



**HAL**  
open science

# Crystal structure of chloroplast fructose-1,6-bisphosphate aldolase from the green alga *Chlamydomonas reinhardtii*

Théo Le Moigne, Edoardo Sarti, Antonin Nourisson, Mirko Zaffagnini,  
Alessandra Carbone, Stéphane Lemaire, Julien Henri

► **To cite this version:**

Théo Le Moigne, Edoardo Sarti, Antonin Nourisson, Mirko Zaffagnini, Alessandra Carbone, et al.. Crystal structure of chloroplast fructose-1,6-bisphosphate aldolase from the green alga *Chlamydomonas reinhardtii*. *Journal of Structural Biology*, 2022, 214 (3), pp.107873. 10.1016/j.jsb.2022.107873 . hal-03700038

**HAL Id: hal-03700038**

<https://hal.science/hal-03700038v1>

Submitted on 22 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **Crystal structure of chloroplast fructose-1,6-bisphosphate aldolase from the green alga**

2 ***Chlamydomonas reinhardtii***

3

4 Théo Le Moigne<sup>1,2,3</sup>, Edoardo Sarti<sup>1</sup>, Antonin Nourisson<sup>4</sup>, Mirko Zaffagnini<sup>5</sup>, Alessandra  
5 Carbone<sup>1</sup>, Stéphane D. Lemaire<sup>1,2</sup>, Julien Henri<sup>1,2,\*</sup>

6

7 1. Sorbonne Université, CNRS, UMR 7238, Institut de Biologie Paris-Seine, Laboratoire de  
8 Biologie Computationnelle et Quantitative, 4 place Jussieu F-75005 Paris, France

9 2. Sorbonne Université, CNRS, UMR 8226, Institut de Biologie Physico-Chimique,  
10 Laboratoire de Biologie Moléculaire et Cellulaire des Eucaryotes, 13 rue Pierre et Marie  
11 Curie, 75005 Paris, France

12 3. Faculty of Sciences, Doctoral School of Plant Sciences, Université Paris-Saclay, 91190  
13 Saint-Aubin, France

14 4. Institut Pasteur, Université de Paris, CNRS UMR3528, Unité Architecture et Dynamique  
15 des Macromolécules Biologiques, F-75015 Paris, France - Sorbonne Université, ED515  
16 Complexité du Vivant, F-75005 Paris, France

17 5. Department of Pharmacy and Biotechnology, University of Bologna, 40126, Bologna, Italy.

18 \* Corresponding author [julien.henri@sorbonne-universite.fr](mailto:julien.henri@sorbonne-universite.fr)

19

20 **Abstract**

21 The Calvin-Benson cycle fixes carbon dioxide into organic triosephosphates through the  
22 collective action of eleven conserved enzymes. Regeneration of ribulose-1,5-bisphosphate,  
23 the substrate of Rubisco-mediated carboxylation, requires two lyase reactions catalyzed by  
24 fructose-1,6-bisphosphate aldolase (FBA). While cytoplasmic FBA has been extensively  
25 studied in non-photosynthetic organisms, functional and structural details are limited for

26 chloroplast FBA encoded by oxygenic phototrophs . Here we determined the crystal structure  
27 of plastidial FBA from the unicellular green alga *Chlamydomonas reinhardtii* (Cr). We  
28 confirm that CrFBA folds as a TIM barrel, describe its catalytic pocket and homo-tetrameric  
29 state. Multiple sequence profiling classified the photosynthetic paralogs of FBA in a distinct  
30 group from non-photosynthetic paralogs. We mapped the sites of thiol- and phospho-based  
31 post-translational modifications known from photosynthetic organisms and predict their  
32 effects on enzyme catalysis.

33

## 34 **Introduction**

35 Photosynthesis is a paramount biological chain reaction sustaining life on Earth through the  
36 conversion of light energy into chemical energy and the subsequent fixation of atmospheric  
37 carbon into triosephosphates (Pfannschmidt and Yang 2012). The Calvin-Benson cycle (CBC)  
38 is an eleven enzymes cycle resulting in the production of six molecules of glyceraldehyde-3-  
39 phosphate (G3P) from three ribulose-1,5-bisphosphate (RuBP) and three carbon dioxides  
40 (Bassham, Benson et al. 1950). In order to regenerate RuBP and sustain the cycle, five G3P  
41 are recycled into three RuBP by a series of reactions involving eight out of eleven CBC  
42 enzymes. Two G3P are converted to dihydroxyacetone phosphate (DHAP), one is reacted  
43 with G3P to form fructose-1,6-bisphosphate (FBP) and the other is ligated with erythrose-4-  
44 phosphate (E4P) to form sedoheptulose-1,7-bisphosphate (SBP) (Sharkey 2019). Both  
45 reactions are catalyzed by fructose-bisphosphate aldolase (FBA, E.C. 4.1.2.13), which  
46 therefore acts in two distinct steps of the CBC. Indeed, it can use two different couples of  
47 substrates and provides substrates to downstream fructose-1,6-bisphosphatase (FBPase) and  
48 sedoheptulose-1,7-bisphosphatase (SBPase). *Pisum sativum* FBAs were isolated from plant,  
49 and their enzymatic activity assayed (Anderson and Pacold 1972) (Razdan, Henrikson et al.  
50 1992). Change in FBA abundance are correlated with the activity of the CBC, supporting a

51 potential control of FBA over photosynthetic carbon fixation (Haake, Zrenner et al. 1998)  
52 (Simkin, McAusland et al. 2015).

53 FBA is ubiquitous in the three phylogenetic kingdoms because it acts in the fourth step of  
54 glycolysis preparatory phase where it splits FBP into DHAP and G3P (Anderson and Advani  
55 1970) (Fothergill-Gilmore and Michels 1993). The nuclear genome of the green alga  
56 *Chlamydomonas reinhardtii* codes for at least four aldolases paralogs: FBA1 (Uniprot entry  
57 A8JCY4, Ensembl gene CHLRE\_01g006950v5), FBA2 (Uniprot entry A8I9H5, Ensembl  
58 gene CHLRE\_02g093450v5), ALDCHL/FBA3 (Uniprot entry Q42690, Ensembl gene  
59 CHLRE\_05g234550v5), and FBA4 (Uniprot entry A8I2Q9, Ensembl gene  
60 CHLRE\_02g115650v5). ALDCHL/FBA3 is predicted to be localized in the chloroplast by its  
61 amino-terminal 17-residue (Tardif, Atteia et al. 2012) or 27-residue transit peptide (Almagro  
62 Armenteros, Salvatore et al. 2019). Mature proteins containing 333/323 residues was  
63 quantified as the 9<sup>th</sup> most abundant protein in *Chlamydomonas* proteome (Schroda, Hemme et  
64 al. 2015), the third most abundant CBC enzyme after Rubisco large and small subunits  
65 (Hammel, Sommer et al. 2020). The estimated concentration of FBA3 is 4.2-10.1 amol/cell  
66 (Wienkoop, Weiss et al. 2010) (Hammel, Sommer et al. 2020), or 75.1-658.5  $\mu\text{mol/L}$  in the  
67 chloroplast stroma (Mettler, Muhlhaus et al. 2014) (Hammel, Sommer et al. 2020). DHAP  
68 and G3P substrates are respectively concentrated at 0.2 mmol/L and 0.01 mmol/L in  
69 *Chlamydomonas* illuminated by 41  $\mu\text{mol photons/m}^2/\text{s}$ , and their concentrations increase at  
70 0.7 mmol/L and 0.03 mmol/L under 145  $\mu\text{mol photons/m}^2/\text{s}$ . Substrate over enzyme ratios are  
71 low (0.4 G3P:FBA, 9 DHAP:FBA at 145  $\mu\text{mol photons/m}^2/\text{s}$  and 75.1  $\mu\text{mol/L}$  enzyme, the  
72 lowest estimate of FBA3 concentration), indicating that FBA functions far from substrates  
73 saturation. This undersaturation of the active site is further confirmed by the relatively high  
74  $K_M$  values reported for a bacterial FBA, which has Michaelis-Menten constants ( $K_M$ ) for  
75 DHAP and G3P of 0.095 mmol/L and 1.17 mmol/L, respectively (Rozova, Khmelenina et al.

76 2010). Moderate turnover numbers ( $k_{cat}$ ) were observed ranging from 0.2 to 65 sec<sup>-1</sup>.  
77 Although there is a lack of knowledge about the kinetic properties of the chloroplast isoforms,  
78 we can hypothesize that the catalytic capacity of FBA in the plastidial context is likely to  
79 exert a kinetic control over CBC metabolic flow.

80 FBA structures are classified into two groups, differentiated by their catalytic mechanism  
81 (Fothergill-Gilmore and Michels 1993) (Reyes-Prieto and Bhattacharya 2007). Class I  
82 aldolases form a Schiff-base intermediate with the substrate DHAP while class II aldolases  
83 are metal- dependent enzymes using divalent cations (Flechner, Gross et al. 1999). The two  
84 classes have been described as evolutionary independent (Marsh and Lebherz 1992). Despite  
85 folding as a similar  $\alpha/\beta$  TIM barrel (SCOPe lineage c.1.10), class I and II aldolases have  
86 dissimilar quaternary structures, class I aldolases being homotetrameric and class II aldolase  
87 homodimeric.

88 Mounting evidence supports that CBC enzymes are target of multiple post-translational  
89 modifications by means of reversible oxidation of cysteine thiols (reviewed in (Michelet,  
90 Zaffagnini et al. 2013) (Zaffagnini, Fermani et al. 2019) (Müller-Schüssele, Bohle et al.  
91 2021)) or threonine/serine phosphorylation ((Wang, Gau et al. 2014) (Werth, McConnell et al.  
92 2019). *CrFBA3* has been retrieved in the thioredoxome of *Chlamydomonas reinhardtii*  
93 supporting the relative affinity of FBA3 for thioredoxin through residue Cys58 (Lemaire,  
94 Guillon et al. 2004) (Perez-Perez, Mauries et al. 2017). *CrFBA3* was identified as a putative  
95 target of GSSG-mediated S-glutathionylation (Zaffagnini, Bedhomme et al. 2012), and it was  
96 also found undergoing S-nitrosylation on three cysteine sites (Cys58, Cys142, and Cys256  
97 (Morisse, Zaffagnini et al. 2014), and phosphorylation on Ser54, Thr57, Ser64, Ser170,  
98 Ser176 (Wang, Gau et al. 2014) (Werth, McConnell et al. 2019). Moreover, FBA was  
99 reported to be modified by lysine crotonylation in *Carica papaya* (Liu, Yuan et al. 2018).  
100 FBA is subjected to lysine acetylation in *Arabidopsis* (Finkemeier, Laxa et al. 2011), while

101 *Chlamydomonas* FBA was found to interact with CP12, a small intrinsically disordered  
102 protein specifically involved in the formation of inhibitory supercomplex with two CBC  
103 enzymes, namely phosphoribulokinase and glyceraldehyde-3-phosphate dehydrogenase (PRK  
104 and GAPDH, respectively) (Erales, Avilan et al. 2008) (Yu, Xie et al. 2020). Post-  
105 translational modifications and protein-protein interactions may support reversible regulatory  
106 mechanisms of protein catalysis but their interpretation requires a precise mapping of  
107 modified groups in the context of the protein structure.

108 A functional classification of the *Cr*FBA paralogs among their homologs could support the  
109 unique role of *Cr*FBA3 in the carbon fixation pathway. The sequence-based computational  
110 method ProfileView (Vicedomini, Bouly et al. 2019) has been recently designed to address  
111 the functional classification of the great diversity of homologous sequences hiding, in many  
112 cases, a variety of functional activities that cannot be anticipated. ProfileView relies on two  
113 main ideas: the use of multiple probabilistic models whose construction explores evolutionary  
114 information in large datasets of sequences (Bernardes, Zaverucha et al. 2016) (Ugarte,  
115 Vicedomini et al. 2018) (Vicedomini, Blachon et al. 2021) (Fortunato, Jaubert et al. 2016)  
116 (Amato, Dell'Aquila et al. 2017), and a new definition of a representation space where to look  
117 at sequences from the point of view of probabilistic models combined together. ProfileView  
118 has been previously applied to classify families of proteins for which functions should be  
119 discovered or characterized within known groups (Vicedomini, Blachon et al. 2021). It was  
120 proven very successful in identifying functional differences between otherwise  
121 phylogenetically similar sequences. Here, we meet a new challenge and use ProfileView to  
122 distinguish paralogs by their function in the FBA family.

