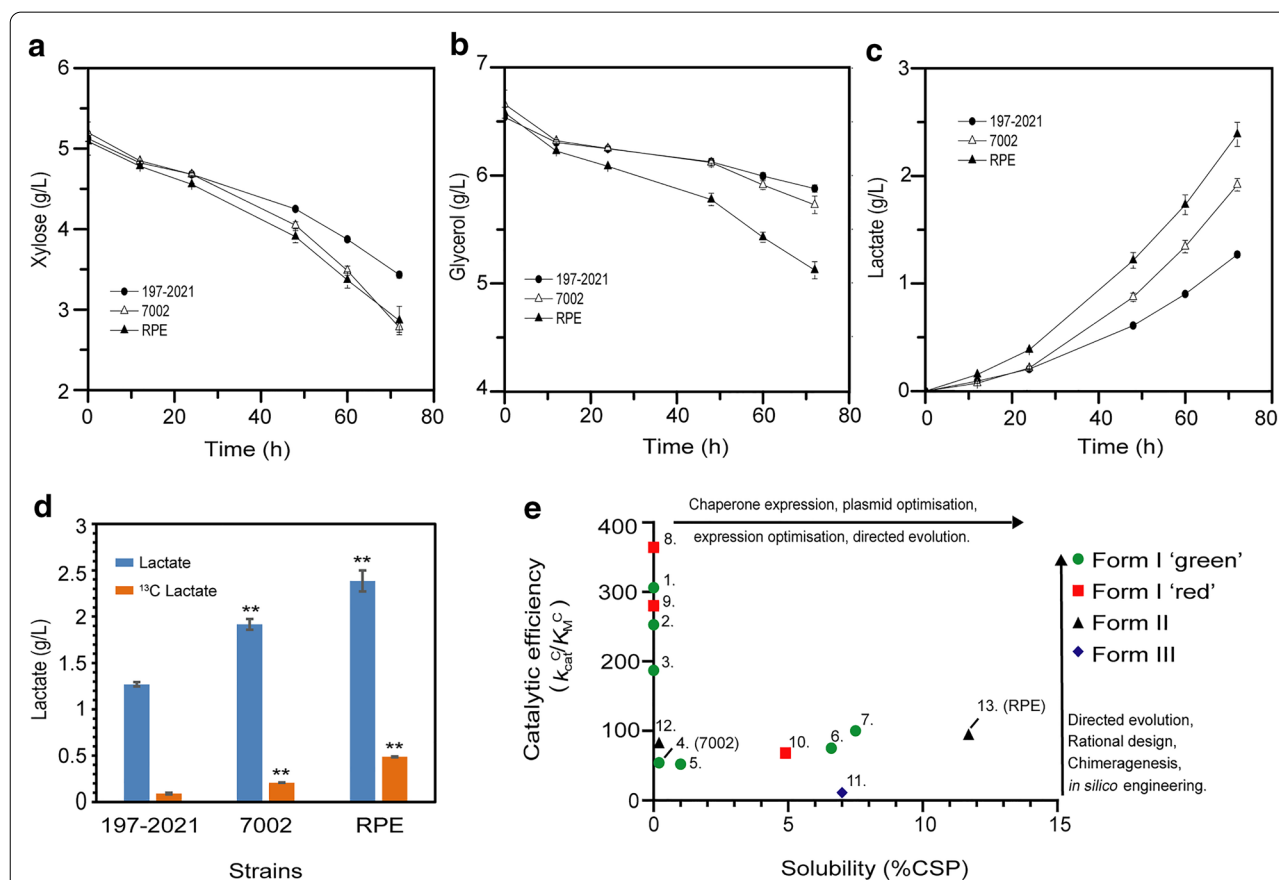
be produced by xylose metabolism through pentose phosphate pathway without Rubisco module. This means D-lactate can be produced by the positive control strain. Once the Rubisco module is introduced, the higher the carboxylation activity of Rubisco, the higher the production rate of lactate, and the more <sup>13</sup>C labeled D-lactate from <sup>13</sup>CO<sub>2</sub> assimilation can be detected (Fig. 4). Thus, the titer and ratio of <sup>13</sup>C-labeled lactate could be used to evaluate the efficiency of CO<sub>2</sub> capture.

