STRUCTURAL BIOLOGY

ISSN: 2059-7983
journals.iucr.org/d

The crystal structure of the malic enzyme from *Candidatus* Phytoplasma reveals the minimal structural determinants for a malic enzyme

C. E. Alvarez, F. Trajtenberg, N. Larrieux, M. Saigo, A. Golic, C. S. Andreo, S. A. Hogenhout, M. A. Mussi, M. F. Drincovich and A. Buschiazzo

Acta Cryst. (2018). D**74**, 332–340

STRUCTURAL BIOLOGY
ISSN 2059-7983

The crystal structure of the malic enzyme from *Candidatus* Phytoplasma reveals the minimal structural determinants for a malic enzyme

C. E. Alvarez,[a]‡ F. Trajtenberg,[b] N. Larrieux,[b] M. Saigo,[a] A. Golic,[a] C. S. Andreo,[a] S. A. Hogenhout,[c] M. A. Mussi,[a] M. F. Drincovich[a]* and A. Buschiazzo[b]*

[a]CEFOBI, Suipacha 531, Rosario, S2000LRJ Santa Fe, Argentina, [b]Institute Pasteur, Mataojo 2020, Montevideo, Uruguay, and [c]Department of Cell and Developmental Biology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, England. *Correspondence e-mail: drincovich@cefobi-conicet.gov.ar, alebus@pasteur.edu.uy

Received 10 November 2017
Accepted 15 February 2018

Edited by R. C. Garratt, University of São Paulo, Brazil

‡ These authors contributed equally to this work.

Keywords: plant pathogens; reductive metabolism; crystallography; homodimer; *Candidatus* Phytoplasma; malic enzyme; phytoplasma; aster yellows; AYWB.

PDB reference: malic enzyme from *Candidatus* Phytoplasma in complex with NAD and Mg^{2+}, 5cee

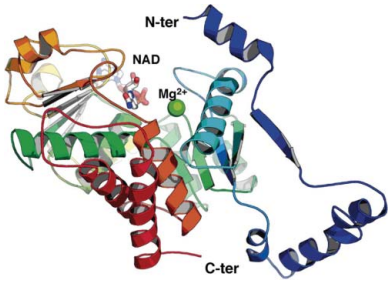
Supporting information: this article has supporting information at journals.iucr.org/d

Phytoplasmas are wall-less phytopathogenic bacteria that produce devastating effects in a wide variety of plants. Reductive evolution has shaped their genome, with the loss of many genes, limiting their metabolic capacities. Owing to the high concentration of C_4 compounds in plants, and the presence of malic enzyme (ME) in all phytoplasma genomes so far sequenced, the oxidative decarboxylation of L-malate might represent an adaptation to generate energy. Aster yellows witches'-broom (*Candidatus* Phytoplasma) ME (AYWB-ME) is one of the smallest of all characterized MEs, yet retains full enzymatic activity. Here, the crystal structure of AYWB-ME is reported, revealing a unique fold that differs from those of 'canonical' MEs. AYWB-ME is organized as a dimeric species formed by intertwining of the N-terminal domains of the protomers. As a consequence of such structural differences, key catalytic residues such as Tyr36 are positioned in the active site of each protomer but are provided by the other protomer of the dimer. A Tyr36Ala mutation abolishes the catalytic activity, indicating the key importance of this residue in the catalytic process but not in the dimeric assembly. Phylogenetic analyses suggest that larger MEs (large-subunit or chimeric MEs) might have evolved from this type of smaller scaffold by gaining small sequence cassettes or an entire functional domain. The *Candidatus* Phytoplasma AYWB-ME structure showcases a novel minimal structure design comprising a fully functional active site, making this enzyme an attractive starting point for rational genetic design.