123 In order to determine the structure-function-regulation relationships of photosynthetic FBA,  
124 we solved the high-resolution structure of the recombinant protein from *Chlamydomonas*, and

125 performed a computational analysis that confirms functional segregation of *CrFBA3* relative  
126 to the other FBA paralogs present in *Chlamydomonas*.

127

## 128 **Results**

### 129 ***CrFBA3* general structure**

130 The recombinant FBA3 from *Chlamydomonas reinhardtii* was purified to homogeneity and  
131 the crystal structure was solved at 2.36 Å resolution. The protein crystallized in the C2 space  
132 group (table 1) and the crystallographic model includes eight polypeptide chains of 329-332  
133 residues numbered according to the UniProtKB entry Q42690. Residual electron density at 1  
134  $\sigma$  additionally allowed us to place 29 sulfate ions, one chloride ion, and 989 water molecules  
135 in the crystal asymmetric unit. Amino acids prior to Met27 and after Thr360 have not been  
136 modelled because of a lack of electron density, as expected for peptidic extensions of low  
137 complexity. Alignment of the eight chains onto each other showed carbon alpha RMSD  
138 values comprised between 0.177 Å and 0.295 Å. All chains are therefore virtually identical  
139 except for some amino acid positions at N-terminal and/or C-terminal extremities, so we will  
140 focus on chain C for the following description, this chain being the most complete among the  
141 eight different chains of the asymmetric unit. Structural alignment of *CrFBA3* to the protein  
142 data bank archive returned 470 matches, with highest similarities towards *Arabidopsis*  
143 *thaliana* glycolytic FBA (PDB ID: 6RS1, Q-score = 0.88; RMSD = 0.93 Å), with  
144 *Plasmodium falciparum* FBA (PDB ID: 4TR9, Q-score = 0.87; RMSD = 0.89 Å), and with  
145 *Toxoplasma gondii* FBA (PDB ID: 4D2J, Q-score = 0.87; RMSD= 0.95 Å). *CrFBA3* is an  $\alpha/\beta$   
146 protein composed of eleven  $\alpha$ -helices and nine  $\beta$ -strands folded as a TIM barrel (CATH  
147 classification: 3.20.20) and numerated from H1 to H11 and S1 to S9, respectively (figure 1).  
148 Regions of the protein with the largest chain-to-chain differences are otherwise highlighted by  
149 the highest B-factors, and are mainly localized on the H2, H3, H4, H5 and H10 and H11  $\alpha$ -

150 helices and the loops that connect them (figure 2). These regions are located close to the  
151 active site of the protein and may allow rearrangements of its accessibility, in order to  
152 potentially accommodate substrates and products according to an induced fit mechanism.

153

#### 154 ***CrFBA3* active site**

155 Active site of *CrFBA3* is localized in the center of the TIM barrel at the C-terminal pole of  
156 the  $\beta$ -strands. This site is a strongly electropositive pocket, containing one or two sulfate ions  
157 captured in the crystallization solution and several water molecules (figure 3). Active site of  
158 class I aldolase have already been described structurally and chemically in various organisms  
159 (Dalby, Dauter et al. 1999) (St-Jean, Lafrance-Vanasse et al. 2005) (Lafrance-Vanasse and  
160 Sygusch 2007) (Gardberg, Abendroth et al. 2011) (Gardberg, Sankaran et al. 2011). Eleven  
161 catalytic residues described in these articles are present in *Chlamydomonas reinhardtii* FBA3  
162 at the following positions: Asp52, Ser54, Thr57, Lys125, Lys164, Arg166, Glu204, Glu206,  
163 Lys246, Ser288, and Arg318. Side chains of all catalytic residues are identically oriented  
164 placed in *Chlamydomonas reinhardtii* and mammals orthologs taken as references  
165 (supplementary figure 3). Human FBA has been described to perform its lyase activity by  
166 forming a Schiff base between the substrate and the lysine in position 229 while being  
167 stabilized by the two positively charged amino acids (Arg148 and Lys146) (Gamblin, Davies  
168 et al. 1991). The hypothesis of a similar catalytic mechanism between *CrFBA3* and other  
169 FBA is comforted by superposing *CrFBA3* structure to two FBA structures co-crystallized  
170 with their substrate in which amino acids from *CrFBA3* are well superposed with those found  
171 in homologs from *Homo sapiens* FBA (*HsFBA*) (PDB ID: 4ALD) and *Oryctolagus cuniculus*  
172 FBA (*OcFBA*) (PDB ID: 1ZAI) with RMSD of 0.813 Å and 0.886 Å, respectively (figure 3).  
173 Residues K229, R148, and K146 of human FBA align with residues K246, R166, and K164



174 of *CrFBA3* (figure 3A-B). The residues side chains are positioned in highly similar  
175 conformations (figure 3D).

176

### 177 ***CrFBA3* is a homo-tetramer**

178 *CrFBA3* crystal has an asymmetric unit containing eight quasi-identical protein chain. The  
179 analysis of the quaternary structure of recombinant *CrFBA3* protein indicated a homo-  
180 tetrameric state in solution, as the principal peak of protein elute at an approximative mass of  
181 160 kDa closely matching the mass of four monomers of 39 kDa each (supplementary figure  
182 1). Analytical size exclusion chromatography coupled to small angle X-ray scattering (SEC-  
183 SAXS) experiments confirmed the tetrameric state in solution (supplementary figure 2).  
184 Alteration of the redox state of *CrFBA3* induced by oxidizing or reducing treatments prior to  
185 SEC-SAXS slightly impacted the estimated mass of the protein in solution being 151-176  
186 kDa and 142-162 kDa for oxidized and reduced protein, respectively, and may suggest a more  
187 compact state for reduced FBA. Altogether, these results agree for a homo-tetrameric state of  
188 *CrFBA3* in solution, which is also described in other structures of class I FBA in other  
189 species (Marsh and Lebherz 1992). In order to evaluate the physiological oligomeric state of  
190 FBA3, we fractionated soluble extracts of *Chlamydomonas* cultures over a calibrated size-  
191 exclusion column and the presence of FBA3 in the eluted fractions was monitored by western  
192 blot using anti-FBA3 antibodies (figure 4). FBA3 eluted in one principal species with an  
193 apparent molecular weight of 170 kDa, consistent with that of a homotetramer.

194

### 195 ***CrFBA3* sites of post-translational modifications**

196 *CrFBA3* possesses six cysteines in its mature sequence. At least three cysteines may be  
197 targeted for redox post-translational modifications mediated by thioredoxins (Perez-Perez,  
198 Mauries et al. 2017). The two couples of cysteines Cys142-Cys183 and Cys218-Cys256 are

199 close enough, with respectively 8.3 Å and 10.2 Å between the two sulphur atoms, to form  
200 disulphide bonds. Recent redox-based proteomic studies identified Cys58, Cys142, and  
201 Cys256 as targets of S-nitrosylation while Cys58 and Cys142 were shown to be S-  
202 glutathionylated (Zaffagnini, Bedhomme et al. 2012). Despite being identified as putative  
203 redox target, Cys58 lateral chain is oriented toward the core of the protein and is not exposed  
204 to solvent (figure 5), raising the hypothesis that redox alteration of the thiol group of this  
205 residue demands local conformational changes or alternatively occurs in a later unfolded state  
206 of the protein. Also, Cys58 redox modification could be a co-translational modification rather  
207 than a post-translational modification occurring during the protein folding process. Cys142 is  
208 close to the catalytic cleft and to the external shell of the protein but is still poorly accessible  
209 to the solvent on our structure unless *CrFBA3* undergoes to a dynamic remodelling of  $\beta$ -  
210 strands 4 and 5 (figure 1). These  $\beta$ -strands are actually the ones with the highest B-factors  
211 (figure 2), supporting the hypothesis of a rearrangement of this part of the protein in function  
212 of the redox conditions. Cys256 is actually the most exposed thiol of *CrFBA3*, making it the  
213 best candidate for a redox control of the enzyme. However, this residue is located at 19.6 Å  
214 from the catalytic Arg166 (fig. 5). Such a distance prevents a direct effect of redox  
215 modification on protein catalysis but could induce local structural shifts that could alter  
216 substrate binding and/or substrate processing. Rabbit muscle aldolase is actually subject to a  
217 dithio-disulfide exchange at Cys72 and Cys338 that induces a conformational change  
218 propagating towards Cys239 at a 28-Å distance (PDB: 1ado and (Heyduk, Michalczyk et al.  
219 1991)).

220 Phosphoproteomic studies found five amino acids that are phosphorylated in *CrFBA3* (Wang,  
221 Gau et al. 2014) (Werth, McConnell et al. 2019) . These amino acids are dispersed in the  
222 primary structure, namely Ser54, Thr57, Ser64, Ser170, and Ser176. All corresponding  
223 sidechains are solvent exposed in *CrFBA3* crystal structure (figure 5). Ser54, Thr57, and

224 Ser64 are localized on the  $\alpha$ -helix 2 which is a key element of the catalytic pocket. In fact,  
225 Ser54 and Thr57 are part of the amino acids needed to accommodate the substrate in the  
226 active site (Dalby, Dauter et al. 1999) (St-Jean, Lafrance-Vanasse et al. 2005) (Gardberg,  
227 Abendroth et al. 2011). Phosphorylation of these residues could interfere with the fixation of  
228 the substrate both via steric and charge obstructions, supporting a negative regulatory  
229 mechanism widely conserved as these residues are conserved from photosynthetic organisms  
230 to animals, archaea, and bacteria (supplementary figure 3). Residues Ser170 and Ser176 are  
231 localized on the loop between  $\beta$ -strand 6 and  $\alpha$ -helix 6 separated by a  $3^{10}$  helix. This loop is  
232 on the other side of the catalytic site, bordering it. While Ser176 is conserved in FBA from all  
233 living organisms, Ser170 is only conserved in photosynthetic organisms.

234 Diverse post-translational modifications were reported on homologous class I aldolases, such  
235 as carbonylation (Lazarus, Buonora et al. 2015), and lysine methylations (Mininno, Brugière  
236 et al. 2012) or acetylation (Kim, Sprung et al. 2006, Barbosa Leite, Severo Gomes et al.  
237 2020). *Homo sapiens* ALDOA is glycated by methylglyoxal at residues Arg42, Lys146,  
238 Lys229, and Lys41 (Donnellan, Young et al. 2022), as well as acetylated at Lys12, Lys41, and  
239 Lys146 (Kim, Sprung et al. 2006). Arg42, Lys146 and Lys229 directly contribute to substrate  
240 binding (PDB entry 4ALD), while Lys12 is located 29.1 Å away from the catalytic site and  
241 Lys41 side chain points 8.7 Å away from the substrate, suggesting that an effect of post-  
242 translational modifications on activity require conformational changes. Strikingly, Lys41,  
243 Arg42 belong to  $\alpha$ -helix H2 (54-SNATCGKRLDS-64) where Ser54, Thr57, and S64 are  
244 phosphorylated while Cys58 is nitrosylated or glutathionylated. Helix 2 appears as a putative  
245 hotspot for multiple post-translational modifications, at an appropriate distance for allosteric  
246 control of FBA catalytic activity.

247

248 ***CrFBA3* is recognized as the Calvin-Benson paralog in *Chlamydomonas reinhardtii***

249 In order to demonstrate the implication of the paralog FBA3 in the CBC, we classified  
250 paralogous sequences from different photosynthetic species (Figure S5), whether belonging to  
251 the CBC or not, to find the subset of sequences involved in the CBC and verified that FBA3  
252 belongs to it.

253 Note that phylogeny-based algorithms are not suitable for this classification because FBA  
254 paralogous sequences share higher sequence identity than homologous sequences in different  
255 species and are clustered together by these computational methods despite their functional  
256 differences (Figure S6). Function-based clustering is known to be a difficult problem.