In the carbon fixation experiment using the CO<sub>2</sub> capture model that we constructed, the consumption of xylose and glycerol as well as the production of lactate

of the positive control strain 197–2021 were the least among the three strains tested. The time-profile of strain 197–2021 serves as a background to calculate how much additional CO<sub>2</sub> can be fixed through Rubiscos in strains RPE and 7002. Interestingly, the xylose consumption of strain RPE was similar to that of strain 7002, but the glycerol consumption of strain RPE was much faster (Fig. 6a, b). As a consequence, the lactate production of strain RPE was the fastest and the highest, followed by strain 7002 and strain 197–2021 (Fig. 6c). The <sup>13</sup>C-labeled lactate concentrations from the fermentation broth collected at 72 h were analyzed, which were 0.09 ± 0.01 g/L, 0.21 ± 0.001 g/L, and 0.49 ± 0.003 g/L for strain 197–2021, strain 7002, and strain RPE, respectively (Fig. 6d). It is conceivable that the lower <sup>13</sup>C-labeled lactate titer of strain 7002 was due to the lower carboxylation activity of 7002 Rubisco resulting from its poor heterologous expression/

assembly in *E. coli*, which could also be reflected in its slow glycerol consumption (Fig. 6b and d). Overall, these results indicate that the Rubisco-based CO<sub>2</sub> capture model functions as designed.

In addition to the production of D-lactate, by-products acetate and a small amount of ethanol were detected, but no formate was detected because of the knockout of gene *pflB*. The carbon recovery of strain PRE reached up to 0.92, higher than that of strain 7002 and control strain 197–2021 (Table 2). As the carbon released from this process could not be quantified precisely, it is reasonable that none of the strains could achieve 100% carbon recovery. We also calculated the NADH/NAD<sup>+</sup> production/consumption based on the carbon flux data from Table 2. The net flux of NADH in strain RPE is higher than that of strain 7002 and the control strain 197–2021 (Table 2), which



**Fig. 6** Characterization of Rubisco-based CO<sub>2</sub> fixation ability in vivo. 197–2021, positive control strain with inactivated PRK and Rubisco; 7002, strain with PRK and 7002 Rubisco; RPE, strain with PRK and RPE Rubisco. Xylose, glycerol consumption **a**, **b** and D-lactate production **c**, **d** as methods described. The mean values and standard derivations of three independent repetitions were shown. *P*-values relative to 197–2021 indicated significance at \**P* < 0.05, \*\**P* < 0.01. **e** Plot of Rubisco carboxylation efficiency against heterologous solubility for the enzymes listed in Additional file 2: Table S3. For an organism expressing Rubisco for CO<sub>2</sub> capture under anoxic conditions (no oxygen competition) optimality corresponds to the upper right region of the plot. Relevant engineering strategies for improving in vivo CO<sub>2</sub> sequestration rates for different Rubisco enzymes are provided as a guide

**Table 2** Substrates consumption rates, products formation rates, carbon recovery and net flux of NADH in Rubisco-based CO<sub>2</sub> capture model

|   | Substrates consumption rates and products formation rates <sup>a</sup> (mmol gDCW <sup>-1</sup> h <sup>-1</sup> ) |              |             | Carbon recovery (C atoms) |              |             |
|---|---|--------------|-------------|---------------------------|--------------|-------------|
|   | 197-2021  | 7002         | RPE         | 197-2021                  | 7002         | RPE         |
| Xylose  | 0.41 ± 0.01   | 0.67 ± 0.03  | 0.50 ± 0.11 | 2.05 ± 0.05               | 3.35 ± 0.15  | 2.50 ± 0.55 |
| Glycerol  | 0.18 ± 0.03   | 0.28 ± 0.06  | 0.46 ± 0.07 | 0.54 ± 0.09               | 0.84 ± 0.18  | 1.38 ± 0.21 |
| Uptake <sup>13</sup> CO <sub>2</sub> <sup>b</sup> | –   | 0.05 ± 0.002 | 0.18 ± 0.04 | –                         | 0.05 ± 0.002 | 0.18 ± 0.04 |
| Lactate   | 0.57 ± 0.02   | 0.89 ± 0.03  | 1.01 ± 0.11 | 1.71 ± 0.06               | 2.67 ± 0.09  | 3.03 ± 0.33 |
| Acetate   | 0.11 ± 0.01   | 0.11 ± 0.02  | 0.18 ± 0.02 | 0.22 ± 0.02               | 0.22 ± 0.04  | 0.36 ± 0.04 |
| Ethanol   | 0.01 ± 0.002  | 0.06 ± 0.01  | 0.06 ± 0.01 | 0.02 ± 0.004              | 0.12 ± 0.02  | 0.12 ± 0.02 |
| Released CO <sub>2</sub> <sup>c</sup>             |   |              |             | 0.12 ± 0.01               | 0.17 ± 0.03  | 0.24 ± 0.03 |
| Total carbon consumption                          |   |              |             | 2.59 ± 0.14               | 4.24 ± 0.33  | 4.06 ± 0.80 |
| Total carbon production                           |   |              |             | 2.07 ± 0.10               | 3.18 ± 0.18  | 3.75 ± 0.42 |
| Carbon balance                                    |   |              |             | 0.80 ± 0.01               | 0.75 ± 0.02  | 0.92 ± 0.08 |
| Net flux of NADH <sup>d</sup>                     |   |              |             | 0.56 ± 0.01               | 0.60 ± 0.03  | 0.74 ± 0.08 |