1. Introduction

Malic enzyme (ME) is present in almost all taxa, including bacteria, yeast, fungi, plants and animals (Hsu, 1982; Edwards & Andreo, 1992; Drincovich *et al.*, 2001; Chang & Tong, 2003; Saigo *et al.*, 2013). ME catalyses the oxidative decarboxylation of L-malate, yielding pyruvate, CO_2 and NAD(P)H, which are all important intermediary metabolites involved in a number of biochemical processes. In each organism, different ME isoforms are implicated in diverse and relevant metabolic pathways. For instance, ME is linked to an increase of the photosynthetic yield in maize and sorghum, amongst other plants with economic impact (Saigo *et al.*, 2013; Alvarez *et al.*, 2013); the regulation of stomata opening in tobacco plants (Laporte *et al.*, 2002); plant defence responses (Saigo *et al.*, 2004); oxidative stress and lignin biosynthesis (Casati *et al.*, 1999; Liu *et al.*, 2007); glutamine metabolism in cancerous animal cells (Yang, Lanks *et al.*, 2002); and the metabolism of pathogens such as phytoplasma (Saigo *et al.*, 2014).

Related to their diverse biological functions, an extremely diverse array of ME variants have been identified over the



years, displaying variable physicochemical and enzymological properties: (i) dissimilar sizes, from 42 to 83 kDa per monomer (Bologna *et al.*, 2007); (ii) different quaternary structures, from homodimers, homotetramers and homohexamers to homooctamers (Chang & Tong, 2003; Detarsio *et al.*, 2008; Tronconi *et al.*, 2010) and even heterodimers (Tronconi *et al.*, 2008); (iii) distinct specificities for the nucleotide cofactor, including NAD- or NADP-specific forms, as well as others able to work with both cofactors, called NAD(P)-MEs (Hsieh *et al.*, 2011); (iv) fusion to unrelated additional protein domains, such as phosphotransacetylase, forming chimeric structures (Bologna *et al.*, 2007); (v) differential capabilities to catalyse the reductive carboxylation of pyruvate (Gerrard Wheeler *et al.*, 2008); and (vi) a differential response to allosteric effectors (Yang, Lanks *et al.*, 2002; Gerrard Wheeler *et al.*, 2009; Arias *et al.*, 2013; Saigo *et al.*, 2013). It has been suggested that the wide diversity of ME properties facilitated the appearance of the C_4 ME isoform, which evolved to enable particularly efficient photosynthesis (Alvarez *et al.*, 2013).

MEs all contain an NAD(P)-binding Rossmann-fold domain and belong to the amino-acid dehydrogenase family. Three molecular structures of ME have been reported to date: those of human mitochondrial NAD(P)-ME (Human-ME; Xu *et al.*, 1999), pigeon liver cytosolic NADP-ME (*Columba livia*; Yang, Zhang *et al.*, 2002) and *Ascaris suum* mitochondrial NAD-ME (Coleman *et al.*, 2002). These MEs have been well characterized and display different substrate specificities, kinetic properties and allosteric regulation. However, their three-dimensional structures are quite similar, even adopting the same tetrameric organization, which can more precisely be described as a dimer of dimers. Each monomer comprises four domains (A–D), which behave as rigid bodies when ligand binding triggers conformational rearrangements (Chang & Tong, 2003; Daily & Gray, 2009). The common topology of these three MEs clusters them into a distinct class of oxidative decarboxylases (Chang & Tong, 2003). A much wider array of MEs exist, and although not yet described in published reports, five crystal structures of four additional MEs have been deposited in the PDB from *Thermotoga maritima* (PDB entries 1vl6 and 2hae; Joint Center for Structural Genomics, unpublished work; New York SGX Research Center for Structural Genomics, unpublished work), *Pyrococcus horikoshii* (PDB entry 2dvm; RIKEN Structural Genomics/Proteomics Initiative, unpublished work), *Streptococcus pyogenes* (PDB entry 2a9f; New York SGX Research Center for Structural Genomics, unpublished work) and *Entamoeba histolytica* (PDB entry 3nv9; A. Chakrabarty, D. Dutta, A. K. Das & S. K. Ghosh, unpublished work), representing a huge group of structurally diverse MEs that are still unexplored.

