



The structure of *Synechococcus elongatus* enolase reveals key aspects of phosphoenolpyruvate binding

Javier M. González, Ricardo Martí-Arbona, Julian C.-H. Chen and Clifford J. Unkefer

Acta Cryst. (2022). F78, 177–184



IUCr Journals

CRYSTALLOGRAPHY JOURNALS ONLINE

This open-access article is distributed under the terms of the Creative Commons Attribution Licence <https://creativecommons.org/licenses/by/4.0/legalcode>, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are cited.





The structure of *Synechococcus elongatus* enolase reveals key aspects of phosphoenolpyruvate binding

Javier M. González,^{a*} Ricardo Martí-Arbona,^b Julian C.-H. Chen^b and Clifford J. Unkefer^{b*}

^aInstituto de Bionanotecnología del NOA (INBIONATEC–CONICET), Universidad Nacional de Santiago del Estero (UNSE), RN9, Km1125, G4206XCP Santiago del Estero, Argentina, and ^bBioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. *Correspondence e-mail: javierg@conicet.gov.ar, cju@lanl.gov

Received 4 February 2022

Accepted 30 March 2022

Edited by R. Sankaranarayanan, Centre for Cellular and Molecular Biology, Hyderabad, India

Keywords: enolases; *Synechococcus elongatus*; cyanobacteria; phosphoenolpyruvate.

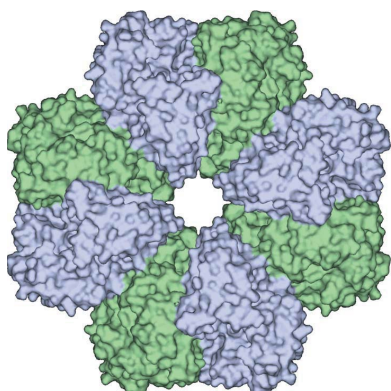
PDB references: enolase from *Synechococcus elongatus*, 4rop; complex with phosphoenolpyruvate, 5j04

Supporting information: this article has supporting information at journals.iucr.org/f

A structure–function characterization of *Synechococcus elongatus* enolase (*SeEN*) is presented, representing the first structural report on a cyanobacterial enolase. X-ray crystal structures of *SeEN* in its apoenzyme form and in complex with phosphoenolpyruvate are reported at 2.05 and 2.30 Å resolution, respectively. *SeEN* displays the typical fold of enolases, with a conformationally flexible loop that closes the active site upon substrate binding, assisted by two metal ions that stabilize the negatively charged groups. The enzyme exhibits a catalytic efficiency of $1.2 \times 10^5 M^{-1} s^{-1}$ for the dehydration of 2-phospho-D-glycerate, which is comparable to the kinetic parameters of related enzymes. These results expand the understanding of the biophysical features of these enzymes, broadening the toolbox for metabolic engineering applications.

1. Introduction

Photosynthetic unicellular microorganisms have been proposed as convenient, cost-effective, carbon-neutral sources for biomass production (Udayan *et al.*, 2022; Alishah Aratboni *et al.*, 2019). Microbial factories based on algae or cyanobacteria offer several advantages over biomass production from higher plants, including the following: (i) microbes may be cultured in pools located in arid uncultivable lands, (ii) microbes exhibit fast growth rates, (iii) unlike lignocellulosic biomass from plants, microbes do not require harsh and costly extraction processes, and (iv) microbes can be genetically engineered with relative ease in order to drive biomass production towards value-added chemicals (Gong *et al.*, 2017; Claassens *et al.*, 2016; Scott *et al.*, 2010). Consequently, a comprehensive understanding of cyanobacterial metabolism, including structural and functional aspects of key enzymes, is required in order to take full advantage of synthetic metabolism for biomass production. Here, we present a structural and functional study of enolase from the cyanobacterium *Synechococcus elongatus* PCC 7942, a workhorse cyanobacterial strain for engineering of photoautotrophic metabolism (Kanno & Atsumi, 2017; Lan *et al.*, 2015). Enolases catalyse the reversible dehydration of 2-phospho-D-glycerate (2PGA) into phosphoenolpyruvate (PEP), an energy-rich phosphoester central to several bioenergetic and biosynthetic metabolic pathways. We show that *S. elongatus* enolase (*SeEN*) displays a typical catalytic efficiency of the order of $10^5 M^{-1} s^{-1}$, as well as a characteristic three-dimensional fold where a flexible loop closes the active site following substrate binding, ensuring the proper environment for catalysis. These results broaden our understanding of the structure and



OPEN ACCESS

Published under a CC BY 4.0 licence

function of metabolic enzymes available for metabolic engineering applications.

2. Materials and methods

2.1. Gene cloning, expression and protein purification

All chemicals and coupling enzymes were obtained from Sigma–Aldrich unless otherwise stated. The oligonucleotide-synthesis and DNA-sequencing reactions were performed by Genewiz Inc. The pET-42a(+) expression vector was purchased from Novagen. *Escherichia coli* Arctic Express (DE3) competent cells were acquired from Stratagene. *SeEN* was synthesized and subcloned into pET-42a(+) by Genewiz Inc. with codon usage optimized for protein expression in *E. coli*. The NdeI and XhoI cloning sites were used to obtain a clone expressing *SeEN* with an 8×His tag at the C-terminus of the protein (Table 1). The expression vector encoding *SeEN* was transformed into *E. coli* Arctic Express (DE3) competent cells and cultured in liquid LB or LB–agar plates supplemented with 50 µg ml⁻¹ kanamycin. A single colony was cultured overnight (~16 h) in 25 ml LB with 50 µg ml⁻¹ kanamycin, and 20 ml was then used to inoculate 2 l of the same medium. Cell cultures were grown at 37°C under shaking to an OD_{600 nm} of ~0.6, after which the temperature was decreased to 16°C and induction was initiated by the addition of isopropyl β-D-1-thiogalactopyranoside to a final concentration of 1.0 mM. The culture was then incubated overnight at 16°C. The bacterial cells were harvested by centrifugation at 7000g for 10 min at 4°C. The pellet was washed and resuspended in 20 mM sodium phosphate, 50 mM NaCl, 5 mM imidazole pH 7.4 (buffer A). 5 µg ml⁻¹ DNase and 0.1 mg ml⁻¹ phenylmethylsulfonyl fluoride protease inhibitor were added to buffer A before cell disruption by sonication. The soluble protein was separated from the cell debris by centrifugation at 12 000g for 30 min at 4°C and then loaded onto two 5 ml HisTrap columns connected in tandem and equilibrated with buffer A. *SeEN* was eluted with a linear gradient of buffer A and 50% buffer B in ten column volumes, where buffer B comprises buffer A plus 500 mM imidazole. Fractions containing *SeEN* were pooled, concentrated, loaded onto a HiLoad 26/60 Superdex 200 preparative-grade gel-filtration column (GE Healthcare) and eluted with 20 mM HEPES, 50 mM NaCl pH 7.2 (buffer C). The purity of the protein during the isolation procedure was checked by SDS–PAGE using Kaleidoscope pre-stained standards (molecular weight 7.6–216 kDa; Bio-Rad).