<sup>a</sup> Calculated by the detected molar concentration divided by dry cell weight, one unit OD<sub>600</sub> = 0.3 gDCW/L (Soini et al. 2008), the cell density of samples remained stable around the initial 2 unit OD<sub>600</sub>

<sup>b</sup> The value of X<sub>2</sub> was calculated by Eqs. 1, 2, 3 based on <sup>13</sup>C-labeled lactate production

<sup>c</sup> Assumed that released CO<sub>2</sub> was equal to the summation of the moles of acetate and ethanol in the dissimilation of pyruvate

<sup>d</sup> Calculated by the ratio of NADH consumption to NADH production, NADH consumption accompanied the production of ethanol and lactic acid, NADH production was from the reaction of glycerol to glyceraldehyde and G3P to 1,3BPG, one molecule of NADH consumed or produced per reaction

is conceivable as the carbon fixation in strain RPE requires more NADH.

## Discussion

We report a highly active Form II Rubisco (RPE) from the endosymbiont of the deep-sea tubeworm *Riftia pachyptila*, which shows a 50.5% higher carboxylation efficiency ( $k_{\text{cat}}^{\text{C}}/K_{\text{M}}^{\text{C}}$ ) than that of the high-performance Form I Rubisco from *Synechococcus* sp. 7002. Crucially, RPE Rubisco expresses to high levels in *E. coli* (12% CSP) without additional chaperones, even compared to other Form II enzymes which generally express well in heterologous hosts (Davidi et al. 2020; Whitney and Andrews 2003). Many studies have reported the active expression of cyanobacterial Form I Rubisco in *E. coli* and other hosts (Lin et al. 2014; Occhialini et al. 2016; Wilson et al. 2018) has long been regarded as the pinnacle of carboxylation efficiency for Rubisco (Davidi et al. 2020). The soluble expression of 7002 Rubisco in *E. coli* in our work was only about 1% CSP, even in the presence of chaperones to boost yield (Fig. 2a). The soluble expression of plant Rubisco in *E. coli* is approximately 2% CSP as measured by [<sup>14</sup>C]-CABP binding (Aigner et al. 2017). The ease of soluble expression and assembly, together with a very high carboxylation efficiency, make RPE Rubisco an outstanding candidate as a biological conduit for CO<sub>2</sub> capture through synthetic biology. Emerging microbial platforms that can survive on CO<sub>2</sub> as a sole carbon source (Gassler et al. 2020; Gleizer et al. 2019)

under bioreactor conditions are ideal systems for RPE Rubisco as the atmospheric conditions can be user controlled. RPE Rubisco was previously reported as showing a poor specificity for substrate CO<sub>2</sub> over O<sub>2</sub> ( $S_{\text{C/O}}=8.6$ ) which is in line with the oxygenation sensitivity displayed by other Rubisco in the Form II clade (Robinson et al. 2003). Form II and Form III Rubiscos are universally poor at discriminating between CO<sub>2</sub> and O<sub>2</sub> as substrates and have low affinity for CO<sub>2</sub> compared to the Form I clade (Davidi et al. 2020; Whitney et al. 2011; Yang et al. 2021). These features render them completely inappropriate for CO<sub>2</sub> capture directly from air, and consequently, any relevance to photosynthetic improvement. However, as we show here, the CO<sub>2</sub> assimilated in the form of D-lactate by *E. coli* harboring RPE Rubisco was 3.6-fold higher than that of 7002 Rubisco, which underscores its potential as a tool for microbial CO<sub>2</sub> capture. Continued engineering of RPE Rubisco to further improve its carboxylation efficiency would push its kinetic profile further towards an increasingly optimal zone for high-efficiency microbial CO<sub>2</sub> capture (Fig. 6e).

Analysis of the original host and habitat in which RPE Rubisco was discovered provides further physiological insights into its superior carboxylation activity and a potential resource for similar isoforms to explore. RPE Rubisco was identified from the chemolithoautotrophic symbiont in the trophosome of giant tubeworm *R. pachyptila* that lives near CO<sub>2</sub>-rich and O<sub>2</sub>-poor deep-sea hydrothermal vents (Robinson et al. 1998). *R.*



*pachyptila* is probably the fastest-growing marine invertebrate (Lutz et al. 1994), but has no mouth or digestive tract. All organic carbon required to support its growth comes from the symbiont in its trophosome. The fast growth rate of the tubeworm in the sparsely populated deep sea can thus be partly explained by the superior CO<sub>2</sub>-fixation efficiency of RPE Rubisco. Rubiscos from similar trophosome and other endosymbionts inhabiting a range of host marine organisms that occupy similar environmental niches as *Riftia* provide may show similar utility for heterologous expression (Fig. 1b). These organisms exhibit frequent heterologous gene transfer (Li et al. 2018) which has likely adapted their Rubiscos enzymes to the remarkable yields observed in this study when expressed in *E. coli*. Interestingly, similar Rubisco found in the Eukaryotic endosymbionts (*Symbiodinium* and *Chromera*) of coral clade with the Rubisco in prokaryotic endosymbionts (Fig. 1b), further highlighting their unique adaptation for heterologous transfer, expression, and high-efficiency host CO<sub>2</sub> assimilation.