Phytoplasmas are wall-less phytopathogenic bacteria that produce devastating effects on a wide variety of plants (Bertaccini, 2007; Hogenhout *et al.*, 2008; Strauss, 2009). Transmitted by insect vectors, phytoplasmas parasitize the phloem of infected plant hosts. As a result of reductive evolution, phytoplasma genomes show the loss of many genes encoding proteins in metabolic routes (Kube *et al.*, 2012). It is thus not clear which compounds they use as carbon/energy sources. Owing to the high concentration of C_4 compounds in plants and the presence of a malate transporter and an ME in all sequenced phytoplasma genomes, it is highly possible that C_4 acids are used by these bacteria during infection, in which case ME is hypothesized to fulfil a critical function. Aster yellows witches'-broom (*Candidatus* Phytoplasma) ME (AYWB-ME) is an interesting ME with a significantly smaller molecular mass (42 kDa compared with ∼62 kDa for the well characterized examples), an apparently dimeric quaternary structure and intriguingly distinct functional properties, such as activation by ADP and glutamine, and inhibition by ATP. The molecular mechanisms underlying the pathogenicity of Phytoplasma remain elusive, and it has been suggested that depletion of host resources by the action of AYWB-ME plays a crucial role (Kube *et al.*, 2012; Saigo *et al.*, 2014). AYWB-ME is an excellent model to define the minimal structural determinants of these enzymes and to resolve the question of how a 30% smaller protein scaffold retains full ME activity. With this aim, we here report the crystal structure of AYWB-ME at 2.5 Å resolution in complex with NAD and Mg^{2+}, confirming the dimeric architecture of this novel class of MEs. Our data illustrate a much broader diversity in ME domain architectures and quaternary-structure organizations, revealing a novel, highly symmetric dimeric organization which contains an extensive inter-monomer interface. As a consequence of such an interleaved structure, the two active sites (one in each monomer) are built by a number of residues from one monomer but include a key tyrosine (Tyr36) contributed by the neighbouring monomer, a tyrosine that we show to be critical for ME activity. This minimal scaffold indeed retains all of the necessary residues to bind the substrates and cofactors, allowing AYWB-ME to adopt a novel ME conformation which defines a new class of MEs.

2. Materials and methods

2.1. Protein expression and purification

Site-directed mutagenesis was performed by MegaPrimer PCR amplification using the primers AYWBF, 5′-GCTA GCATGAACATCAAAGAAAAAGCAT-3′, and AY_WB_Y36A_rev, 5′-GAGTGGCAACTAATGATAAATCATC-3′. The MegaPrimer product was used with the primer AYWBR, 5′-CTCGAGTCATTTTCTTACTACTCCAGT-3′, to obtain the mutated sequence (AYWB-ME_Tyr36Ala). The amplified products were cloned in pGEM T-Easy (Promega), and the NheI–XhoI fragment was further subcloned into cognate sites of pET-28a (Novagen; Table 1). The expression construct was designed to code for an N-terminal hexahistidine (6×His) tag for purification by immobilized metal-affinity chromatography. The recombinant AYWB-ME_Tyr36Ala protein was expressed in *Escherichia coli* BL21 (DE3) cells previously transformed with pET-AYWB-ME_Tyr36Ala, and the purification protocol was similar to that used for the wild-type protein (Saigo *et al.*, 2014).

Recombinant protein expression and purification of AYWB-ME were performed as described previously (Saigo *et*

Table 1
Macromolecule-production information.

Source organism	Aster yellows witches'-broom phytoplasma (strain AYWB)
DNA source	Aster yellows witches'-broom phytoplasma (strain AYWB)
Forward primer (AYWBF)	5′-GCTAGCATGAACATCAAAGAAAAAGCAT-3′
Reverse primer (AYWBR)	5′-CTCGAGTCATTTTCTTACTACTCCAGT-3′
Cloning vector	pGEM T-Easy (Promega)
Expression vector	pET-28a (Novagen)
Expression host	*E. coli* BL21 (DE3)
Complete amino-acid sequence of the construct produced	MGSSHHHHHHSSGLVPRGSHMASMNIKEKALEMHEKNKGKVGVVSKVKVQNLDDLSLVYTPGVAEPCLKIKENPSDVYRYTMKGNMVGVITNGTAVLGLGNIGPKASLPVMEGKAILFKELAGIDSFPICIDSTDSQEIVNIVSKISTVFGAINLEDIKSPQCIEIEDALKAKLDIPVFHDDQHGTAIVVAAGILNALKVVKKSIEDVQVVINGAGSAGMAIAKMLLLKVNNVVLVDKTGTLYKGVANLNEPQKKLVEVTNKYQEKGTLKEVLKGKDIFIGVSAPGIVTAEMVATMAKDAIVFALANPVPEIMPDEAKKGGARIVATGRSDFPNQVNNCLAFPGVFRGTLDAKATQITEEMKKAATYALKNIIKEQDLNENNILPTSFNKEVVKQIALAVCKVAKETGVVRK

al., 2014). Briefly, AYWB-ME was expressed in *E. coli* BL21 (DE3) cells and purified by Ni–NTA affinity chromatography followed by size-exclusion chromatography.