2.2. Protein molecular-weight estimation

A Tricorn 5/150 column (GE Healthcare) was filled with Superdex 200 prep-grade resin and equilibrated with buffer consisting of 50 mM Tris–HCl, 50 mM NaCl pH 8.0. The column was calibrated with a 12–200 kDa molecular-weight kit for gel-filtration chromatography (Sigma–Aldrich). To determine the elution volume of each protein in the kit and of *SeEN*, the protein standards were dissolved in equilibration buffer to a final concentration of 1 mg ml⁻¹ and *SeEN* was

Table 1
Macromolecule-production information.

Source organism	<i>Synechococcus elongatus</i> PCC7942
DNA source	Synthetic
Expression vector	pET-42a
Expression host	<i>Escherichia coli</i> Arctic Express (DE3)
Primers	Cloning into NdeI and XhoI restriction sites of vector pET-42a
Complete amino-acid sequence of the construct produced	MPDDYGTQIAEITAREILDSRGRPTVEAEV HLEDGVSGLAQVPSGASTGTFFEAHELRRDD DPSRYGGKGVQKAVENVS AIEDALIGLSA LDQEGLDKAMIALDGT PNKKNLGANAILA VSLATAHAAATSLNPL YRYLGGPLANVL PVPMMNVINGGAHADNNVDFQEFMIMPVG APSFKEALRWGAEV FHALAKVLDKGLAT GVGDEGGFAPNLGNSKEALELLLTAEAA GYKPGEQVALAMDVASSEFYKNGLYTCDG VSHEPAGMIGILADLVSYPIVSIEDGLQ EDDWSNWKTLTQQLGSTVQLVGGDDLFVTN PDRLQSGIEQQVGNVAVLIKLNQIGTLTET LRTIDLATRSYRSVISHRSGETEDTTIA DLAVATRAGQIKTGLSRSERIAKYNRL RTEAALGENALYAGAI GLGPKGRLEHHHH HHHH

diluted to 1 mg ml⁻¹. 100 µl of each protein was injected into the column. The size of purified *SeEN* was then estimated through interpolation in the calibration curve.

2.3. Steady-state kinetic assays

The specific activity of *SeEN* for the dehydration of 2-phospho-D-glycerate (2PGA) was followed by coupling the production of phosphoenolpyruvate (PEP) to the oxidation of NADH using pyruvate kinase (PyrKin) and alanine dehydrogenase (ALDH). In this assay, the enolase dehydrates 2PGA to PEP, which is used to phosphorylate ADP to ATP using PyrKin, producing pyruvate. Next, pyruvate is reduced to alanine by ALDH using NADH, the concentration decay of which is followed spectrophotometrically at 340 nm using a Synergy H4 hybrid microplate reader (BioTek Instruments). The standard assay consisted of 100 mM HEPES pH 8.0, 100 mM KCl, 1.0 mM MgCl₂, 1.2 mM ADP, 54 µg ml⁻¹ ALDH, 7.5 µg ml⁻¹ PyrKin, 0.64 mM NADH, 1.0 µg ml⁻¹ *SeEN* and 2PGA in a final volume of 250 µl at 30°C. The steady-state kinetic parameters, *k*_{cat} and *K*_m, were determined by the initial rates method, fitting velocity data to the standard Michaelis–Menten equation.

2.4. Crystallization

Crystals of *SeEN* were grown using the sitting-drop vapour-diffusion method in 96-3 IntelliPlate trays (Art Robbins Instruments; Table 2). Drops consisted of 1.0 µl of a 1:1 mixture of protein sample (25 mg ml⁻¹ in 10 mM HEPES pH 8, 20 mM NaCl) and reservoir solution prepared with an Oryx8 liquid pipettor robot (Douglas Instruments). Initial hits appeared after 3–4 days of incubation at 20°C in a medium consisting of 0.1 M HEPES pH 7.5, 10–15% (w/v) PEG 200, 0.15–0.2 M calcium acetate (mother liquor). Crystals were mounted in nylon loops (Hampton Research) and flash-cooled in liquid nitrogen using 7 M sodium formate pH 7.0 as a cryoprotectant. In order to obtain the complex with PEP,

Table 2
Crystallization.

Method	Sitting drop
Plate type	96-3 IntelliPlate
Temperature (K)	293
Protein concentration (mg ml ⁻¹)	25
Buffer composition of protein solution	10 mM HEPES pH 8, 20 mM NaCl
Composition of reservoir solution	0.1 M HEPES pH 7.5, 10–15% (w/v) PEG 200, 0.15–0.2 M calcium acetate
Volume and ratio of drop	1 µl, 1:1
Volume of reservoir (ml)	0.1

Table 3
Data collection and processing.

Values in parentheses are for the highest resolution shell.

	Apoenzyme	PEP complex
Diffraction source	BL7-1, SSRL	08ID-1, CLS
Wavelength (Å)	1.12709	0.97949
Resolution range (Å)	37.7–2.05 (2.10–2.05)	82.1–2.30 (2.38–2.30)
Total No. of reflections	328759 (25708)	406551 (39756)
No. of unique reflections	30631 (2261)	44630 (4315)
Completeness (%)	98.7 (99.9)	99.8 (100)
Temperature (K)	100	100
Detector	ADSC Quantum 315r CCD	Rayonix MX300 CCD
Space group	<i>I</i> 422	<i>P</i> 4 ₂ 12
<i>a</i> , <i>b</i> , <i>c</i> (Å)	164.3, 164.3, 75.5	164.2, 164.2, 72.7
α , β , γ (°)	90, 90, 90	90, 90, 90
Mosaicity (°)	0.37	0.66
R_{merge}	0.115 (0.631)	0.196 (0.829)
R_{meas}	0.121 (0.664)	0.208 (0.878)
$R_{\text{p.i.m.}}$	0.038 (0.207)	0.068 (0.285)
Multiplicity	10.1 (10.2)	9.1 (9.2)
$\langle I/\sigma(I) \rangle$	9.0 (2.2)	8.7 (3.1)
CC _{1/2}	0.998 (0.928)	0.992 (0.919)
Overall <i>B</i> factor from Wilson plot (Å ²)	34.3	24.3

crystals were soaked in a solution of mother liquor with 50 mM PEP at pH 7.4 for 30 min, cryoprotected in 7 M sodium formate at pH 7.0 and flash-cooled in liquid nitrogen.

2.5. Data collection, processing and structure solution

Diffraction data for apo *SeEN* were collected remotely on beamline BL7-1 at Stanford Synchrotron Radiation Light Source (SSRL) and data for the PEP complex were collected on beamline 08ID-1 at the Canadian Light Source (CLS) (Table 3). Reflections were indexed and integrated with *XDS* (Kabsch, 2010) and *iMosflm* (Battye *et al.*, 2011), while data scaling and truncation was conducted with *AIMLESS* (Evans & Murshudov, 2013). The structure was solved by molecular replacement using *Phaser* (McCoy *et al.*, 2007). The structure of enolase from *E. coli* (PDB entry 1e9i; Kühnel & Luisi, 2001), which shares 62% sequence identity with *SeEN*, was used as a search model, where all noncovalently bound ligands, alternate conformers and solvent molecules were removed and a random shift of 0.3 Å was applied to the remaining protein atom coordinates to minimize model bias. Structure refinement was performed with *REFMAC5* (Murshudov *et al.*, 2011) and the *CCP4* suite of programs (Winn *et al.*, 2011) (Table 4). Initial density maps were used for automated building with *Buccaneer* (Cowtan, 2006) followed

Table 4
Structure solution and refinement.