257 Here, we use ProfileView (Vicedomini, Bouly et al. 2022), a sequence-based computational  
258 method designed to functionally classify sets of homologous sequences and based on multiple  
259 profile Hidden Markov Models (pHMM; (Eddy 1998, Bernardes, Zaverucha et al. 2016,  
260 Ugarte, Vicedomini et al. 2018)). We demonstrate that ProfileView distinguishes paralogous  
261 sequences in the *C. reinhardtii* genome with respect to their function.

262 Starting from a set of 1244 FBA sequences (PF00247) covering a large diversity of species,  
263 ProfileView builds a library of 1244 pHMMs. After filtering them it uses 40 of them to define  
264 a multidimensional space of 2x40 dimensions where the FBA homologous sequences to be  
265 classified are represented. This set comprises 2601 homologous FBA sequences, from both  
266 prokaryotic and eukaryotic photosynthetic organisms. Among these sequences, there are the  
267 four *C. reinhardtii* FBA paralogs that constitute the object of our enquiry. Within the set, we  
268 also find FBA paralogs from other species that have been classified as CBC-related and non-  
269 CBC-related. ProfileView constructs a functional tree (Figure 6) organizing the input  
270 sequences by function (see Methods), where, intuitively, distinguished subtrees of sequences  
271 correspond to distinguished functions. In order to functionally interpret ProfileView  
272 distinguished subtrees, we used available functional knowledge on FBA homologous

273 sequences included in the tree and belonging to other species (*Volvox carteri*, *Arabidopsis*  
274 *thaliana*, *Synechocystis* sp. PCC 6803, see Table 2). We checked whether CBC-related  
275 paralogs were grouped under a common subtree or not. As shown in Figure 6A, ProfileView  
276 reconstructs three different subtrees, colored green, yellow and red. CBC and non-CBC-  
277 related sequences are located in different subtrees: the green one collects all known CBC  
278 sequences, whereas the yellow and red ones collect all known non-CBC ones. The two,  
279 yellow and red, subtrees contain almost exclusively sequences from unicellular and  
280 multicellular organisms, respectively. This means that the *CrFBA3* sequence is neatly  
281 separated from its paralogs *CrFBA1*, *CrFBA2* and *CrFBA4*: FBA3 appears in the Calvin-  
282 Benson subtree (Figure 6B), whereas the others appear in the unicellular non-Calvin-Benson  
283 subtree (Figure 6C).

284

## 285 **Conclusion**

286 We describe here the structure of chloroplastic FBA3 from *Chlamydomonas reinhardtii*. If  
287 nothing seems different in the catalytic mechanism of *CrFBA3* in comparison with already  
288 described class I FBA from other species, post-translational regulatory target sites of *CrFBA3*  
289 appear original and chloroplast-specific. The two Cys58 and Cys142, both target of S-  
290 glutathionylation and S-nitrosylation, are only conserved in photosynthetic organisms and  
291 therefore could be specific targets of a distinct and stress-related regulation of CBC. The  
292 amino acid Ser170, target of phosphorylation events, is in the same case as it is conserved  
293 only in photosynthetic organism. Alpha helix 2 contain residues Ser54, Thr57, Cys58, and  
294 Ser64 that are subject to post-translational modifications, it also has high B-factor in  
295 comparison to the other secondary elements and is close to the active site. These  
296 characteristics making it a probable regulation cluster and an important area in the activity of  
297 *CrFBA3* (Seoane and Carbone 2021).

298 This work also provides a computational analysis that allows a functional classification of the  
299 *CrFBA* paralogs among their homologs. This classification supports the unique role of  
300 *CrFBA3* in the Calvin-Benson cycle.

301 The computational approach addressing the functional classification of FBA paralogous  
302 proteins is presented in this work for the first time. Indeed, the ProfileView method  
303 (Vicedomini, Bouly et al. 2019) has similarly been used for functional classification of other  
304 protein sequences across species. Here, we demonstrate that, based on it, one can classify  
305 paralogous sequences within a species by function, a problem which might be considered  
306 different from the classification problem formulated on homology. The results obtained on the  
307 FBA protein family open the door to new applications in protein sequence classification and  
308 their putative physiological role.

309

## 310 **Experimental procedures**

### 311 Cloning

312 Nuclear encoded amino acids sequence of *Chlamydomonas reinhardtii* Fructose-bisphosphate  
313 aldolase 3 (Cre05.g234550, UniRef100 entry Q42690) was analyzed by TargetP2.0  
314 (Emanuelsson, Brunak et al. 2007) (Almagro Armenteros, Salvatore et al. 2019), ChloroP  
315 (Emanuelsson, Nielsen et al. 1999) and Predalgo (Tardif, Atteia et al. 2012) to predict the  
316 transit peptide . The subsequent mature sequence of chloroplastic protein coding for amino  
317 acid 28 to 377 has been amplified by PCR from the *Chlamydomonas reinhardtii* EST index  
318 (Asamizu and Nakamura 2004). PCR product was then inserted into pET3d-His6 by  
319 restriction and ligation at 5'-NdeI and 3'-BamHI sites to obtain a pET3d-His6-*CrFBA3*  
320 expression plasmid. Plasmid sequences were validated by Sanger sequencing.

321  
322 Protein expression and purification  
323 Transformation of *Escherichia coli* strain BL21(DE3) Rosetta-2 pLysS (Novagen Merck,  
324 Darmstadt Germany) were made with a pET3d-His6-CrFBA3 plasmid. This strain was grown  
325 in 2YT medium supplemented with ampicillin (100 µg/mL) at 37 °C until the culture reached  
326 exponential phase at an optical density at 600 nm of 0.8. The culture medium was  
327 supplemented with 0.2 mmol/L of IPTG and a temperature shift to 15 °C for 19 hours was  
328 made. Culture was next centrifugated at 4,000 g for 10 minutes at 4 °C. The cell pellet was  
329 resuspended in a buffer composed of 20 mM Tris-HCl (pH7.9), 100 mM NaCl, 5% glycerol  
330 and lyzed at a pressure of 2400 bar in a CF high-pressure homogenizer (Constant Systems  
331 Ltd., Daventry United Kingdom). The total extract of proteins was then centrifuged for 20  
332 min at 20,000 g and the soluble fraction loaded on an affinity chromatography with 2 ml of  
333 NiNTA resin (Sigma-Aldrich Merck, Darmstad Germany). The resin was washed with 20 ml  
334 of 20 mmol/L Tris-HCl (pH7.9) supplemented with 100 mM NaCl (buffer A) complemented  
335 with respectively 10 mmol/L and 20 mmol/L of imidazole. CrFBA3 was then eluted in 2  
336 consecutive steps with buffer A complemented with 250 mmol/L and 500 mmol/L of  
337 imidazole. Fractions were then analyzed by SDS-PAGE on a 12% acrylamide gel revealed by  
338 Coomassie blue staining. Fractions containing pure CrFBA3 were pooled, concentrated by  
339 ultrafiltration on 5,000 MWCO filter units (Millipore Merck, Darmstadt Germany) until a  
340 final volume of 10 ml prior to be injected on a size exclusion column HiLoad 26/600  
341 Superdex 200 pg (Cityva life sciences) in buffer A. Fractions containing pure CrFBA were  
342 pooled, concentrated by ultrafiltration on 5,000 MWCO filter units (Millipore Merck,  
343 Darmstadt Germany) to a final concentration of 9 mg.mL<sup>-1</sup> that was measured by NanoDrop  
344 2000 spectrophotometer (Thermo Fisher Scientific, Waltham MA USA) with theoretical Mw  
345 = 39,390 g.mol<sup>-1</sup> and ε<sub>280</sub> = 50,037 mol<sup>-1</sup>.L.cm<sup>-1</sup>.

346

347 Protein crystallization and structure determination

348 Purified *CrFBA3* was tested for crystallization at two concentration of 9 mg/ml and 4.5  
349 mg/ml on commercial sparse-screening conditions (Qiagen, Hilden Germany) based on the  
350 work of Jancarik and Kim (Jancarik, Scott et al. 1991) with a mixture of 50 nL of protein and  
351 50 nL of precipitant solution equilibrated against 40  $\mu$ L of reservoir solution at 20 °C. Several  
352 conditions showed small crystals in presence of ammonium sulfate. Condition A4 of Qiagen  
353 “the classics” kit (5% (v/v) isopropanol, 2 M ammonium sulfate) was selected to be optimized  
354 in a 2D gradient from 0% to 10% isopropanol and from 2 M to 2.5 M of ammonium sulfate  
355 with 2  $\mu$ l drops of protein/precipitant solution ratio 1/1 equilibrated against 1 mL of reservoir  
356 solution. Protein concentration was brought to 2.6 mg/mL to avoid formation of clustered  
357 crystals and monocrystals were grown in a solution of 5 % isopropanol and 2 M ammonium  
358 sulfate. Crystals were flash frozen in liquid nitrogen for diffraction experiment at Proxima-2  
359 beamline of SOLEIL synchrotron (Saint-Aubin, France). A 99.41 % complete dataset at 2.36  
360 Å resolution was obtained from 3600 images and indexed in the C2 space group, integrated,  
361 scaled and converted with XDSME (Legrand 2017). Structure was phased by molecular  
362 replacement with PHENIX (Adams, Afonine et al. 2010) (Adams, Afonine et al. 2011)  
363 PHASER-MR (McCoy, Grosse-Kunstleve et al. 2007) using a search model obtained from the  
364 Phyre2 server (Kelley, Mezulis et al. 2015). Eight *CrFBA* monomers were found in the  
365 asymmetric unit. Model was then refined by iterative cycle of manual building in WinCOOT  
366 (Emsley and Cowtan 2004) (Emsley, Lohkamp et al. 2010) followed by refinement with  
367 PHENIX.REFINE (Afonine, Grosse-Kunstleve et al. 2012) until completion of a structure  
368 passing MOLPROBITY (Chen, Arendall et al. 2010) evaluation with 99.5% residues in  
369 Ramachandran restrains RMS(bond) = 0.008, RMS(angles) = 0.93 and final Rwork = 0.2002,



370 Rfree = 0.2429 (Table 1). Structure representations were drawn with PYMOL 2.4  
371 (Schrodinger, New York USA).

372

373 Structural data

374 Reflections and coordinates files of the final model are registered in the protein data bank  
375 under accession code 7B2N.

376

377 Computational analysis

378 Generation of functional trees.

379 ProfileView (Vicedomini, Bouly et al. 2019) takes as input a set of homologous sequences  
380 and a protein domain, and returns a classification of the sequences in functional subgroups  
381 together with functional motifs characterising the subgroups. The first main idea of  
382 ProfileView is to extract conserved patterns from the space of available sequences through the  
383 construction of many probabilistic models for a protein family that should sample the  
384 diversity of the available homologous sequences and reflect shared structural and functional  
385 characteristics. These models are built as conservation profiles (pHMM [eddy1998profile]).

386 A pHMM can either generate new sequences, which are similar to the sequences used for its  
387 construction, or it can evaluate how likely it would be for a given query sequence to be  
388 generated by that model. To do so, it associates the query with an absolute likelihood score,  
389 which allows scores associated with different queries to be meaningfully compared.

390 ProfileView exploits this feature of pHMM for functionally describing an arbitrary set of  
391 query sequences. Indeed, the second main idea of ProfileView is to use its library of  
392 probabilistic models to embed input sequences into a multidimensional representation space,  
393 where each dimension is associated with a distinguished probabilistic model. Namely, each  
394 input sequence to be classified receives a score from each pHMM expressing how close the

395 pHMM is to the sequence. All the scores are recorded as a vector entry for the sequence.  
396 Protein sequences are thus represented as points on this space, and distances can be calculated  
397 between each pair. These distances do not reflect mere sequence identities, but rather capture  
398 functional aspects that would have been overlooked by simpler, sequence-related metrics.  
399 Lastly, the distances are used to generate a tree that functionally relates all the query  
400 sequences. The internal nodes of the tree are, whenever possible, annotated by representative  
401 pHMM and by functional motifs. We expect this tree to discriminate paralogs taking part in  
402 Calvin-Benson cycle, such as *C. reinhardtii* FBA3, from paralogs involved in other chemical  
403 pathways (FBA1, FBA2, FBA4).