RPE Rubisco was from chemolithoautotrophic microbes living in deep sea. Contrary to the general impression that (photo) autotrophic organisms or extremophiles always grow slowly, it is reported that the doubling times of *Thiomicrospira crunogena* and *Hydrogenomonas thermophila*, two chemolithoautotrophic microbes from deep-sea hydrothermal vents, were only about 1 h (Dobranski et al. 2005) and 1.2 h (Takai et al. 2004), respectively. These doubling times were even faster than that of *Saccharomyces cerevisiae*, a well-known fast-growing heterotrophic microbe, indicating that the carbon fixation and metabolism in these chemolithoautotrophic microbes must be very efficient. Although photosynthesis by oxygenic photoautotrophs is the main contributor to CO<sub>2</sub> fixation on earth, our results suggest that CO<sub>2</sub> fixation by chemosynthesis of the chemolithoautotrophs might be more efficient and its potential may therefore be underestimated. A recent study (Davidi et al. 2020) also reported four Form II Rubisco (L<sub>2</sub>)<sub>n</sub> variants with  $k_{cat}^C$  higher than 7002 Rubisco, and three out of these four were also from chemolithoautotrophs, thus providing further evidence to our argument. There were no data on the heterologous expression and protein structure of these newly screened Form II Rubiscos in this report though (Davidi et al. 2020). It would thus be interesting to see whether these highly active Form II Rubiscos would be expressed/assembled in *E. coli* as well as RPE Rubisco, and whether they would exhibit similar potential for CO<sub>2</sub> capture.

Previously, the in vivo Rubisco-based carbon fixation efficiency was evaluated by indirect calculation (Tseng et al. 2018; Yang et al. 2016). In these models, it was difficult to precisely evaluate Rubisco-based CO<sub>2</sub> capture

ability due to pyruvate decarboxylation after CO<sub>2</sub> fixation. The Rubisco-based CO<sub>2</sub> capture model developed in this work secures the fixed CO<sub>2</sub> in the form of D-lactate, which is a carbon-conserving process. Using this model, we revealed that RPE Rubisco can help viable *E. coli* cells capture more CO<sub>2</sub> than 7002 Rubisco. It should be pointed out that xylose can also be metabolized through the pentose phosphate pathway to produce D-lactate. 36% of xylose in strain RPE was consumed to the metabolic flux of the CO<sub>2</sub>-fixation bypass pathway. However, only 7.5% of xylose in strain 7002, which was one-fifth of that of the RPE Rubisco. This suggests the Rubisco activity determines both the speed and efficiency of CO<sub>2</sub> capture.

Although there is great sequence and structural diversity among the Form I Rubiscos from oxygenic photoautotrophs and the Form II Rubiscos from chemolithoautotrophs, the two types of Rubiscos share some common features. The first one is the well-known trade-off between their carboxylation rate and affinity toward CO<sub>2</sub>. The most active Form I Rubisco from the oxygenic photoautotrophic cyanobacteria shows the lowest affinity toward CO<sub>2</sub>, with a  $K_M^C$  in the range of 200–250  $\mu$ M (Galmes et al. 2014; Hanson 2016; Whitney et al. 2011). RPE Rubisco, which has a higher carboxylation rate than the cyanobacterial Rubisco, also exhibits slightly higher CO<sub>2</sub> affinity, with a  $K_M^C$  of 172.4  $\mu$ M. The second one is the CO<sub>2</sub>-rich surroundings. The Rubisco enzymes from oxygenic photoautotrophic cyanobacteria and C<sub>4</sub> plants showing relatively high carboxylation activity but low CO<sub>2</sub> affinity usually rely on a carbon concentrating mechanism. This also holds true for the RPE Rubisco from a chemolithoautotroph. It is reported that partial pressure of CO<sub>2</sub> in the sea water was significantly elevated from 0.024 kPa in the vast region to 2.9 kPa around the tubeworm (Childress et al. 1993). Together with the reported high concentrations of carbonic anhydrase in the worm's plume and trophosome tissue, the internal total CO<sub>2</sub> concentration of *R. pachyptila* can reach up to 31 mM (Childress et al. 1993). Under such a high internal CO<sub>2</sub> concentration, the low affinity of RPE Rubisco toward CO<sub>2</sub> will not limit its carboxylation efficiency. *E. coli* does not have a carbon concentrating mechanism, but the slightly higher CO<sub>2</sub> affinity enabled *E. coli* harboring RPE Rubisco to capture CO<sub>2</sub> under the supply of 10% CO<sub>2</sub> in the environment. This implies *E. coli* harboring RPE Rubisco does not need a CO<sub>2</sub>-rich environment to play its carbon capture function.