Values in parentheses are for the highest resolution shell.

	Apoenzyme	PEP complex
PDB code	4rop	5j04
Resolution range (Å)	37.7–2.05 (2.10–2.05)	82.1–2.30 (2.38–2.30)
No. of reflections, working set	30631	42379
No. of reflections, test set	1561 [4.8%]	2189 [4.9%]
Final R_{cryst} (%)	20.9 (29.2)	21.8 (24.4)
Final R_{free} (%)	23.3 (33.3)	24.3 (29.2)
Chains	1	2
No. of non-H atoms		
Protein	3144	6278
Ligand	4 [Ca ²⁺]	35 [Ca ²⁺ , ACT, PEP]
Water	154	191
Total	3302	6504
R.m.s. deviations		
Bonds (Å)	0.015	0.015
Angles (°)	1.92	1.91
Average <i>B</i> factor (Å ²)	38.7	32.3
Ramachandran favoured	409 [96.4%]	830 [97.2%]
Ramachandran allowed	12 [3.32%]	12 [2.61%]
Ramachandran outliers	1 [0.24%]	2 [0.24%]
Real-space correlation coefficients		
PEP	—	0.95 [A], 0.93 [B]
Ca ²⁺ , site Ca1	0.99	0.96 [A and B]
Ca ²⁺ , site Ca2	—	0.94 [A], 0.83 [B]

by manual building with *Coot* (Emsley *et al.*, 2010), using σ_A -weighted $2F_o - F_c$ and $F_o - F_c$ Fourier difference maps. Structure validation was performed with the PDB Validation Service (Westbrook *et al.*, 2003) and the built-in functions in *Coot*. Figures were prepared with *PyMOL* 2.3.0 (Schrödinger) and *CorelDRAW X7* (Corel). Metal-ion sites were checked with *CMM* (Handing *et al.*, 2018). Real-space correlation coefficients were calculated with *MolProbity* (Williams *et al.*, 2018).

2.6. Bioinformatic analysis

The construction of sequence-similarity networks (SSNs) was conducted with the *EFI-EST* server (Gerlt *et al.*, 2015) for Pfam (Mistry *et al.*, 2021) families PF03952 (59 600 sequences) and PF00113 (55 120 sequences), which correspond to the N-terminal capping domains and C-terminal TIM-barrel domains of enolases, respectively. Edges were drawn for node pairs sharing a *BLAST* alignment $-\log(E\text{-score})$ of 100 or better, which corresponds to a sequence identity of approximately 50% or higher. The SSN size was further simplified by sampling the UniRef90 database, grouping together nodes sharing up to 40% sequence identity into representative nodes (rep nodes). Finally, an SSN with 1231 rep nodes and 46 008 edges was obtained for the enolase family. *Cytoscape* 3.8.2 (Shannon *et al.*, 2003) was utilized for network analysis, layout calculation and the evaluation of neighbourhood connectivity distribution (González, 2021). The oligomeric state of selected PDB enolase structures was estimated taking into account the buried molecular surface area (BSA) calculated with the *PISA* server (Krissinel & Henrick, 2007) (Supplementary Table S1). In order to analyse the enolase sequences with known experimentally determined structures that were present in the SSN, a subnetwork was constructed with those rep nodes

containing at least one entry with an associated PDB entry, resulting in a network with 18 rep nodes and 101 edges (Supplementary Fig. S1), comprising 89 PDB entries mapped to 32 UniProt entries (Supplementary Table S2). These 32 enolase sequences were aligned with the online version of *MAFFT* (Katoch *et al.*, 2019) and a 1000-bootstrap neighbour-joining unrooted cladogram was also obtained with the server, considering only 333 conserved positions in an alignment of 556 positions and 406 gap-free sites, using the JTT substitution model. *FigTree* 1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) was used for tree display and *CorelDRAW* (Corel) was used for figure preparation.

3. Results and discussion

SeEN was purified to homogeneity, as shown by a single band on SDS-PAGE immediately above that of carbonic anhydrase (45.7 kDa) from the Kaleidoscope pre-stained standards, in line with the theoretical value of 46.7 kDa for histidine-tagged *SeEN*. The kinetic parameters for dehydration of 2PGA were estimated using a coupled system with PyrKin and ALDH. Briefly, after dehydration of 2PGA to PEP, pyruvate is both produced from PEP and reduced to alanine, leading to NADH consumption, which is followed spectrophotometrically at 340 nm. Using this setup, a Michaelis constant K_m of $61 \pm 4 \mu\text{M}$ and a k_{cat} of $7.4 \pm 0.1 \text{ s}^{-1}$ ($k_{\text{cat}}/K_m = 1.2 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$) were obtained for the 2PGA dehydration reaction of *SeEN*. Such a catalytic efficiency value is comparable with available parameters for enolases (Table 5).

Crystals of recombinant *SeEN* were grown in a medium consisting of 0.1 M HEPES, 0.2 M calcium acetate, 15% (w/v) PEG 200 pH 7.5 (Table 2, Supplementary Fig. S2). Apo *SeEN* crystals belonged to the body-centred tetragonal space group *I422*, with one monomer per asymmetric unit; soaking with PEP changed to the space group to primitive tetragonal, *P42₁2*, with a dimer in the asymmetric unit, in which each monomer was refined with a single molecule of PEP. X-ray diffraction data-collection and refinement statistics are summarized in Tables 3 and 4. In general, the most significant conformational differences were observed in loop 43–51 (PSGASTGTF), which adopts an open (apo) or closed (PEP-bound) form, with an root-mean-square deviation (r.m.s.d.) of 0.221 Å for the fitting of 2519 atoms between the two forms. Additional subtle conformational changes occur upon substrate binding, leading to a shift in the space group from *I422* (one polypeptide chain per asymmetric unit) to *P42₁2* (two polypeptide chains per asymmetric unit), with an r.m.s.d. of 0.074 Å for the fitting of 2567 atoms between chains.

SeEN displays the typical structural features of this enzyme family. The enzyme monomer comprises the customary C-terminal TIM-barrel and N-terminal capping domains, with a metal-binding active site located between them (Fig. 1). An overall fold inspection of *SeEN* readily shows the characteristic unusual $\alpha_2\beta_2(\alpha\beta)_6$ topology, a variant of the symmetric $(\alpha\beta)_8$ -barrel TIM fold, which was witnessed early on for the yeast enzyme, which was the first enolase crystal structure to be determined (Lebioda *et al.*, 1989). Buried surface area

Table 5

Kinetic parameters of *SeEN* as obtained in this work, compared with those of selected enolases.