404 The *C. reinhardtii* FBA3 protein sequence (UniProt code Q42690) contains only one protein  
405 domain which encompasses the most conserved part of the protein. It is the large Pfam  
406 domain Glycolytic PF00274. In order to generate ProfileView functional tree for the 2601  
407 homologous FBA sequences and including all *C. reinhardtii* FBA paralogous sequences, we  
408 devised the following pipeline:

409 (1) Selection of the sequences for the construction of the probabilistic models

410 To generate the ProfileView library of probabilistic models for the Glycolytic PF00274  
411 domain, we downloaded the Pfam “full” alignment (Pfam version 34 -  
412 <https://pfam.xfam.org/family/Glycolytic>). We favored this set of sequences because of their  
413 relatively limited number of gaps (high sequence length homogeneity), hand curation and  
414 uniform sampling of homologous sequences on their phylogenetic tree. Of the 3886  
415 sequences, 29% and 22% of them come from Proteobacteria and Metazoa respectively, and  
416 thus are not related to the Calvin-Benson cycle. Only 16% of sequences come from  
417 Viridiplantae. Nonetheless, it is important to select sequences covering most diversity as  
418 possible, in order to construct the most diversified model space for functional discrimination.  
419 No filter based on taxonomic classification was used in this step of the construction.

420 (2) Clustering of the selected sequences

421 Sequences having more than 20% gaps associated with amino acids of the query sequence  
422 were removed from the set, in order to filter out sequences with missing domains. The  
423 remaining 2356 sequences were clustered using MMseqs2 with parameters --min-seq-id 0.65  
424 -c 0.8. The parameters were chosen as in (Ugarte, Vicedomini et al. 2018) and on its  
425 observation that optimal performances are obtained with 1000-1500 sequences. The resulting  
426 set contains 1244 sequences and spans the original diversity of taxa (supplementary figure 5).

427 (3) Generation of the pHMM library

428 We used the ProfileView-build routine to generate a library composed of 1244 pHMM  
429 models. These models are compiled from an equal amount of automatically-generated  
430 alignments, containing a variable number of sequences gathered in Uniprot (Uniclust30  
431 08/2018).

432 (4) Generating the functional tree

433 To determine the functional differences between FBA paralogs in photosynthetic species, we  
434 considered the complete Pfam “uniprot” dataset (15967 sequences) and selected sequences to  
435 be classified in the ProfileView tree based on taxonomy: we selected taxa with Uniprot  
436 classification Cryptophyceae, Glaucocystophyceae, Rhodophyta, and Viridiplantae. In order  
437 to have the maximum number of available sequences in the tree, no length- nor sequence  
438 identity-based filtering was applied. We identified 2601 sequences (of which 96% are  
439 Viridiplantae, 3% Cryptophyceae, 2% Rhodophyta, and only 2 Glaucocystophyceae). Note  
440 that the overlap between the 1244 sequences which are “seed” for the model construction and  
441 the 2601 classified sequences amounts to 252 sequences.

442

443 Acknowledgements

444 This work was funded by ANR grants CALVINDESIGN (ANR-17-CE05-0001),  
445 CALVINTERACT (ANR-19-CE11-0009) and LABEX DYNAMO (ProjetIA-11-LABX-  
446 0011). We thank the *Institut de Biologie Physico-Chimique* (CNRS FR 550) for access to  
447 crystallization facility. We acknowledge SOLEIL for provision of synchrotron radiation  
448 facilities at beamlines Proxima-1, Proxima-2, and SWING. We thank Dr. Olivier Vallon for  
449 helpful discussions.

450

451 Conflict of interest

452 The authors declare that they have no conflict of interest.

453

454 Table 1. X-ray diffraction and crystallographic model statistics.

	<b>Fructose biphosphate aldolase</b>
<b>Wavelength (Å)</b>	0.9840
<b>Resolution range (Å)</b>	46.67-2.36 (2.44-2.36)
<b>Space group</b>	C2
<b>Unit cell (Å, °)</b>	109.49-251.07-126.411 90-90.11-90
<b>Total reflections</b>	988770 (90253)
<b>Unique reflexions</b>	138497 (13155)
<b>Multiplicity</b>	7.1 (6.9)
<b>Completeness (%)</b>	99.42 (94.42)
<b>Mean I/sigma(I)</b>	13.34 (1.46)
<b>Wilson B-factor (Å<sup>2</sup>)</b>	40.19
<b>R-merge</b>	0.2 (1.495)
<b>R-meas</b>	0.2159 (1.617)
<b>R-pim</b>	0.08067 (0.6089)
<b>CC1/2</b>	0.995 (0.566)
<b>CC*</b>	0.999 (0.85)
<b>Reflections used in refinement</b>	138424 (13104)
<b>Reflections used for R-free</b>	1984 (189)
<b>R-work</b>	0.2025 (0.3002)
<b>R-free</b>	0.2500 (0.3513)
<b>CC (work)</b>	0.954 (0.775)
<b>CC (free)</b>	0.921 (0.643)

<b>RMS (bonds) (Å)</b>	0.004
<b>RMS (angles) (°)</b>	0.63
<b>Ramachandran favoured (%)</b>	96.28
<b>Ramachandran allowed (%)</b>	3.34
<b>Ramachandran outliers (%)</b>	0.38
<b>Rotamer outliers (%)</b>	0.43
<b>Clashscore</b>	6.97
<b>Average B-factor (Å<sup>2</sup>)</b>	46.74

455

456

457 Table 2. Uniprot codes and the corresponding classification found by crossreferencing online  
458 databases. Classification was manually conducted by considering that Calvin-Benson FBA  
459 bear a chloroplast-addressing peptide, as predicted by ChloroP (Emanuelsson Nielsen et al.  
460 1999) and Predalgo (Tardif et al. 2012), and present a circadian expression available for  
461 *Chlamydomonas* (Strenkert Merchant et al. PNAS 2019) and *Arabidopsis* (Romanowski et al.  
462 2020).

463

Uniprot code	Uniprot accession numbers	Classification
Q42690_CHLRE	Q42690, A8JE10, Q36725	<i>C. reinhardtii</i> FBA3 (positive test)
A8JCY4_CHLRE	A8JCY4	<i>C. reinhardtii</i> FBA1 (negative test)
A8I9H5_CHLRE	A8I9H5	<i>C. reinhardtii</i> FBA2 (negative test)
A8I2Q9_CHLRE	A8I2Q9	<i>C. reinhardtii</i> FBA4

		(negative test)
A0A0K9Q9C0_SPIOL	A0A0K9Q9C0	Non-Calvin-Benson
A0A1P8B8S7_ARATH	A0A1P8B8S7	Non-Calvin-Benson
ALFP2_ARATH	Q944G9, Q5XEU6, Q9SVJ6	Non-circadian chloroplastic
A0A1P8B8Q8_ARATH	A0A1P8B8Q8	Non-Calvin-Benson
ALFC5_ARATH	O65581	Non-Calvin-Benson
ALFC8_ARATH	Q9LF98	Non-Calvin-Benson
A0A0K9QNV9_SPIOL	A0A0K9QNV9	Non-Calvin-Benson
D8TKY4_VOLCA	D8TKY4	Circadian chloroplastic
ALFC7_ARATH	P22197, O65582, Q53YH0	Non-Calvin-Benson
D8UAC8_VOLCA	D8UAC8	Non-Calvin-Benson
ALFP1_ARATH	Q9SJU4, Q2V473, Q93WF5, Q94C97	Circadian chloroplastic
F4IGL5_ARATH	F4IGL5	Circadian chloroplastic
F4IGL7_ARATH	F4IGL7	Circadian chloroplastic
ALFP3_ARATH	Q9ZU52	Non-circadian chloroplastic
A0A0K9R8B2_SPIOL	A0A0K9R8B2	Chloroplastic
D8U593_VOLCA	D8U593	Non-Calvin-Benson
ALFC4_ARATH	F4KGQ0, Q8LET3, Q8W4G8, Q9LZR9	Non-Calvin-Benson

D8UIJ0_VOLCA	D8UIJ0	Non-Calvin-Benson
A0A0K9QFF9_SPIOL	A0A0K9QFF9	Chloroplastic
F4JUI5_ARATH	F4JUI5	Non-circadian chloroplastic
A0A0K9QU20_SPIOL	A0A0K9QU20	Chloroplastic
ALFC6_ARATH	Q9SJI9	Non-Calvin-Benson
B3H6D7_ARATH	B3H6D7	Non-Calvin-Benson
A0A1P8BBB5_ARATH	A0A1P8BBB5	Non-Calvin-Benson
A0A0K9R0B2_SPIOL	A0A0K9R0B2	Chloroplastic

464

465 **Figure 1 : A.** Front view of *CrFBA3* represented as cartoon and colored from N-terminal to  
466 C-terminal from blue to red. Alpha helices and beta strands are named from H1 to H11 and  
467 from S1 to S9. Active site residues side chains are represented in sticks. **B.** Back view (180°  
468 rotation).

469

470 **Figure 2 :** Representation of the eight monomeric chains of *CrFBA3* of the asymmetric unit.  
471 Each amino acids is colored from blue to red and is thicker in function of their B-factor value,  
472 respectively with the lower B-factor blue and small and the higher thick and red.

473

474 **Figure 3 : A.** Stick representation of amino acids of the catalytic site of *CrFBA3*. Atoms are  
475 colored with the CPK colouring code with an exception of carbons which are colored in  
476 green. Two sulphate ions inside the pocket are also represented. **B.** Stick representation of  
477 amino acids of the catalytic site of *HsFBA* and a fructose-1,6-bisphosphate molecule from the  
478 PDB structure 4ALD aligned on the *CrFBA3* structure. Atoms are colored as in A. with  
479 carbons in light pink. **C.** Stick representation of amino acids of the catalytic site of *OcFBA*



480 and a fructose-1,6-bisphosphate molecule from the PDB structure 1ZAI aligned on the  
481 *CrFBA3* structure. Atoms are colored as in A. with carbons in teal. **D.** Superposition of the  
482 amino acids from A. B. and C.

483

484 **Figure 4 :** **A.** Cartoon representation of the homotetramer of *CrFBA3*. The different subunits  
485 are represented in red, green, orange and blue. **B.** Electrostatic surface of *CrFBA3* calculated  
486 by PyMOL APBS and represented in a gradient from blue (electropositive) to red  
487 (electronegative). *CrFBA3* is in the same orientation as in A. **C.** Size-exclusion chromatogram  
488 of *Chlamydomonas reinhardtii* cellular extracts separated on superose 6 10/300 column.  
489 Theoretical molecular masses from column calibration are indicated in kDa. **D.** Western blot  
490 with anti-*CrFBA3* antibody on chromatography fractions.

491

492 **Figure 5.** Redox post-translational modification sites colored in gold and phosphorylation  
493 sites colored in red on *CrFBA3*. Fructose-1,6-bisphosphate is modelled in cyan sticks, by  
494 alignment with the co-crystal structure of human FBA complexed with FBP (PDB entry  
495 4ALD). **A.** Surface representation of the FBA monomer. **B.** Cartoon representation of the FBA  
496 monomer, same orientation as in A.  $\mu$ =nitrosylation,  $\xi$ =glutathionylation.

497

498 **Figure 6:** FBA functional tree. ProfileView tree showing functional relationship between  
499 FBA sequences of photosynthetic organisms: **A.** The tree separates FBA paralogs participating  
500 in Calvin-Benson (green subtree) and not participating in Calvin-Benson (orange and red  
501 subtrees, for unicellular and multicellular organisms respectively). For *C. reinhardtii*, the  
502 paralog FBA3 is localized in the Calvin-Benson subtree, whereas the other three paralogs  
503 FBA1, FBA2, FBA4 are grouped together in a different subtree. **B.** Zoom on the Calvin-  
504 Benson subtree containing the *C. reinhardtii* FBA3 sequence. **C.** Detail of the Unicellular

505 Non-Calvin-Benson subtree, where the other *C. reinhardtii* FBA paralogs are found. In all  
506 panels, circles correspond to *C. reinhardtii* sequences, triangles to *V. carteri*, squares to *S.*  
507 *oleosa*, and stars to *A. thaliana*.