It is generally known that extremely high expression of Rubisco is required in natural hosts to compensate for the shortage of its low carboxylation activity, which leads plant leaf protein content to consist up to 50% as Rubisco. Moreover, Rubisco from plant and cyanobacteria required their own specific chaperones to achieve such

a high-level expression. The simple structure, high carboxylation efficiency, easy heterologous soluble expression/assembly characteristics of RPE Rubisco make it an interesting enzyme for further carbon fixation research. Additional improvement to the CO<sub>2</sub> fixation rates in this study may be achieved by co-expression of RPE Rubisco activase genes (*cbbO/cbbQ*), as enzymatic deactivation in the absence of Rubisco activase has been shown to significantly reduce chemoheterotrophic Rubisco performance in heterologous systems (Gunn et al. 2020).

## Conclusion

The Form II RPE Rubisco from the endosymbiont of a deep-sea tubeworm *Riftia pachyptila* was identified and characterized in this study. RPE Rubisco has the potential to be used for carbon capture due to its higher carboxylation efficiency, easy heterologous soluble expression/assembly and simple hexamer structure. The CO<sub>2</sub> assimilation efficiency of *E. coli* harboring RPE Rubisco was 3.6-fold higher than that of 7002 Rubisco using a designed CO<sub>2</sub> capture model, demonstrating the application potential of RPE in microbial CO<sub>2</sub> capture.

## Abbreviations

Rubisco: Ribulose-1,5-bisphosphate carboxylase/oxygenase; RPE Rubisco: The endosymbiont of *Riftia pachyptila* Rubisco; 7002 Rubisco: *Synechococcus* sp. PCC7002 Rubisco; PRK: Phosphoribulokinase; CBB: Calvin–Benson–Bassham; Rbcl: The large subunit gene of each selected Rubisco; CABP: 2-Carboxyarabinitol-1,5-diphosphate; RuBP: D-Ribulose-1,5-bisphosphate; 3PGA: Glycerate-3-phosphate; Ribulose-5P: Ribulose-5-phosphate; F6P: Fructose-6-phosphate; G3P: Glyceraldehyde 3-phosphate; 1,3BPG: 1,3-Bisphosphoglycerate.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40643-021-00439-6>.

**Additional file 1.** Additional Rubisco sequences information.

**Additional file 2.** Additional tables and figures.

## Acknowledgements

We are grateful to Manajit Hayer-Hartl in Max Planck Institute of Biochemistry for her great help in <sup>14</sup>C-CABP synthesis and <sup>14</sup>CO<sub>2</sub> activity assay, we also thankful to Yan-ping Zhang in our lab for her insightful and helpful discussion in metabolic flux analysis. We thank Hong-Jie Zhang (Core Facility for Protein Research, Institute of Biophysics, Chinese Academy of Sciences) for the technique support using radioactivity detection. We thank Ying Fu from Public Technology service Center, Institute of Microbiology, Chinese Academy of Sciences for <sup>13</sup>C-labeled lactate detection by LC-MS/MS. We thank Field Rose Rogers of Massachusetts Institute of Technology for helpful advice on figure preparations.

## Authors' contributions

ZC and YL conceived and designed the project. JZ performed enzyme screening, characterization and structure analysis. GL designed the carbon capture model and performed associated analysis. AIC constructed the phylogenetic tree. RHW performed soluble expression analysis and analyzed data. JZ, GL, ZC, RHW and YL wrote the manuscript. All authors read and approved the final manuscript.

## Funding

This work was supported by the Key Research Program of the Chinese Academy of Sciences (Grant No. ZDRW-ZS-2016-3), the Youth Innovation Promotion Association of CAS (Zhen Cai), and the National Natural Science Foundation of China (No. 215060111), and Centre of Excellence for Biotechnology, Chinese Academy of Sciences—The World Academy of Sciences for the advancement of science in developing countries). AIC was financed by DAAD/BECAS Chile (No. 91643526).

## Availability of data and materials

The data and the materials are all available in this article as well as the Additional files 1 and 2.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interest

The authors declare no competing interest with regard to this manuscript.