Enzyme organism	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($\text{M}^{-1} \text{ s}^{-1}$)	Reference
Yeast	78	300	2.6×10^5	Reed <i>et al.</i> (1996)
<i>Klebsiella pneumoniae</i>	52.7	425	1.2×10^5	Pietkiewicz <i>et al.</i> (2009)
Human β -enolase	81.7	199	4.1×10^5	Pietkiewicz <i>et al.</i> (2009)
<i>Staphylococcus aureus</i>	830	365	2.3×10^5	Wu <i>et al.</i> (2015)
<i>S. elongatus</i>	7.4	61	1.2×10^5	This work

(BSA) calculations allow us to estimate that *SeEN* can adopt at least an octameric quaternary structure (Supplementary Table S1). Enolases generally adopt a dimeric or octameric structure, although some exist in solution as a mixture of monomers, dimers and octamers, depending on the buffer composition, as reported for *Helicobacter pylori* enolase (López-López *et al.*, 2018).

Indeed, analysis of *SeEN* with *PISA* indicates that the octamers are stabilized by two kinds of contacts, which we refer to here as ‘main contacts’ and ‘pseudo-contacts’ (Fig. 1). Main contacts display a BSA of 3570 \AA^2 and define the heart-shaped dimers typical of dimeric enolases. On the other hand, pseudo-contacts exhibit a BSA of 3020 \AA^2 and define pseudo-dimers which lead to the formation of octamers in combination with the main contacts. Assuming that the strength of a given contact is proportional to the BSA between interacting monomers, we anticipate that *SeEN* would also exhibit a dynamical equilibrium between monomers, dimers and octamers comparable with the scenario in *H. pylori* enolase, since the BSA for contacts in dimers is only 18% larger than the BSA for pseudo-contacts in pseudo-dimers, *i.e.* pseudo-contacts are marginally weaker than main contacts. In fact, during purification by gel filtration *SeEN* showed a major elution peak corresponding to the monomeric form, suggesting that oligomerization *in vivo* requires specific conditions.

Diffraction crystals of *SeEN* were obtained in the as-purified apo form (PDB entry 4rop) and in complex with PEP (PDB entry 5j04), allowing an analysis of the structural changes that take place upon substrate binding. Considering that both crystal forms grow in a mother liquor containing 150–200 mM calcium acetate, metals were refined as Ca^{2+} ions, even though the available literature indicates that enolases generally contain Mg^{2+} ions as cofactors. Nevertheless, taking into account the chemical similarity between Ca^{2+} and Mg^{2+} ions, it is likely that both metals have comparable structural effects. In the resting state of the enzyme, a ‘catalytic’ Ca^{2+} ion appears to be bound at the bottom of a wide cleft (Ca1 site), coordinated to Asp246, Glu287 and Asp314. Crystal soaking with PEP leads to binding of a second metal ion (Ca2 site), which partly stabilizes the negative charge of the PEP phosphate moiety while anchoring the flexible loop in the closed conformation and is coordinated to the Ala46 carbonyl. The second metal ion has been termed ‘structural’ since it aids in stabilizing the closed loop through its interaction with specific residues (the carbonyl of Ala46 in the loop PSGASTGTF) and the phosphate moiety of PEP (Fig. 2). In addition, the

negatively charged phosphate group in PEP appears to form a salt bridge to the guanidinium side chain of Arg368 and a hydrogen bond to the backbone and side chain of Ser369, further stabilizing the ligand and giving rise to an intricate hydrogen-bond network around it. On comparing chains *A* and *B* of the PEP-complex structure, bound PEP appears to exhibit some conformational freedom, which is accompanied by hydrogen bonds to Arg368 and Ser369.

In order to evaluate the context in which *Se*EN is located within the enolase family of proteins, a sequence-similarity network (SSN) was constructed with all enolase sequences in the UniProt database using the *EFI-EST* server (Gerlt *et al.*, 2015). Currently, the enolase family comprises more than 100 000 sequences (Pfam families PF03952 and PF00113); therefore, SSN analysis is a convenient approach to explore such a large data set. As described in Section 2, an SSN with 1231 rep nodes and 46 008 edges was obtained for the enolase

family. SSN neighbourhood connectivity (NC) reveals a major connected component with two clusters, averaging about 150 and 250 neighbours per node, and a residual group of unclustered sequences (Fig. 3). Considering that a sequence similarity of approximately 50% or higher was assumed to draw edges, a connected component indicates a group of sequences that share this similarity cutoff. The higher the NC value, the higher the shared sequence similarity; therefore, NC values define functionally analogous families. About 73% of all enolases belong to the major cluster, 20% to a less interconnected cluster and 7% remain unclustered. Notably, all enolases functionally and structurally characterized thus far belong to the major cluster, while the cyanobacterial *Se*EN reported here is borderline between clusters 1 and 2. In other words, 27% of the sequence space of all enolases remains experimentally unexplored. The 89 enolase structures currently available in the Protein Data Bank can be mapped to