508

509

510 References

511

- 512 Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G.  
513 J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C.  
514 Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart (2010). "PHENIX: a comprehensive  
515 Python-based system for macromolecular structure solution." Acta Crystallogr D Biol Crystallogr  
516 **66**(Pt 2): 213-221.
- 517 Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, N. Echols, J. J. Headd, L. W. Hung, S. Jain, G. J.  
518 Kapral, R. W. Grosse Kunstleve, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, R. J. Read, D. C.  
519 Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart (2011). "The Phenix software for  
520 automated determination of macromolecular structures." Methods **55**(1): 94-106.
- 521 Afonine, P. V., R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C.  
522 Terwilliger, A. Urzhumtsev, P. H. Zwart and P. D. Adams (2012). "Towards automated crystallographic  
523 structure refinement with phenix.refine." Acta Crystallogr D Biol Crystallogr **68**(Pt 4): 352-367.
- 524 Almagro Armenteros, J. J., M. Salvatore, O. Emanuelsson, O. Winther, G. von Heijne, A. Elofsson and  
525 H. Nielsen (2019). "Detecting sequence signals in targeting peptides using deep learning." Life Sci  
526 Alliance **2**(5).
- 527 Amato, A., G. Dell'Aquila, F. Musacchia, R. Annunziata, A. Ugarte, N. Maillet, A. Carbone, M. Ribera  
528 d'Alcalà, R. Sanges, D. Iudicone and M. I. Ferrante (2017). "Marine diatoms change their gene  
529 expression profile when exposed to microscale turbulence under nutrient replete conditions." Sci  
530 Rep **7**(1): 3826.
- 531 Anderson, L. E. and V. R. Advani (1970). "Chloroplast and cytoplasmic enzymes: three distinct  
532 isoenzymes associated with the reductive pentose phosphate cycle." Plant Physiol **45**(5): 583-585.
- 533 Anderson, L. E. and I. Pacold (1972). "Chloroplast and Cytoplasmic Enzymes: IV. Pea Leaf Fructose  
534 1,6-Diphosphate Aldolases." Plant Physiol **49**(3): 393-397.
- 535 Asamizu, E. and Y. Nakamura (2004). "[System for expressed sequence tags (EST) analysis]."  
536 Tanpakushitsu Kakusan Koso **49**(11 Suppl): 1847-1852.
- 537 Barbosa Leite, A., A. A. Severo Gomes, A. C. de Castro Nascimento Sousa, M. Roberto de Mattos  
538 Fontes, S. Schenkman and N. Silvio Moretti (2020). "Effect of lysine acetylation on the regulation of  
539 Trypanosoma brucei glycosomal aldolase activity." Biochem J **477**(9): 1733-1744.
- 540 Bassham, J. A., A. A. Benson and M. Calvin (1950). "The path of carbon in photosynthesis." J Biol  
541 Chem **185**(2): 781-787.
- 542 Bernardes, J., G. Zaverucha, C. Vaquero and A. Carbone (2016). "Improvement in Protein Domain  
543 Identification Is Reached by Breaking Consensus, with the Agreement of Many Profiles and Domain  
544 Co-occurrence." PLoS Comput Biol **12**(7): e1005038.
- 545 Chen, V. B., W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray,  
546 J. S. Richardson and D. C. Richardson (2010). "MolProbity: all-atom structure validation for  
547 macromolecular crystallography." Acta Crystallogr D Biol Crystallogr **66**(Pt 1): 12-21.

548 Dalby, A., Z. Dauter and J. A. Littlechild (1999). "Crystal structure of human muscle aldolase  
549 complexed with fructose 1,6-bisphosphate: mechanistic implications." Protein Sci **8**(2): 291-297.  
550 Donnellan, L., C. Young, B. S. Simpson, M. Acland, V. S. Dhillon, M. Costabile, M. Fenech, P. Hoffmann  
551 and P. Deo (2022). "Proteomic Analysis of Methylglyoxal Modifications Reveals Susceptibility of  
552 Glycolytic Enzymes to Dicarboxyl Stress." Int J Mol Sci **23**(7).  
553 Eddy, S. R. (1998). "Profile hidden Markov models." Bioinformatics **14**(9): 755-763.  
554 Emanuelsson, O., S. Brunak, G. von Heijne and H. Nielsen (2007). "Locating proteins in the cell using  
555 TargetP, SignalP and related tools." Nat Protoc **2**(4): 953-971.  
556 Emanuelsson, O., H. Nielsen and G. von Heijne (1999). "ChloroP, a neural network-based method for  
557 predicting chloroplast transit peptides and their cleavage sites." Protein Sci **8**(5): 978-984.  
558 Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta  
559 Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-2132.  
560 Emsley, P., B. Lohkamp, W. G. Scott and K. Cowtan (2010). "Features and development of Coot." Acta  
561 Crystallogr D Biol Crystallogr **66**(Pt 4): 486-501.  
562 Eroles, J., L. Avilan, S. Lebreton and B. Gontero (2008). "Exploring CP12 binding proteins revealed  
563 aldolase as a new partner for the phosphoribulokinase/glyceraldehyde 3-phosphate  
564 dehydrogenase/CP12 complex--purification and kinetic characterization of this enzyme from  
565 *Chlamydomonas reinhardtii*." FEBS J **275**(6): 1248-1259.  
566 Finkemeier, I., M. Laxa, L. Miguet, A. J. Howden and L. J. Sweetlove (2011). "Proteins of diverse  
567 function and subcellular location are lysine acetylated in *Arabidopsis*." Plant Physiol **155**(4): 1779-  
568 1790.  
569 Flechner, A., W. Gross, W. F. Martin and C. Schnarrenberger (1999). "Chloroplast class I and class II  
570 aldolases are bifunctional for fructose-1,6-bisphosphate and sedoheptulose-1,7-bisphosphate cleavage  
571 in the Calvin cycle." FEBS Lett **447**(2-3): 200-202.  
572 Fortunato, A. E., M. Jaubert, G. Enomoto, J. P. Bouly, R. Raniello, M. Thaler, S. Malviya, J. S.  
573 Bernardes, F. Rappaport, B. Gentili, M. J. Huysman, A. Carbone, C. Bowler, M. R. d'Alcala, M. Ikeuchi  
574 and A. Falciatore (2016). "Diatom Phytochromes Reveal the Existence of Far-Red-Light-Based Sensing  
575 in the Ocean." Plant Cell **28**(3): 616-628.  
576 Fothergill-Gilmore, L. A. and P. A. Michels (1993). "Evolution of glycolysis." Prog Biophys Mol Biol  
577 **59**(2): 105-235.  
578 Gamblin, S. J., G. J. Davies, J. M. Grimes, R. M. Jackson, J. A. Littlechild and H. C. Watson (1991).  
579 "Activity and specificity of human aldolases." J Mol Biol **219**(4): 573-576.  
580 Gardberg, A., J. Abendroth, J. Bhandari, B. Sankaran and B. Staker (2011). "Structure of fructose  
581 bisphosphate aldolase from *Bartonella henselae* bound to fructose 1,6-bisphosphate." Acta  
582 Crystallogr Sect F Struct Biol Cryst Commun **67**(Pt 9): 1051-1054.  
583 Gardberg, A., B. Sankaran, D. Davies, J. Bhandari, B. Staker and L. Stewart (2011). "Structure of  
584 fructose bisphosphate aldolase from *Encephalitozoon cuniculi*." Acta Crystallogr Sect F Struct Biol  
585 Cryst Commun **67**(Pt 9): 1055-1059.  
586 Gascuel, O. (1997). "BIONJ: an improved version of the NJ algorithm based on a simple model of  
587 sequence data." Mol Biol Evol **14**(7): 685-695.  
588 Haake, V., R. Zrenner, U. Sonnewald and M. Stitt (1998). "A moderate decrease of plastid aldolase  
589 activity inhibits photosynthesis, alters the levels of sugars and starch, and inhibits growth of potato  
590 plants." Plant J **14**(2): 147-157.  
591 Hammel, A., F. Sommer, D. Zimmer, M. Stitt, T. Muhlhaus and M. Schroda (2020). "Overexpression of  
592 Sedoheptulose-1,7-Bisphosphatase Enhances Photosynthesis in *Chlamydomonas reinhardtii* and Has  
593 No Effect on the Abundance of Other Calvin-Benson Cycle Enzymes." Front Plant Sci **11**: 868.  
594 Heyduk, T., R. Michalczyk and M. Kochman (1991). "Long-range effects and conformational flexibility  
595 of aldolase." J Biol Chem **266**(24): 15650-15655.  
596 Jancarik, J., W. G. Scott, D. L. Milligan, D. E. Koshland, Jr. and S. H. Kim (1991). "Crystallization and  
597 preliminary X-ray diffraction study of the ligand-binding domain of the bacterial chemotaxis-  
598 mediating aspartate receptor of *Salmonella typhimurium*." J Mol Biol **221**(1): 31-34.

599 Kelley, L. A., S. Mezulis, C. M. Yates, M. N. Wass and M. J. Sternberg (2015). "The Phyre2 web portal  
600 for protein modeling, prediction and analysis." Nat Protoc **10**(6): 845-858.

601 Kim, S. C., R. Sprung, Y. Chen, Y. Xu, H. Ball, J. Pei, T. Cheng, Y. Kho, H. Xiao, L. Xiao, N. V. Grishin, M.  
602 White, X.-J. Yang and Y. Zhao (2006). "Substrate and Functional Diversity of Lysine Acetylation  
603 Revealed by a Proteomics Survey." Molecular Cell **23**(4): 607-618.

604 Lafrance-Vanasse, J. and J. Sygusch (2007). "Carboxy-terminus recruitment induced by substrate  
605 binding in eukaryotic fructose bis-phosphate aldolases." Biochemistry **46**(33): 9533-9540.

606 Lazarus, R. C., J. E. Buonora, D. M. Jacobowitz and G. P. Mueller (2015). "Protein carbonylation after  
607 traumatic brain injury: cell specificity, regional susceptibility, and gender differences." Free Radic Biol  
608 Med **78**: 89-100.

609 Legrand, P. (2017). "XDSME: XDS Made Easier." GitHub repository.

610 Lemaire, S. D., B. Guillon, P. Le Marechal, E. Keryer, M. Miginiac-Maslow and P. Decottignies (2004).  
611 "New thioredoxin targets in the unicellular photosynthetic eukaryote *Chlamydomonas reinhardtii*."  
612 Proc Natl Acad Sci U S A **101**(19): 7475-7480.

613 Liu, K., C. Yuan, H. Li, K. Chen, L. Lu, C. Shen and X. Zheng (2018). "A qualitative proteome-wide lysine  
614 crotonylation profiling of papaya (*Carica papaya* L.)." Sci Rep **8**(1): 8230.

615 Marsh, J. J. and H. G. Leberer (1992). "Fructose-bisphosphate aldolases: an evolutionary history."  
616 Trends Biochem Sci **17**(3): 110-113.

617 McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read (2007).  
618 "Phaser crystallographic software." J Appl Crystallogr **40**(Pt 4): 658-674.

619 Mettler, T., T. Muhlhaus, D. Hemme, M. A. Schottler, J. Rupprecht, A. Idoine, D. Veyel, S. K. Pal, L.  
620 Yaneva-Roder, F. V. Winck, F. Sommer, D. Vosloh, B. Seiwert, A. Erban, A. Burgos, S. Arvidsson, S.  
621 Schonfelder, A. Arnold, M. Gunther, U. Krause, M. Lohse, J. Kopka, Z. Nikoloski, B. Mueller-Roeber, L.  
622 Willmitzer, R. Bock, M. Schroda and M. Stitt (2014). "Systems Analysis of the Response of  
623 Photosynthesis, Metabolism, and Growth to an Increase in Irradiance in the Photosynthetic Model  
624 Organism *Chlamydomonas reinhardtii*." Plant Cell **26**(6): 2310-2350.