## Author details

<sup>1</sup>CAS Key Laboratory of Microbial Physiological and Metabolic Engineering, State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China. <sup>2</sup>University of the Chinese Academy of Sciences, Beijing, China. <sup>3</sup>Department of Cellular Biochemistry, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany. <sup>4</sup>Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

Received: 2 June 2021 Accepted: 27 August 2021

Published online: 07 September 2021

## References

- Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58:1948–1954. <https://doi.org/10.1107/S0907444902016657>
- Aigner H, Wilson RH, Bracher A, Calisse L, Bhat JY, Hartl FU, Hayer-Hartl M (2017) Plant RuBisCo assembly in *E. coli* with five chloroplast chaperones including BSD2. *Science* 358:1272. <https://doi.org/10.1126/science.aap9221>
- Andersson I, Tjader AC, Cedergren-Zeppezauer E, Branden CI (1983) Crystallization and preliminary x-ray studies of spinach ribulose 1,5-bisphosphate carboxylase/oxygenase complexed with activator and a transition state analogue. *J Biol Chem* 258:14088–14090. [https://doi.org/10.1016/S0021-9258\(17\)43825-7](https://doi.org/10.1016/S0021-9258(17)43825-7)
- Banda DM, Pereira JH, Liu AK, Orr DJ, Hammel M, He C, Parry MAJ, Carmo-Silva E, Adams PD, Banfield JF, Shih PM (2020) Novel bacterial clade reveals origin of form I Rubisco. *Nat Plants* 6:1158–1166. <https://doi.org/10.1038/s41477-020-00762-4>
- Cai Z, Liu GX, Zhang JL, Li Y (2014) Development of an activity-directed selection system enabled significant improvement of the carboxylation efficiency of Rubisco. *Protein Cell* 5:552–562. <https://doi.org/10.1007/s13238-014-0072-x>
- Childress JJ, Lee RW, Sanders NK, Felbeck H, Oros DR, Toulmond A, Desbruyeres D, Kennicutt MC, Brooks J (1993) Inorganic carbon uptake in hydrothermal vent tubeworms facilitated by high environmental pCO<sub>2</sub>. *Nature* 362:147–149. <https://doi.org/10.1038/362147a0>
- Cleland WW, Andrews TJ, Gutteridge S, Hartman FC, Lorimer GH (1998) Mechanism of Rubisco: The carbamate as general base. *Chem Rev* 98:549–561. <https://doi.org/10.1021/cr970010r>
- Davidi D, Shamshoum M, Guo ZJ, Bar-On YM, Prywes N, Oz A, Jablonska J, Flamholz A, Wernick DG, Antonovsky N, de Pina B, Shachar L, Hochhauser D, Peleg Y, Albeck S, Sharon I, Mueller-Cajar O, Milo R (2020) Highly active