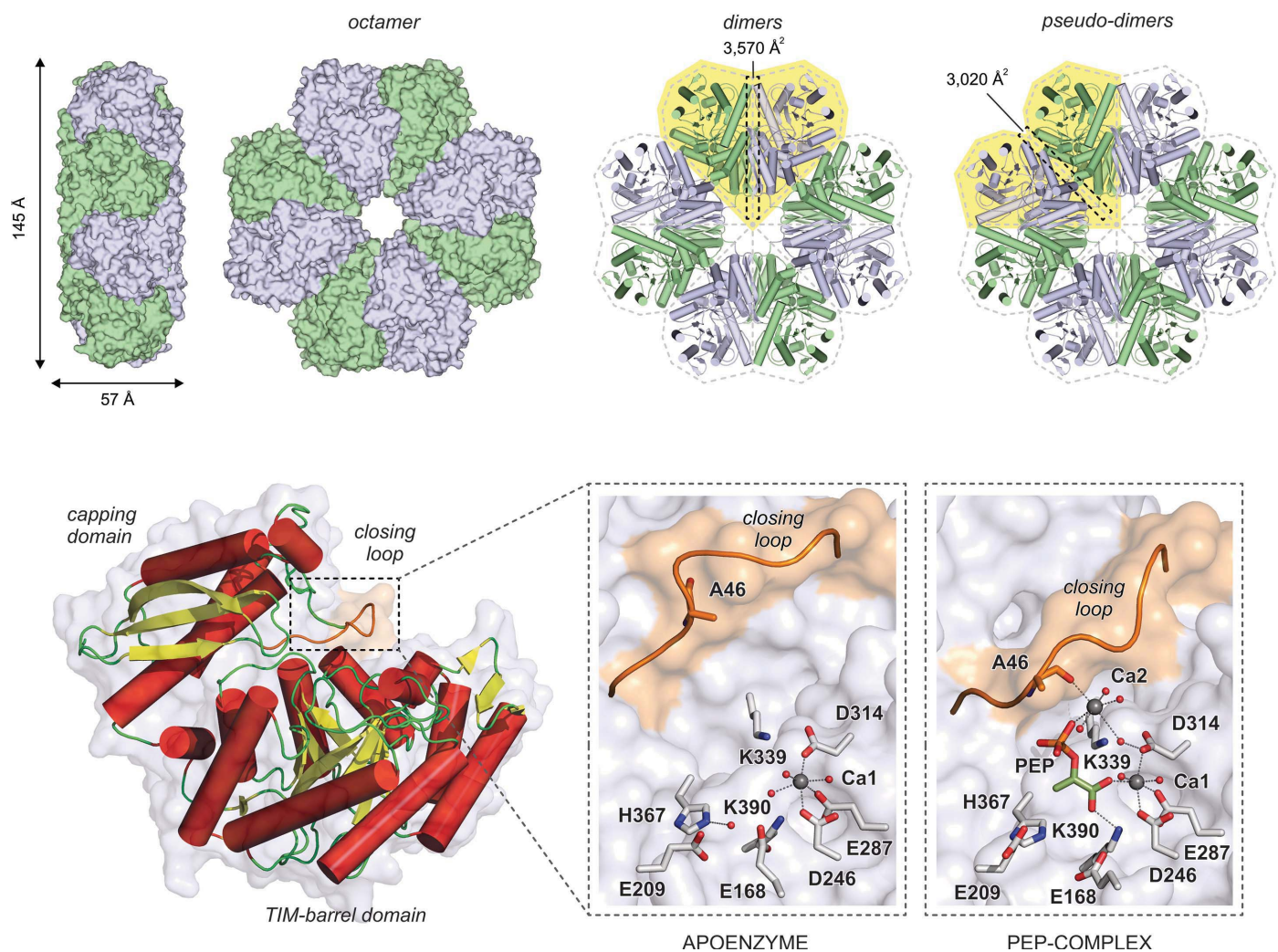


Figure 1
Overall crystal structure of *Se*EN. The enzyme is likely to adopt an octameric quaternary structure comprising four heart-shaped dimers or pseudo-dimers (shaded yellow). In the resting state, *Se*EN shows one Ca²⁺ ion (Ca1 site) coordinated to Asp246, Glu287 and Asp314. The segment 43-PSGASTGTF-51 is conformationally flexible and closes on the active-site cleft upon sequential entrance of the substrate and a second Ca²⁺ ion (Ca2 site). In this way, the backbone carbonyl of Ala46 becomes a ligand of Ca2. Proton extraction from 2PGA is assumed to occur from behind by Lys339 acting as a strong base, whereas hydration of PEP is believed to take place from the opposite side by a water molecule hydrogen-bonded to Glu209 and Glu168.

32 UniProt entries (Supplementary Table S2). A dendrogram based on a sequence alignment of such sequences places *SeEN* in a clade of prokaryotic, typically octameric enolases (Fig. 3). This observation is consistent with BSA analysis of the crystal lattice molecular packing. Even though one or two molecules

per asymmetric unit are found in the structures of *SeEN* in the apo and PEP-bound forms, respectively, both structures indicate that *SeEN* is able to adopt an octameric quaternary structure. Nevertheless, it is worth noting that the enolase sequences might not reflect solely their catalytic performance

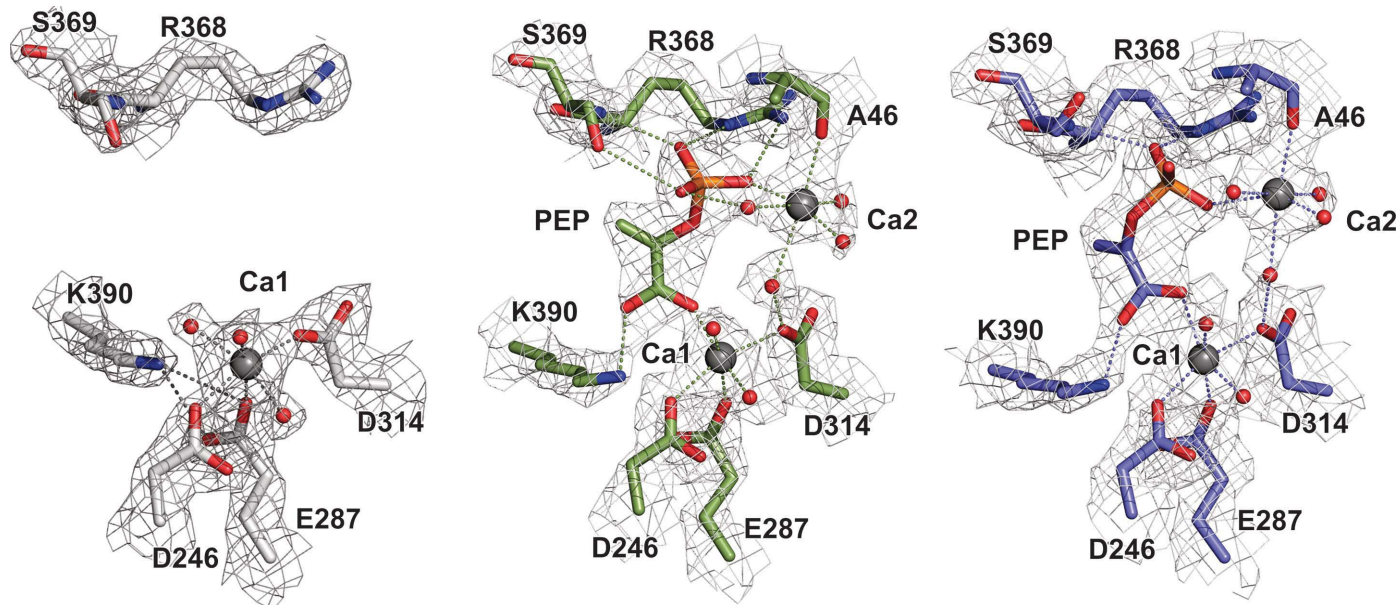


Figure 2 Electron-density distribution around the ligand-binding sites of apo *SeEN* (PDB entry 4rop) (*a*) and PEP-bound *SeEN* (PDB entry 5j04) chains *A* and *B* (*b*, *c*). For simplicity, only those residues that directly interact (dotted lines) with the Ca^{2+} ions and PEP are displayed. Wireframe surfaces indicate $2F_o - F_c$ electron-density maps contoured at 1.5σ . Grey spheres indicate the Ca1 and Ca2 metal-binding sites. Red spheres indicate water molecules.