625 Michelet, L., M. Zaffagnini, S. Morisse, F. Sparla, M. E. Perez-Perez, F. Francia, A. Danon, C. H.  
626 Marchand, S. Fermani, P. Trost and S. D. Lemaire (2013). "Redox regulation of the Calvin-Benson  
627 cycle: something old, something new." Front Plant Sci **4**: 470.

628 Mininno, M., S. Brugière, V. Pautre, A. Gilgen, S. Ma, M. Ferro, M. Tardif, C. Alban and S. Ravanel  
629 (2012). "Characterization of chloroplastic fructose 1,6-bisphosphate aldolases as lysine-methylated  
630 proteins in plants." J Biol Chem **287**(25): 21034-21044.

631 Morisse, S., M. Zaffagnini, X. H. Gao, S. D. Lemaire and C. H. Marchand (2014). "Insight into protein S-  
632 nitrosylation in *Chlamydomonas reinhardtii*." Antioxid Redox Signal **21**(9): 1271-1284.

633 Müller-Schüssele, S. J., F. Bohle, J. Rossi, P. Trost, A. J. Meyer and M. Zaffagnini (2021). "Plasticity in  
634 plastid redox networks: evolution of glutathione-dependent redox cascades and glutathionylation  
635 sites." BMC Plant Biol **21**(1): 322.

636 Perez-Perez, M. E., A. Mauries, A. Maes, N. J. Tourasse, M. Hamon, S. D. Lemaire and C. H. Marchand  
637 (2017). "The Deep Thioredoxome in *Chlamydomonas reinhardtii*: New Insights into Redox  
638 Regulation." Mol Plant **10**(8): 1107-1125.

639 Pfanschmidt, T. and C. Yang (2012). "The hidden function of photosynthesis: a sensing system for  
640 environmental conditions that regulates plant acclimation responses." Protoplasma **249 Suppl 2**:  
641 S125-136.

642 Razdan, K., R. L. Heinrikson, H. Zurcher-Neely, P. W. Morris and L. E. Anderson (1992). "Chloroplast  
643 and cytoplasmic enzymes: isolation and sequencing of cDNAs coding for two distinct pea chloroplast  
644 aldolases." Arch Biochem Biophys **298**(1): 192-197.

645 Reyes-Prieto, A. and D. Bhattacharya (2007). "Phylogeny of Calvin cycle enzymes supports Plantae  
646 monophyly." Mol Phylogenet Evol **45**(1): 384-391.

647 Rozova, O. N., V. N. Khmelenina, Mustakhimov, II, A. S. Reshetnikov and Y. A. Trotsenko (2010).  
648 "Characterization of recombinant fructose-1,6-bisphosphate aldolase from *Methylococcus capsulatus*  
649 Bath." Biochemistry (Mosc) **75**(7): 892-898.

650 Schroda, M., D. Hemme and T. Muhlhaus (2015). "The Chlamydomonas heat stress response." Plant J  
651 **82**(3): 466-480.

652 Seoane, B. and A. Carbone (2021). "The complexity of protein interactions unravelled from structural  
653 disorder." PLoS Comput Biol **17**(1): e1008546.

654 Sharkey, T. D. (2019). "Discovery of the canonical Calvin-Benson cycle." Photosynth Res **140**(2): 235-  
655 252.

656 Simkin, A. J., L. McAusland, L. R. Headland, T. Lawson and C. A. Raines (2015). "Multigene  
657 manipulation of photosynthetic carbon assimilation increases CO<sub>2</sub> fixation and biomass yield in  
658 tobacco." J Exp Bot **66**(13): 4075-4090.

659 St-Jean, M., J. Lafrance-Vanasse, B. Liotard and J. Sygusch (2005). "High resolution reaction  
660 intermediates of rabbit muscle fructose-1,6-bisphosphate aldolase: substrate cleavage and induced  
661 fit." J Biol Chem **280**(29): 27262-27270.

662 Tardif, M., A. Atteia, M. Specht, G. Cogne, N. Rolland, S. Brugiere, M. Hippler, M. Ferro, C. Bruley, G.  
663 Peltier, O. Vallon and L. Cournac (2012). "PredAlgo: a new subcellular localization prediction tool  
664 dedicated to green algae." Mol Biol Evol **29**(12): 3625-3639.

665 Ugarte, A., R. Vicedomini, J. Bernardes and A. Carbone (2018). "A multi-source domain annotation  
666 pipeline for quantitative metagenomic and metatranscriptomic functional profiling." Microbiome  
667 **6**(1): 149.

668 Vicedomini, R., C. Blachon, F. Oteri and A. Carbone (2021). "MyCLADE: a multi-source domain  
669 annotation server for sequence functional exploration." Nucleic Acids Res **49**(W1): W452-W458.

670 Vicedomini, R., J. P. Bouly, E. Laine, A. Falciatore and A. Carbone (2019). "ProfileView: multiple  
671 probabilistic models resolve protein families functional diversity." bioRxiv: 717249.

672 Vicedomini, R., J. P. Bouly, E. Laine, A. Falciatore and A. Carbone (2022). "Multiple Profile Models  
673 Extract Features from Protein Sequence Data and Resolve Functional Diversity of Very Different  
674 Protein Families." Molecular Biology and Evolution **39**(4).

675 Wang, H., B. Gau, W. O. Slade, M. Juergens, P. Li and L. M. Hicks (2014). "The global  
676 phosphoproteome of Chlamydomonas reinhardtii reveals complex organellar phosphorylation in the  
677 flagella and thylakoid membrane." Mol Cell Proteomics **13**(9): 2337-2353.

678 Werth, E. G., E. W. McConnell, I. Couso Lianez, Z. Perrine, J. L. Crespo, J. G. Umen and L. M. Hicks  
679 (2019). "Investigating the effect of target of rapamycin kinase inhibition on the Chlamydomonas  
680 reinhardtii phosphoproteome: from known homologs to new targets." New Phytol **221**(1): 247-260.

681 Wienkoop, S., J. Weiss, P. May, S. Kempa, S. Irgang, L. Recuenco-Munoz, M. Pietzke, T. Schwemmer,  
682 J. Rupprecht, V. Egelhofer and W. Weckwerth (2010). "Targeted proteomics for Chlamydomonas  
683 reinhardtii combined with rapid subcellular protein fractionation, metabolomics and metabolic flux  
684 analyses." Mol Biosyst **6**(6): 1018-1031.

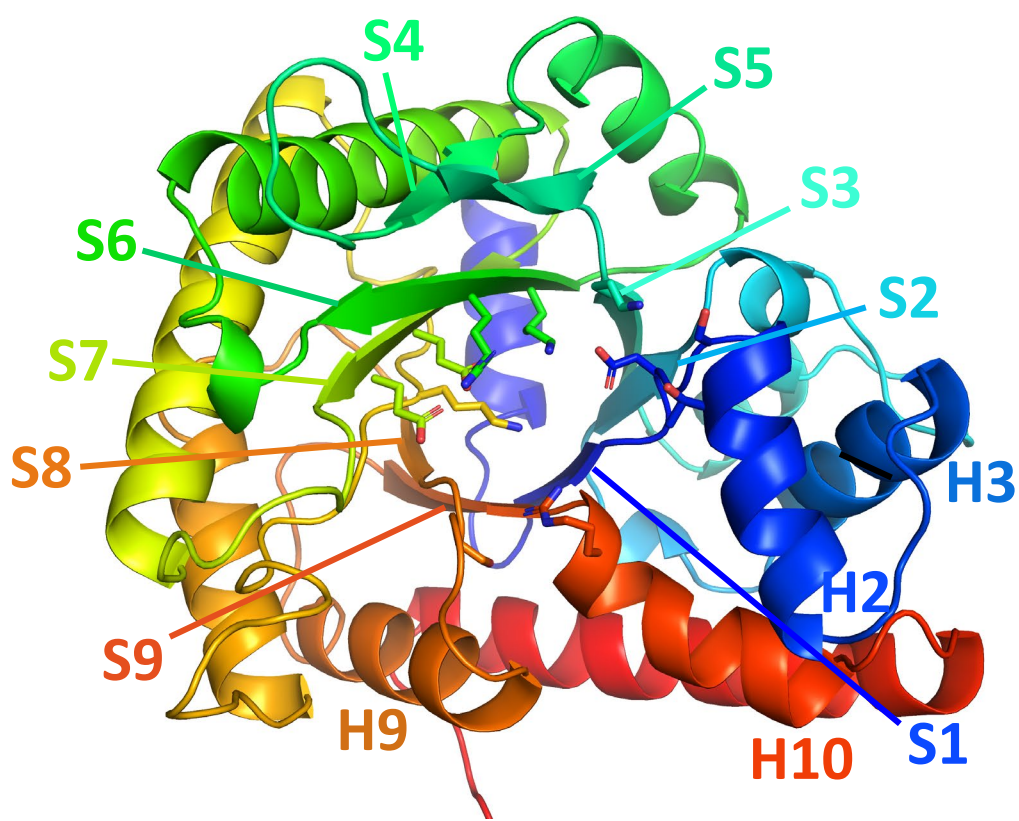
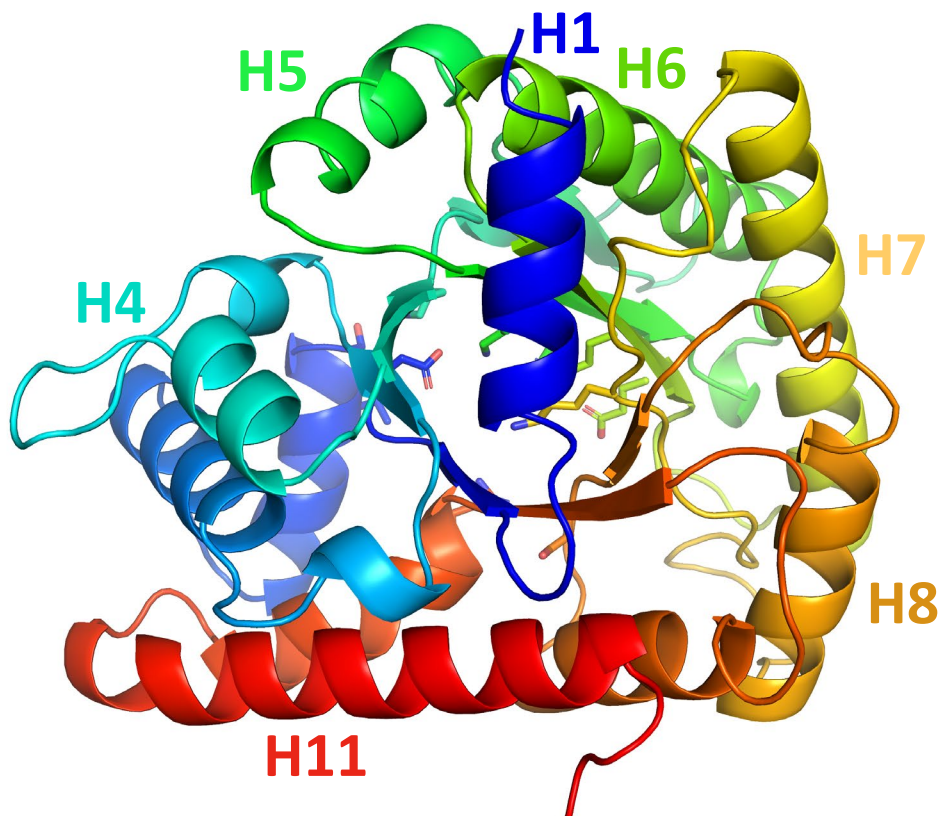
685 Yu, A., Y. Xie, X. Pan, H. Zhang, P. Cao, X. Su, W. Chang and M. Li (2020). "Photosynthetic  
686 Phosphoribulokinase Structures: Enzymatic Mechanisms and the Redox Regulation of the Calvin-  
687 Benson-Bassham Cycle." Plant Cell **32**(5): 1556-1573.