- rubiscos discovered by systematic interrogation of natural sequence diversity. *Embo J.* <https://doi.org/10.15252/embj.2019104081>
- Dobranski KP, Longo DL, Scott KM (2005) The carbon-concentrating mechanism of the hydrothermal vent chemolithoautotroph *Thiomicrospira crunogena*. *J Bacteriol* 187:5761–5766. <https://doi.org/10.1128/Jb.187.16.5761-5766.2005>
- Durao P, Aigner H, Nagy P, Mueller-Cajar O, Hartl FU, Hayer-Hartl M (2015) Opposing effects of folding and assembly chaperones on evolvability of Rubisco. *Nat Chem Biol* 11:148–U94. <https://doi.org/10.1038/Nchembio.1715>
- Emlyn-Jones D, Woodger FJ, Price GD, Whitney SM (2006) RbcX can function as a Rubisco chaperonin, but is non-essential in *Synechococcus* PCC7942. *Plant Cell Physiol* 47:1630–40. <https://doi.org/10.1093/pcp/pcl028>
- Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–32. <https://doi.org/10.1107/S0907444904019158>
- Galmes J, Kapralov MV, Andralojc PJ, Conesa MA, Keys AJ, Parry MA, Flexas J (2014) Expanding knowledge of the Rubisco kinetics variability in plant species: environmental and evolutionary trends. *Plant Cell Environ* 37:1989–2001. <https://doi.org/10.1111/pce.12335>
- Gasser T, Sauer M, Gasser B, Egermeier M, Troyer C, Causon T, Hann S, Mat-tanovich D, Steiger MG (2020) The industrial yeast *Pichia pastoris* is converted from a heterotroph into an autotroph capable of growth on CO<sub>2</sub>. *Nat Biotechnol* 38:210. <https://doi.org/10.1038/s41587-019-0363-0>
- Gleizer S, Ben-Nissan R, Bar-On YM, Antonovsky N, Noor E, Zohar Y, Jona G, Krieger E, Shamshoum M, Bar-Even A, Milo R (2019) Conversion of *Escherichia coli* to generate all biomass carbon from CO<sub>2</sub>. *Cell* 179:1255. <https://doi.org/10.1016/j.cell.2019.11.009>
- Gong FY, Liu GX, Zhai XY, Zhou J, Cai Z, Li Y (2015) Quantitative analysis of an engineered CO<sub>2</sub>-fixing *Escherichia coli* reveals great potential of heterotrophic CO<sub>2</sub> fixation. *Biotechnol Biofuels*. <https://doi.org/10.1186/s13068-015-0268-1>
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704. <https://doi.org/10.1080/10635150390235520>
- Gunn LH, Martin Avila E, Birch R, Whitney SM (2020) The dependency of red Rubisco on its cognate activase for enhancing plant photosynthesis and growth. *Proc Natl Acad Sci U S A* 117:25890–6. <https://doi.org/10.1073/pnas.2011641117>
- Hanson DT (2016) Breaking the rules of Rubisco catalysis. *J Exp Bot* 67:3180–2. <https://doi.org/10.1093/jxb/erw197>
- Higgins CF, Hiles ID, Salmond GP, Gill DR, Downie JA, Evans IJ, Holland IB, Gray L, Buckel SD, Bell AW et al (1986) A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria. *Nature* 323:448–50. <https://doi.org/10.1038/323448a0>
- Hinzke T, Kleiner M, Meister M, Schluter R, Hentschker C, Pane-Farre J, Hildebrandt P, Felbeck H, Sievert SM, Bonn F, Volker U, Becher D, Schweder T, Markert S (2021) Bacterial symbiont subpopulations have different roles in a deep-sea symbiosis. *Elife*. <https://doi.org/10.7554/eLife.58371>
- Hudson GS, Morell MK, Arvidsson YBC, Andrews TJ (1992) Synthesis of spinach phosphoribulokinase and ribulose 1,5-bisphosphate in *Escherichia coli*. *Aust J Plant Physiol* 19:213–21. <https://doi.org/10.1071/Pp9920213>
- Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059–66. <https://doi.org/10.1093/nar/gk436>
- Kubien DS, Brown CM, Kane HJ (2011) Quantifying the amount and activity of Rubisco in leaves. *Methods Mol Biol* 684:349–62. [https://doi.org/10.1007/978-1-60761-925-3\\_27](https://doi.org/10.1007/978-1-60761-925-3_27)
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) Procheck - a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–91. <https://doi.org/10.1107/S0021889892009944>
- Li YN, Liles MR, Halanych KM (2018) Endosymbiont genomes yield clues of tubeworm success. *Isme J* 12:2785–95. <https://doi.org/10.1038/s41396-018-0220-z>
- Lin MT, Occhialini A, Andralojc PJ, Parry MA, Hanson MR (2014) A faster Rubisco with potential to increase photosynthesis in crops. *Nature* 513:547–50. <https://doi.org/10.1038/nature13776>
- Lutz RA, Shank TM, Fornari DJ, Haymon RM, Lilley MD, Vondamm KL, Desbruyeres D (1994) Rapid growth at deep-sea vents. *Nature* 371:663–4. <https://doi.org/10.1038/371663a0>
- Mueller-Cajar O, Morell M, Whitney SM (2007) Directed evolution of Rubisco in *Escherichia coli* reveals a specificity-determining hydrogen bond in the form II enzyme. *Biochemistry-US* 46:14067–74. <https://doi.org/10.1021/bi700820a>
- Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240–55. <https://doi.org/10.1107/S0907444996012255>
- Occhialini A, Lin MT, Andralojc PJ, Hanson MR, Parry MAJ (2016) Transgenic tobacco plants with improved cyanobacterial Rubisco expression but no extra assembly factors grow at near wild-type rates if provided with elevated CO<sub>2</sub>. *Plant J* 85:148–60. <https://doi.org/10.1111/tpj.13098>
- Orr DJ, Worrall D, Lin MT, Carmo-Silva E, Hanson MR, Parry MAJ (2020) Hybrid cyanobacterial-tobacco Rubisco supports autotrophic growth and procarboxysomal aggregation. *Plant Physiol* 182:807–18. <https://doi.org/10.1104/pp.19.01193>
- Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326. [https://doi.org/10.1016/S0076-6879\(97\)76066-X](https://doi.org/10.1016/S0076-6879(97)76066-X)
- Pierce J, Tolbert NE, Barker R (1980) Interaction of ribulosebisphosphate carboxylase/oxygenase with transition-state analogues. *Biochemistry-US* 19:934–42. <https://doi.org/10.1021/bi00546a018>
- Price GD, Pengelly JLL, Forster B, Du JH, Whitney SM, von Caemmerer S, Badger MR, Howitt SM, Evans JR (2013) The cyanobacterial CCM as a source of genes for improving photosynthetic CO<sub>2</sub> fixation in crop species. *J Exp Bot* 64:753–68. <https://doi.org/10.1093/jxb/ers257>
- Robinson JJ, Stein JL, Cavanaugh CM (1998) Cloning and sequencing of a form II ribulose-1,5-bisphosphate carboxylase/oxygenase from the bacterial symbiont of the hydrothermal vent tubeworm *Riftia pachyptila*. *J Bacteriol* 180:1596–1599. <https://doi.org/10.1128/JB.180.6.1596-1599.1998>
- Robinson JJ, Scott KM, Swanson ST, O'Leary MH, Horken K, Tabita FR, Cavanaugh CM (2003) Kinetic isotope effect and characterization of form II Rubisco from the chemoautotrophic endosymbionts of the hydrothermal vent tubeworm *Riftia pachyptila*. *Limnol Oceanogr* 48:48–54. <https://doi.org/10.4319/lo.2003.48.1.0048>
- Satagopan S, Chan S, Perry LJ, Tabita FR (2014) Structure-function studies with the unique hexameric Form II Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) from *Rhodospseudomonas palustris*. *J Biol Chem* 289:21433–50. <https://doi.org/10.1074/jbc.M114.578625>
- Sela I, Ashkenazy H, Katoh K, Pupko T (2015) GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res* 43:W7–14. <https://doi.org/10.1093/nar/gkv318>
- Soini J, Ukkonen K, Neubauer P (2008) High cell density media for *Escherichia coli* are generally designed for aerobic cultivations - consequences for large-scale bioprocesses and shake flask cultures. *Microb Cell Fact* 7:26. <https://doi.org/10.1186/1475-2859-7-26>
- Tabita FR, Satagopan S, Hanson TE, Kreeel NE, Scott SS (2008) 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* 59:1515–24. <https://doi.org/10.1093/jxb/erm361>
- Takai K, Nealson KH, Horikoshi K (2004) *Hydrogenimonas thermophila* gen. nov., sp nov., a novel thermophilic, hydrogen-oxidizing chemolithoautotroph within the epsilon-Proteobacteria, isolated from a black smoker in a central indian ridge hydrothermal field. *Int J Syst Evol Micr* 54:25–32. <https://doi.org/10.1099/ijs.0.02787-0>
- Tseng IT, Chen YL, Chen CH, Shen ZX, Yang CH, Li SY (2018) Exceeding the theoretical fermentation yield in mixotrophic Rubisco-based engineered *Escherichia coli*. *Metab Eng* 47:445–52. <https://doi.org/10.1016/j.mbsen.2018.04.018>
- Varaljay VA, Satagopan S, North JA, Witte B, Dourado MN, Anantharaman K, Arbing MA, McCann SH, Oremland RS, Banfield JF, Wrighton KC, Tabita FR (2016) Functional metagenomic selection of ribulose 1,5-bisphosphate carboxylase/oxygenase from uncultivated bacteria. *Environ Microbiol* 18:1187–99. <https://doi.org/10.1111/1462-2920.13138>
- Whitney SM, Andrews TJ (2003) Photosynthesis and growth of tobacco with a substituted bacterial rubisco mirror the properties of the introduced enzyme. *Plant Physiol* 133:287–94. <https://doi.org/10.1104/pp.103.026146>
- Whitney SM, Houtz RL, Alonso H (2011) Advancing our understanding and capacity to engineer nature's CO<sub>2</sub>-sequestering enzyme, Rubisco. *Plant Physiol* 155:27–35. <https://doi.org/10.1104/pp.110.164814>