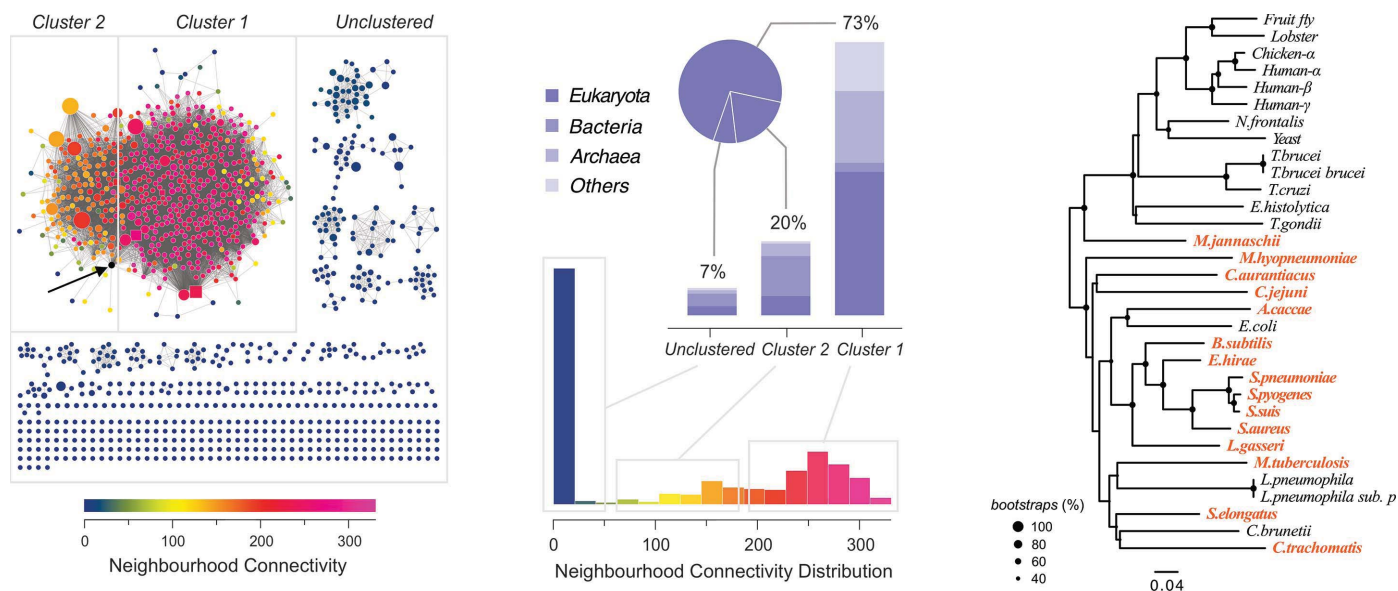


Figure 3 (*a*) Sequence-similarity network (SSN) of all enolase sequences currently available in the UniProt database (1231 rep nodes; 46 008 edges). Edges connecting nodes indicate at least 50% sequence similarity. Node size is proportional to the number of IDs per rep node, whereas square rep nodes indicate the presence of at least one experimentally determined enolase structure (*i.e.* associated PDB entries). The node containing the *SeEN* sequence is indicated by a black arrow. (*b*) The NC distribution indicates that all functionally and structurally characterized enolases belong to the largest cluster, accounting for 73% of the family, of which about 50% is composed of eukaryotic enolases. (*c*) Unrooted phylogenetic tree of 32 structurally characterized enolases (see Supplementary Table S1). Organism names highlighted in orange indicate enolase structures that are predicted to exhibit an octameric quaternary structure.

in terms of dehydration of 2PGA or their ability to adopt octameric structures, but also additional functions unrelated to dehydration of 2PGA. In fact, these enzymes belong to a group of ancient proteins that have evolved so-called moonlighting functions, participating in broad protein–protein interaction networks inside and outside the cell (Paludo *et al.*, 2015).

The enolase reaction mechanism has been extensively studied, particularly in the yeast enzyme (Reed *et al.*, 1996), and shows remarkable mechanistic similarities to muconate-lactonizing enzyme and mandelate racemase (Neidhart *et al.*, 1990; Gerlt *et al.*, 2005, 2012). Their catalytic mechanism is defined by a partial reaction in which an α -proton of a substrate carboxylic acid is removed by a nearby strong base, forming a C1=C2 double bond, with the additional negative charge developing in the carboxylate moiety stabilized by a metal ion such as Mg²⁺ (Fig. 4). In this way, the α -carbon becomes an electron-rich species that reacts with another proton in a second stage. In enolases, a β -elimination takes place, in which a β -OH is lost as a water molecule and an α - β (C2=C3) double bond is formed. These characteristic reaction steps also take place within a conserved protein scaffold. The capping domain, including its flexible loop, defines the volume, shape and polarity of the active site, thereby selecting

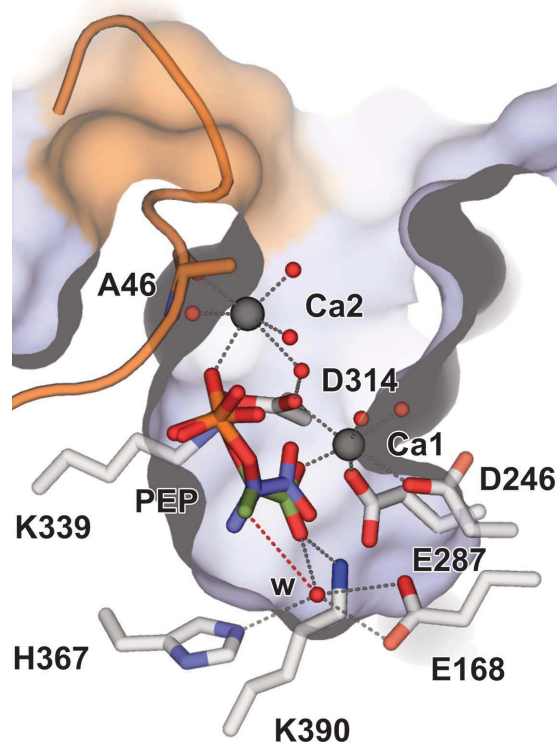


Figure 4
Binding of PEP in the active site of *SeEN*. The PEP molecule is shown in the two conformations found in chains *A* (green) and *B* (blue) of PDB entry 5j04. The water molecule 'w' is 3.5 Å from C3 of the PEP molecule (red dotted line) and is hydrogen-bonded to His367 and Glu168. Compare with Fig. 2. Presumably, during the dehydration of 2PGA or the hydration of PEP (bottom), a water molecule is exchanged from the Glu168 side, whereas proton extraction occurs from the opposite side, with Lys339 acting as a strong base.

the substrate that is presented to the catalytic residues. A reaction mechanism for *SeEN* can be pictured considering the proposed reaction mechanism of yeast enolase. Binding of PEP at the bottom of the cleft defines two sides: an inner solvent-inaccessible side, where proton extraction by Lys339 acting as a strong base would take place, and a solvent-accessible side, from where PEP hydration is believed to occur through activation by a hydrogen-bond network involving Glu168 and His367.

Production of value-added biomass using photosynthetic microbes such as algae and cyanobacteria has been proposed as a cost-effective and environment-friendly alternative to biomass from fossil fuels (Claassens *et al.*, 2016; Quintana *et al.*, 2011; Scott *et al.*, 2010). The results presented here expand our comprehension of cyanobacterial metabolism, enlarging our toolbox of enzymes for metabolic engineering applications.