688 Zaffagnini, M., M. Bedhomme, H. Groni, C. H. Marchand, C. Puppo, B. Gontero, C. Cassier-Chauvat, P.  
689 Decottignies and S. D. Lemaire (2012). "Glutathionylation in the photosynthetic model organism  
690 Chlamydomonas reinhardtii: a proteomic survey." Mol Cell Proteomics **11**(2): M111.014142.

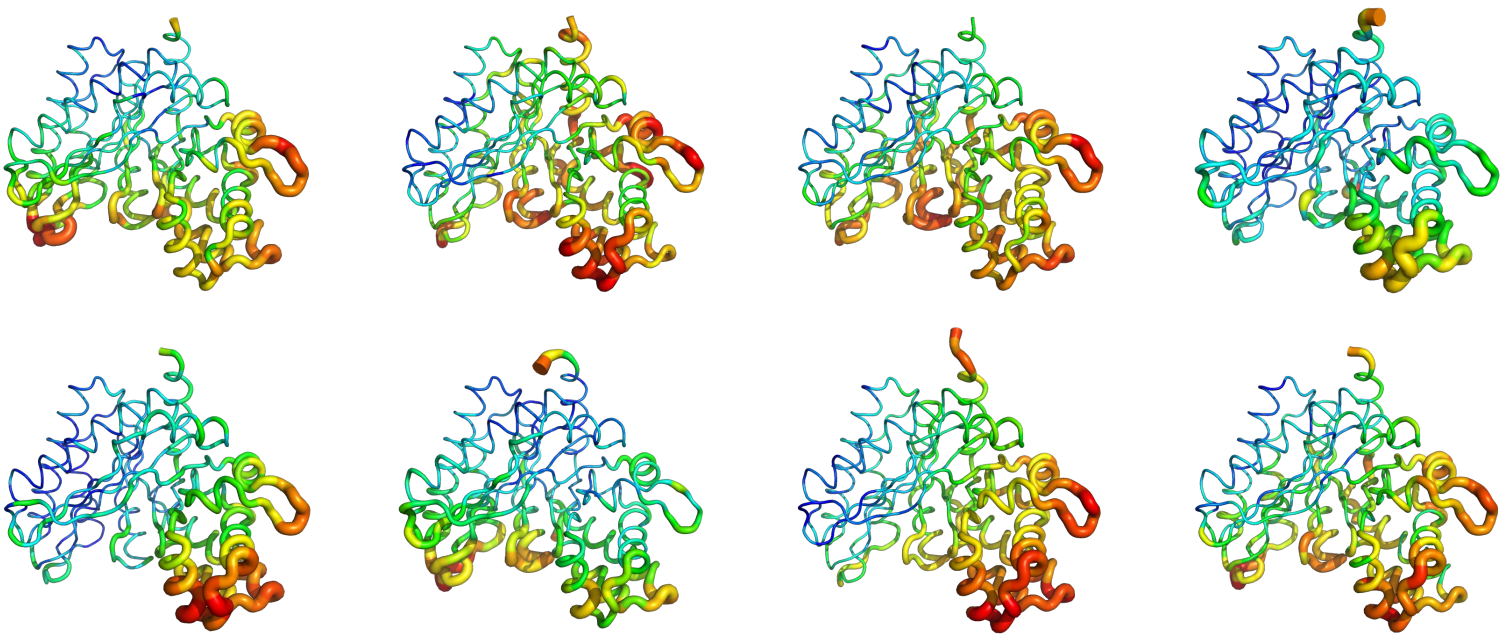
691 Zaffagnini, M., M. Bedhomme, H. Groni, C. H. Marchand, C. Puppo, B. Gontero, C. Cassier-Chauvat, P.  
692 Decottignies and S. D. Lemaire (2012). "Glutathionylation in the photosynthetic model organism  
693 Chlamydomonas reinhardtii: a proteomic survey." Mol Cell Proteomics **11**(2): M111.014142.

694 Zaffagnini, M., S. Fermani, C. H. Marchand, A. Costa, F. Sparla, N. Rouhier, P. Geigenberger, S. D.  
695 Lemaire and P. Trost (2019). "Redox Homeostasis in Photosynthetic Organisms: Novel and  
696 Established Thiol-Based Molecular Mechanisms." Antioxid Redox Signal **31**(3): 155-210.

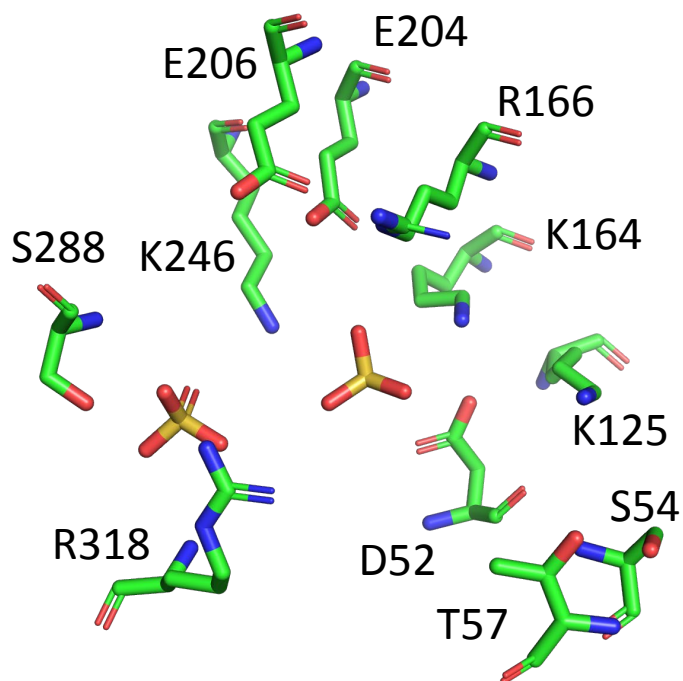
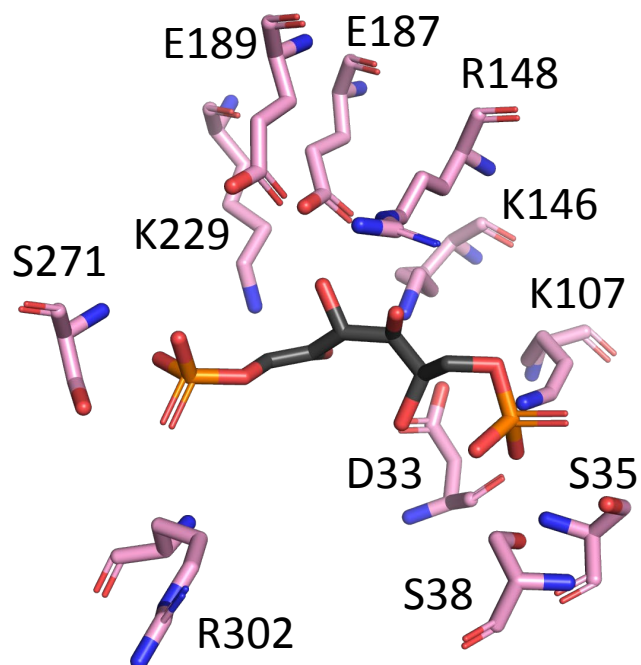
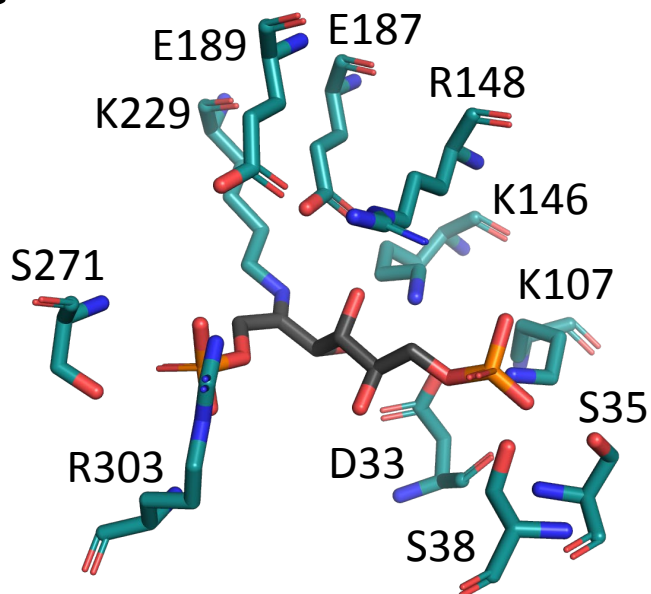
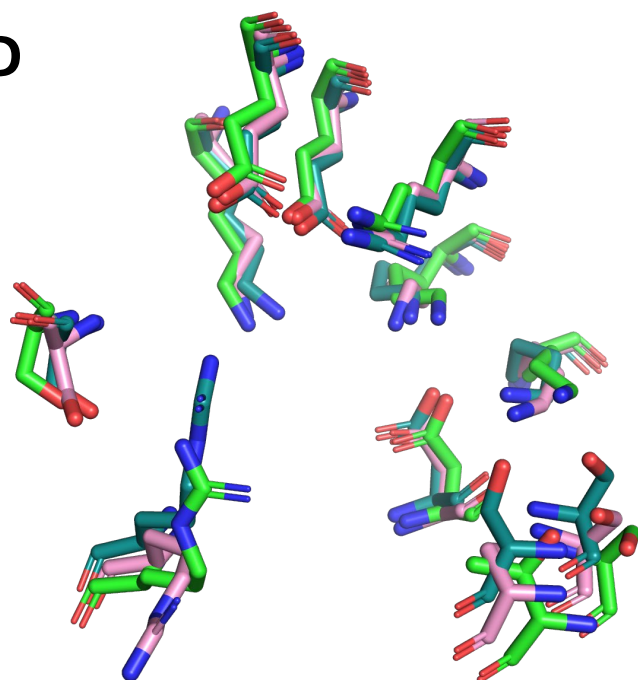
697

**A****B**

**Figure 1.** **A.** Front view of CrFBA3 represented as cartoon and colored from N-terminal to C-terminal from blue to red. Alpha helices and beta strands are named from H1 to H11 and from S1 to S9. Active site residues side chains are represented in sticks. **B.** Back view (180° rotation).