- Wilson RH, Martin-Avila E, Conlan C, Whitney SM (2018) An improved *Escherichia coli* screen for Rubisco identifies a protein-protein interface that can enhance CO<sub>2</sub>-fixation kinetics. *J Biol Chem* 293:18–27. <https://doi.org/10.1074/jbc.M117.810861>
- Wilson RH, Hayer-Hartl M, Bracher A (2019a) Crystal structure of phosphoribulokinase from *Synechococcus* sp. strain PCC 6301. *Acta Crystallogr F Struct Biol Commun* 75:278–89. <https://doi.org/10.1107/S2053230X19002693>
- Wilson RH, Thieulin-Pardo G, Hartl FU, Hayer-Hartl M (2019b) Improved recombinant expression and purification of functional plant Rubisco. *Febs Lett* 593:611–21. <https://doi.org/10.1002/1873-3468.13352>
- Yang CH, Liu EJ, Chen YL, Ou-Yang FY, Li SY (2016) The comprehensive profile of fermentation products during in situ CO<sub>2</sub> recycling by Rubisco-based engineered *Escherichia coli*. *Microb Cell Fact* 15:133. <https://doi.org/10.1186/s12934-016-0530-7>
- Yang F, Zhang J, Cai Z, Zhou J, Li Y (2021) Exploring the oxygenase function of Form II Rubisco for production of glycolate from CO(2). *AMB Express* 11:65. <https://doi.org/10.1186/s13568-021-01224-6>
- Zhou Y, Whitney S (2019) Directed evolution of an improved Rubisco; In vitro analyses to decipher fact from fiction. *Int J Mol Sci*. <https://doi.org/10.3390/ijms20205019>
- Zhuang ZY, Li SY (2013) Rubisco-based engineered *Escherichia coli* for in situ carbon dioxide recycling. *Bioresour Technol* 150:79–88. <https://doi.org/10.1016/j.biortech.2013.09.116>

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)