Acknowledgements

We thank Pawel Grochulski, Shaun Labiuk and the staff of Canadian Light Source (CLS) beamline 08ID-1 for assistance with data collection. The CLS is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), National Research Council, the Canadian Institutes of Health Research (CIHR) and the University of Saskatchewan. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the National Institutes of Health, National Institute of General Medical Sciences (P30GM133894). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH. JMG is a staff member of CONICET (Argentine National Council of Research and Technology). This is article release LA-UR-22-20772.

Funding information

This work has been supported by grant PICT2017-4590 (Argentine Agency for Promotion of Research, Development and Innovation) and Laboratory-Directed Research and Development-Directed Research grants 20130091DR and 20120776PRD4 from Los Alamos National Laboratory.

References

- Alishah Aratboni, H., Rafei, N., Garcia-Granados, R., Alemzadeh, A. & Morones-Ramírez, J. R. (2019). *Microb. Cell Fact.* **18**, 178.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
- Claassens, N. J., Sousa, D. Z., dos Santos, V. A. P. M., de Vos, W. M. & van der Oost, J. (2016). *Nat. Rev. Microbiol.* **14**, 692–706.
- Cowtan, K. (2006). *Acta Cryst.* **D62**, 1002–1011.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.

- Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* **D69**, 1204–1214.
- Gerlt, J. A., Babbitt, P. C., Jacobson, M. P. & Almo, S. C. (2012). *J. Biol. Chem.* **287**, 29–34.
- Gerlt, J. A., Babbitt, P. C. & Rayment, I. (2005). *Arch. Biochem. Biophys.* **433**, 59–70.
- Gerlt, J. A., Bouvier, J. T., Davidson, D. B., Imker, H. J., Sadkhin, B., Slater, D. R. & Whalen, K. L. (2015). *Biochim. Biophys. Acta*, **1854**, 1019–1037.
- Gong, Z., Nielsen, J. & Zhou, Y. J. (2017). *Biotechnol. J.* **12**, 1700014.
- González, J. M. (2021). *Heliyon*, **7**, e05867.
- Handing, K. B., Niedzialkowska, E., Shabalin, I. G., Kuhn, M. L., Zheng, H. & Minor, W. (2018). *Nat. Protoc.* **13**, 1062–1090.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 133–144.
- Kanno, M. & Atsumi, S. (2017). *ACS Synth. Biol.* **6**, 69–75.
- Katoh, K., Rozewicki, J. & Yamada, K. D. (2019). *Brief. Bioinform.* **20**, 1160–1166.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Kühnel, K. & Luisi, B. F. (2001). *J. Mol. Biol.* **313**, 583–592.
- Lan, E. I., Chuang, D. S., Shen, C. R., Lee, A. M., Ro, S. Y. & Liao, J. C. (2015). *Metab. Eng.* **31**, 163–170.
- Lebioda, L., Stec, B. & Brewer, J. M. (1989). *J. Biol. Chem.* **264**, 3685–3693.
- López-López, M. D. J., Rodríguez-Luna, I. C., Lara-Ramírez, E. E., López-Hidalgo, M., Benítez-Cardoza, C. G. & Guo, X. (2018). *Biomed. Res. Int.* **2018**, 9538193.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D. & Bateman, A. (2021). *Nucleic Acids Res.* **49**, D412–D419.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Neidhart, D. J., Kenyon, G. L., Gerlt, J. A. & Petsko, G. A. (1990). *Nature*, **347**, 692–694.
- Paludo, G. P., Lorenzatto, K. R., Bonatto, D. & Ferreira, H. B. (2015). *Comput. Biol. Chem.* **58**, 1–8.
- Pietkiewicz, J., Bednarz-Misa, I., Jermakow, K. & Gamian, A. (2009). *Adv. Clin. Exp. Med.* **18**, 221–233.
- Quintana, N., Van der Kooy, F., Van de Rhee, M. D., Voshol, G. P. & Verpoorte, R. (2011). *Appl. Microbiol. Biotechnol.* **91**, 471–490.
- Reed, G. H., Poyner, R. R., Larsen, T. M., Wedekind, J. E. & Rayment, I. (1996). *Curr. Opin. Struct. Biol.* **6**, 736–743.
- Scott, S. A., Davey, M. P., Dennis, J. S., Horst, I., Howe, C. J., Lea-Smith, D. J. & Smith, A. G. (2010). *Curr. Opin. Biotechnol.* **21**, 277–286.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. & Ideker, T. (2003). *Genome Res.* **13**, 2498–2504.
- Udayan, A., Sirohi, R., Sreekumar, N., Sang, B. I. & Sim, S. J. (2022). *Bioresour. Technol.* **344**, 126406.
- Westbrook, J., Feng, Z., Burkhardt, K. & Berman, H. M. (2003). *Methods Enzymol.* **374**, 370–385.
- Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S. & Richardson, D. C. (2018). *Protein Sci.* **27**, 293–315.
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst.* **D67**, 235–242.
- Wu, Y., Wang, C., Lin, S., Wu, M., Han, L., Tian, C., Zhang, X. & Zang, J. (2015). *Acta Cryst.* **D71**, 2457–2470.



STRUCTURAL BIOLOGY
COMMUNICATIONS

Volume 78 (2022)

Supporting information for article:

The structure of *Synechococcus elongatus* enolase reveals key aspects of phosphoenolpyruvate binding

Javier M. González, Ricardo Martí-Arbona, Julian C.-H. Chen and Clifford J. Unkefer

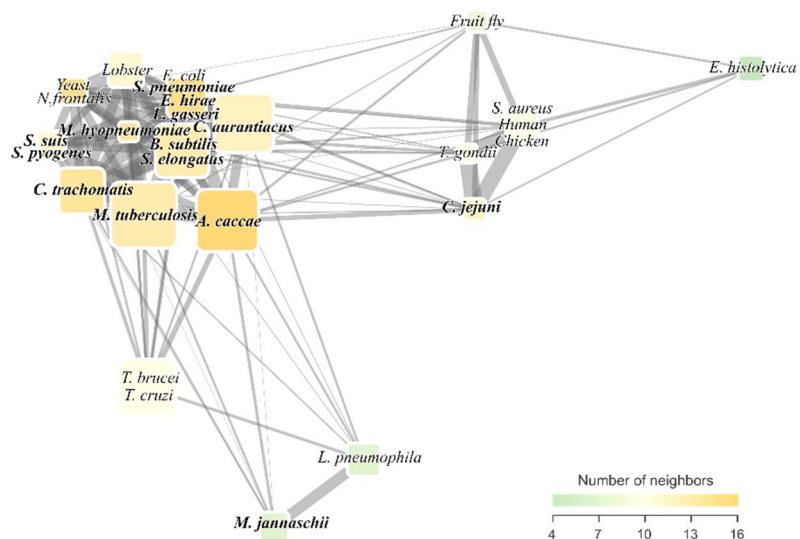


Figure S1 Sequence similarity subnetwork (18 rep nodes by 101 edges) comprising those rep nodes and edges in the family-wide SSN including at least one enolase sequence with known PDB structure. Colouring according to number of neighbours (also known as node degree) and edge thickness proportional to sequence similarity allow a better visualization of clusters in the network. Compare with the dendrogram in Figure 3.