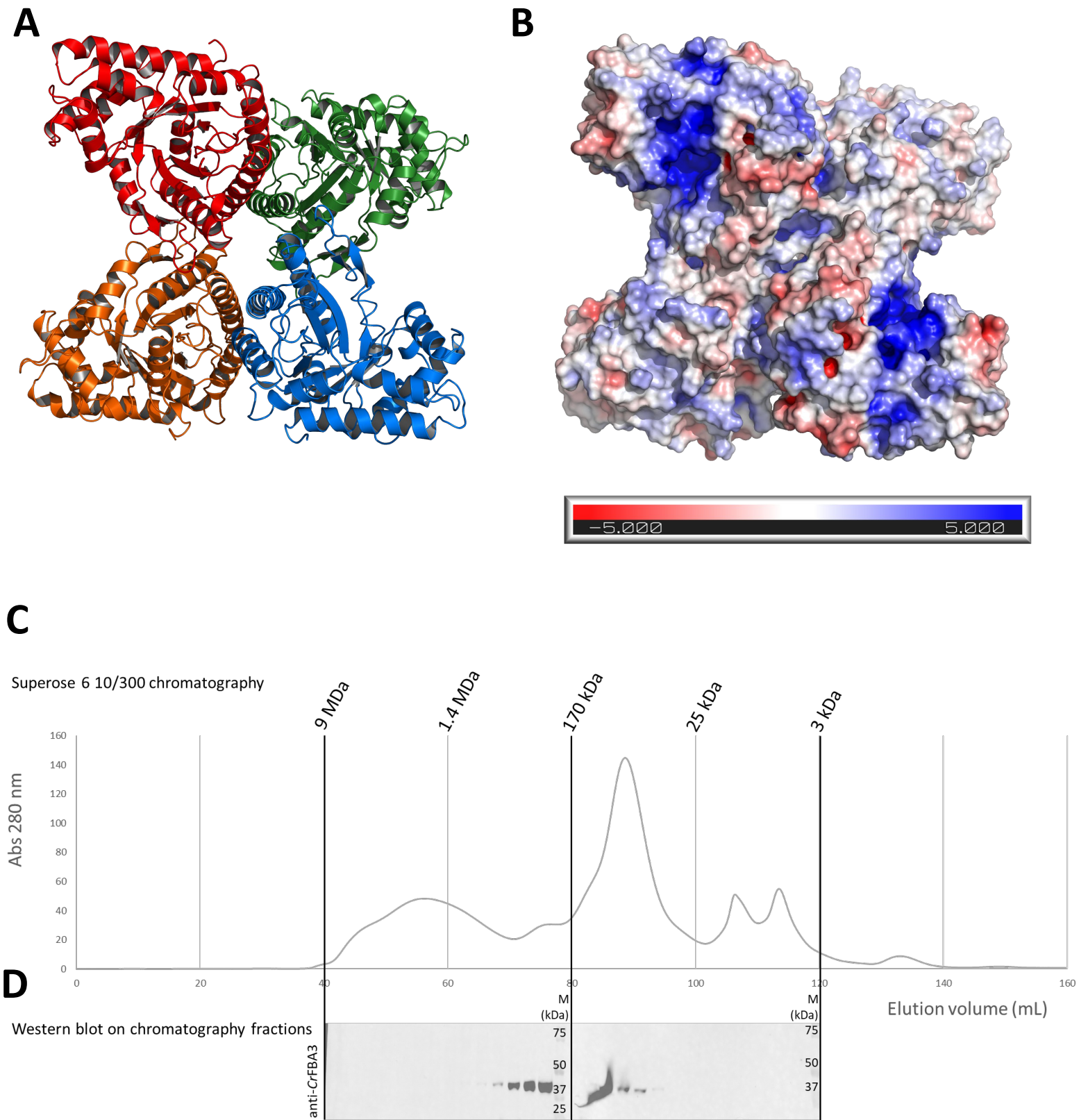


**Figure 2.** Representation of the eight monomeric chains of *CrFBA3* of the asymmetric unit. Each amino acids is colored from blue to red and is thicker in function of their B-factor value, respectively with the lower B-factor blue and small and the higher thick and red.

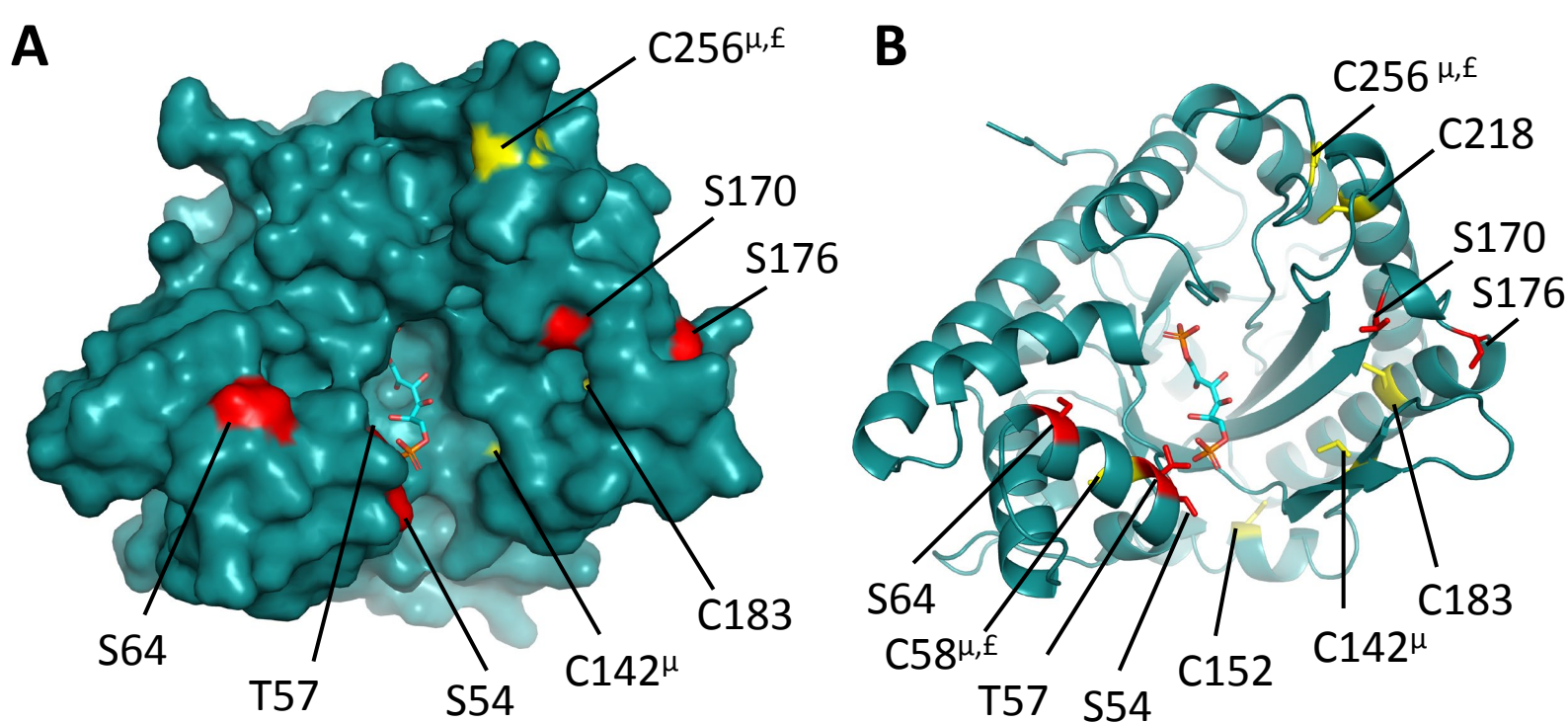
**A****B****C****D**

**Figure 3.** **A.** Stick representation of amino acids of the catalytic site of *CrFBA3*. Atoms are colored with the CPK colouring code with an exception of carbons which are colored in green. Two sulphate ions inside the pocket are also represented. **B.** Stick representation of amino acids of the catalytic site of *HsFBA* and a fructose-1,6-bisphosphate molecule from the PDB structure 4ALD aligned on the *CrFBA3* structure. Atoms are colored as in A. with carbons in light pink. **C.** Stick representation of amino acids of the catalytic site of *OcFBA* and a fructose-1,6-bisphosphate molecule from the PDB structure 1ZAI aligned on the *CrFBA3* structure. Atoms are colored as in A. with carbons in teal. **D.** Superposition of the amino acids from A. B. and C.

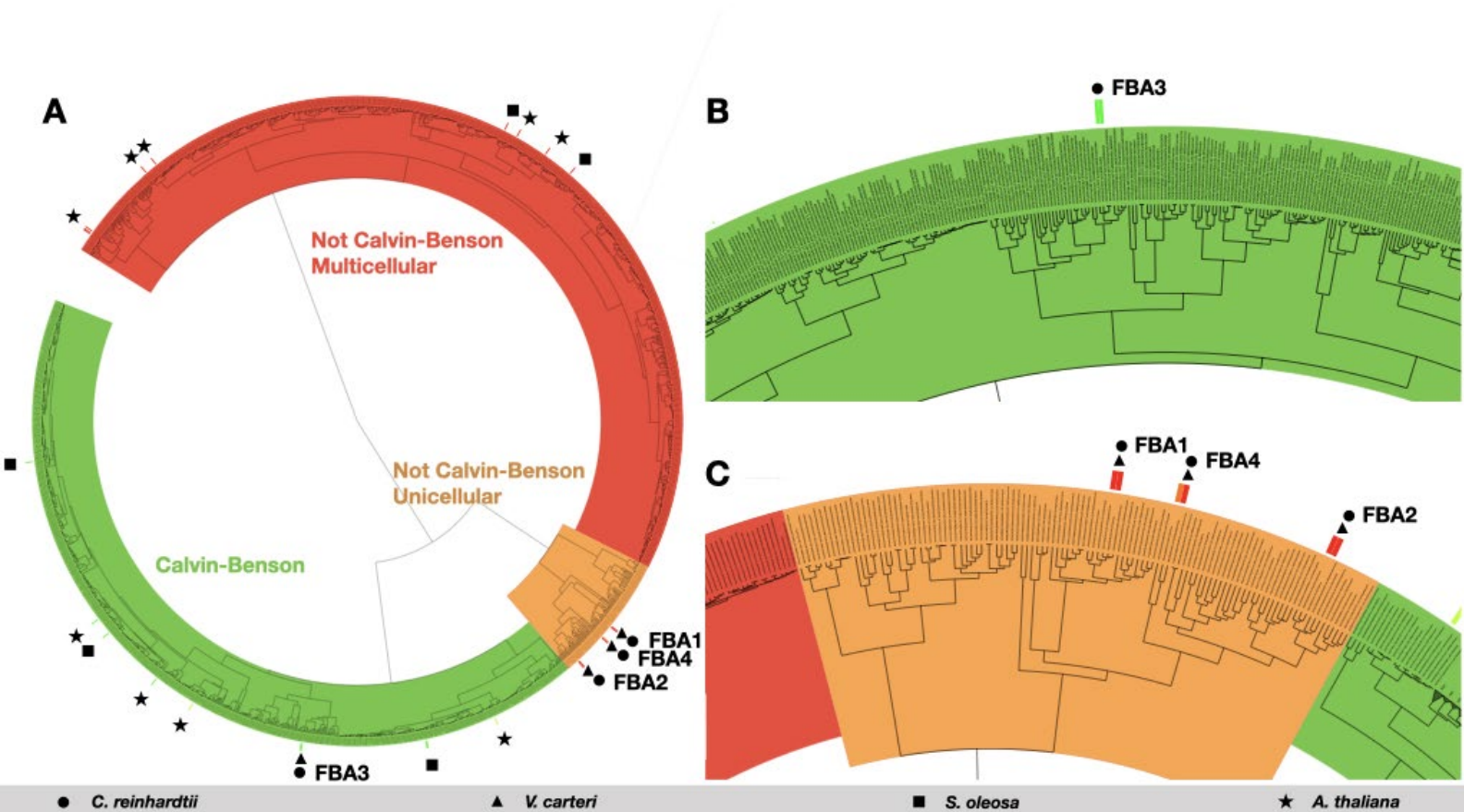




**Figure 4.** **A.** Cartoon representation of the homotetramer of *CrFBA3*. The different subunits are represented in red, green, orange and blue. **B.** Electrostatic surface of *CrFBA3* calculated by PyMOL APBS and represented in a gradient from blue (electropositive) to red (electronegative). *CrFBA3* is in the same orientation as in **A**. **C.** Size-exclusion chromatogram of *Chlamydomonas reinhardtii* cellular extracts separated on **superose** 6 10/300 column. Theoretical molecular masses from column calibration are indicated. **D.** Western blot with anti-*CrFBA3* antibody on chromatography fractions.



**Figure 5.** Redox post-translational modification sites colored in gold and phosphorylation sites colored in red on *CrFBA3*. Fructose-1,6-bisphosphate is modelled in cyan sticks, by alignment with the co-crystal structure of human FBA complexed with FBP (PDB entry 4ALD). **A.** Surface representation of the FBA monomer. **B.** Cartoon representation of the FBA monomer, same orientation as in A.  $\mu$ =nitrosylation,  $\epsilon$ =glutathionylation.



**Figure 6.** FBA functional tree. ProfileView tree showing functional relationship between FBA sequences of photosynthetic organisms: **A.** The tree separates FBA paralogs participating in Calvin-Benson (green subtree) and not participating in Calvin-Benson (orange and red subtrees, for unicellular and multicellular organisms respectively). For *C. reinhardtii*, the paralog FBA3 is localized in the Calvin-Benson subtree, whereas the other three paralogs FBA1, FBA2, FBA4 are grouped together in a different subtree. **B.** Zoom on the Calvin-Benson subtree containing the *C. reinhardtii* FBA3 sequence. **C.** Detail of the Unicellular Non-Calvin-Benson subtree, where the other *C. reinhardtii* FBA paralogs are found. In all panels, circles correspond to *C. reinhardtii* sequences, triangles to *V. carteri*, squares to *S. oleosa*, and stars to *A. thaliana*.