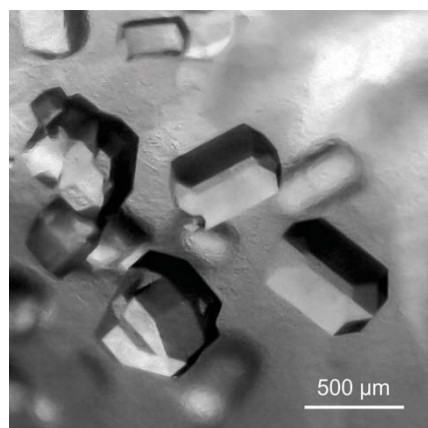


Figure S2 *SeEN* crystals.

Table S1 Buried area (in Å²) of selected enolase crystallographic structures for quaternary structure estimation, as calculated with the PISA server (Krissinel & Henrick, 2007).

Organism	UniProt	PDB	Dimer	Octamer
<i>Trypanosoma cruzi</i>	Q4DZ98	4G7F	0	0
<i>Coxiella burnetii</i>	Q83B44	3TQP	3,760	0
<i>Toxoplasma gondii</i>	B9PH47	3OTR	3,770	0
<i>Escherichia coli</i> (strain K12)	P0A6P9	1E9I	4,060	0
<i>Legionella pneumophila</i>	A0A2S6F4U1	6NBM	4,260	0
<i>Homarus gammarus</i> (European lobster)	P09104	1PDY	4,530	0
Human γ -enolase	Q97QS2	1TE6	4,820	0
<i>Saccharomyces cerevisiae</i> (Baker's yeast)	P00924	1EBG	4,920	0
Human α -enolase	Q38BV6	2PSN	4,960	0
<i>Trypanosoma brucei brucei</i>	Q9NDH8	2PTZ	5,670	0
<i>Trypanosoma brucei brucei</i>	P56252	2PU1	5,780	0
<i>Entamoeba histolytica</i>	P51555	3QTP	5,920	0
<i>Legionella pneumophila sub. pneumophila</i>	Q5ZTX1	6NB2	6,130	0
<i>Drosophila melanogaster</i> (Fruit fly)	P15007	5WRO	6,920	0
Human β -enolase	P13929	2XSX	8,670	0
<i>Streptococcus suis</i>	K7N5M7	4EWJ	3,350	22,870
<i>Lactobacillus gasseri</i>	Q042F4	4MKS	3,410	23,150
<i>Chloroflexus aurantiacus</i>	A9WCM4	4YWS	3,290	23,630
<i>Synechococcus elongatus</i>	Q5N3P4	4ROP	3,570	23,920
<i>Streptococcus pyogenes</i>	Q5XD01	3ZLH	3,650	24,770
<i>Streptococcus pneumoniae</i>	Q60173	1W6T	4,190	26,920
<i>Anaerostipes caccae</i>	B0MAG5	3UJ2	4,810	26,960
<i>Bacillus subtilis</i> (strain 168)	P37869	4A3R	3,480	27,120
<i>Campylobacter jejuni</i>	P42448	3QN3	4,750	29,170
<i>Staphylococcus aureus</i>	O69174	5BOE	5,270	31,760
<i>Enterococcus hirae</i>	Q8GR70	1IYX	4,780	32,180
<i>Mycobacterium tuberculosis</i>	P9WNL1	6L7D	5,380	32,670
<i>Mycoplasma hyopneumoniae</i>	A0A223MA21	6J36	5,840	33,530
<i>Methanocaldococcus jannaschii</i>	P06733	2PA6	4,310	34,170
<i>Chlamydia trachomatis</i>	B0BA40	6O4N	5,670	38,850

Table S2 PDB and associated UniProt entries present in the enolase SSN (Figure 3).

Organism	PDB	Entry
<i>Anaerostipes caccae</i>	3UJ2	B0MAG5
<i>Bacillus subtilis</i> (strain 168)	4A3R	P37869
<i>Campylobacter jejuni</i>	3QN3	P42448
<i>Chlamydia trachomatis</i>	6O4N	B0BA40
<i>Chloroflexus aurantiacus</i>	4YWS, 4Z17, 4Z1Y	A9WCM4
<i>Coxiella burnetii</i>	3TQP	Q83B44
<i>Drosophila melanogaster</i> (Fruit fly)	5WRO	P15007
<i>Entamoeba histolytica</i>	3QTP	P51555
<i>Enterococcus hirae</i>	1IYX	Q8GR70
<i>Escherichia coli</i> (strain K12)	1E9I, 2FYM, 3H8A, 5OHG, 6BFY, 6BFZ, 6D3Q, 6NPF	P0A6P9
<i>Gallus gallus</i> (Chicken)	2GUA	P51913
<i>Homarus gammarus</i> (European lobster)	1PDY, 1PDZ	P56252
Human α -enolase	2PSN, 3B97, 5JLZ, 5LAX, 5NI9, 5NIG, 5OCK	P06733
Human β -enolase	2XSX	P13929
Human γ -enolase	1TE6, 2AKM, 2AKZ, 3UCC, 3UCD, 3UJE, 3UJF, 3UJR, 3UJS, 4ZA0, 4ZCW, 5EU9, 5IDZ, 5TD9, 5TIJ	P09104
<i>Lactobacillus gasseri</i>	4MKS	Q042F4
<i>Legionella pneumophila</i>	6NBM	A0A2S6F4U1
<i>Legionella pneumophila sub. pneumophila</i>	6NB2	Q5ZTX1
<i>Methanocaldococcus jannaschii</i>	2PA6	Q60173
<i>Mycobacterium tuberculosis</i>	6L7D	P9WNL1
<i>Mycoplasma hyopneumoniae</i>	6J36	A0A223MA21
<i>Neocallimastix frontalis</i> (Rumen fungus)	2GE6	P42894
<i>Saccharomyces cerevisiae</i> (Baker's yeast)	1EBG, 1EBH, 1ELS, 1L8P, 1NEL, 1ONE, 1P43, 1P48, 2AL1, 2AL2, 2ONE, 2XGZ, 2XH0, 2XH2, 2XH4, 2XH7, 3ENL, 4ENL, 5ENL, 6ENL, 7ENL	P00924
<i>Staphylococcus aureus</i>	5BOE, 5BOF	O69174
<i>Streptococcus pneumoniae</i>	1W6T	Q97QS2
<i>Streptococcus pyogenes</i>	3ZLF, 3ZLG, 3ZLH	Q5XD01
<i>Streptococcus suis</i>	4EWJ	K7N5M7
<i>Synechococcus elongatus</i>	4ROP, 5J04	Q5N3P4
<i>Toxoplasma gondii</i>	3OTR	B9PH47
<i>Trypanosoma brucei brucei</i>	1OEP, 2PU1	Q9NDH8
<i>Trypanosoma brucei brucei</i>	2PTW, 2PTX, 2PTY, 2PTZ, 2PU0	Q38BV6
<i>Trypanosoma cruzi</i>	4G7F	Q4DZ98