2.2. Protein crystallization

The initial identification of crystallization conditions for AYWB-ME was carried out by the sitting-drop vapour-diffusion method using a robotic workstation (Honeybee963, Isogen Life Science). Sitting drops were set up using 400 nl of a 1:1 mixture of protein and reservoir solution, equilibrating against 150 µl reservoir solution in Greiner plates at both 4 and 20°C. The solutions screened were from commercial kits: Crystal Screen and Crystal Screen 2 (Hampton Research) and The PEGs Suite (Qiagen). The reservoir solution that gave the best crystal hit consisted of 0.2 *M* $MgCl_2$, 0.1 *M* Tris pH 8.5, 30%(*w*/*v*) PEG 4000. The crystal-growth habit was similar at both crystallization temperatures, so we continued working only at 20°C. Initial AYWB-ME crystals were manually optimized by varying the precipitant and protein concentrations using VDX plates (Hampton Research) with a hanging-drop setup. The best crystallization condition was ultimately obtained using reservoir solution consisting of 30%(*w*/*v*) PEG 4000, 0.1 *M* Tris–HCl pH 8.5, 0.2 *M* $MgCl_2$ at 20°C, with protein at 8.8 mg ml^{-1} in 100 m*M* Tris–HCl pH 8.0, 5 m*M* NAD, 5 m*M* ADP, 40 m*M* pyruvate, 10 m*M* $MgCl_2$, 10%(*v*/*v*) glycerol (Table 2). Cryoprotection was achieved by slowly adding cryoprotection solution [20%(*v*/*v*) glycerol, 30%(*w*/*v*) PEG 4000, 0.1 *M* Tris–HCl pH 8.5, 0.2 *M* $MgCl_2$, 5 m*M* NAD, 5 m*M* ADP, 40 m*M* pyruvate] to the drop in small volumes (∼5% of the drop volume stepwise until >15% glycerol was reached); the crystals were then rapidly soaked in 100% cryoprotection solution and flash-cooled in liquid N_2 for storage until data collection.

Table 2
Crystallization.

Method	Vapour diffusion, hanging drop
Plate type	24-well VDX plate (Hampton Research)
Temperature (K)	277
Protein concentration (mg ml^{-1})	8.8
Buffer composition of protein solution	100 m*M* Tris pH 8.0, 5 m*M* NAD, 5 m*M* ADP, 40 m*M* pyruvate, 10 m*M* $MgCl_2$, 10%(*v*/*v*) glycerol
Composition of reservoir solution	0.2 *M* $MgCl_2$, 0.1 *M* Tris pH 8.5, 30% PEG 4000
Volume and ratio of drop	2 µl, 1:1 mixture of protein and reservoir solutions
Volume of reservoir (ml)	1

2.3. Diffraction data-processing, structure-determination and refinement procedures

X-ray diffraction data collection was performed using a MicroMax-007 HF generator (Rigaku) with a rotating copper anode, VariMax HF optics (Rigaku) and a MAR345 image-plate detector (MAR Research). The crystals used for X-ray diffraction studies grew in the trigonal space group $P3_121$, with unit-cell parameters $a = b = 48.2$, $c = 291.1$ Å, and were hemihedrally twinned (twinning law $h = -h$, $k = -k$, $l = l$, subsequently used throughout the refinement procedure). Twinning was detected during data processing by the standard indices that are now routinely used, as implemented in *TRUNCATE* and *POINTLESS* (Evans, 2011). Briefly, statistical analyses of the distribution of diffraction intensities in all resolution shells strongly suggested merohedral twinning, which was especially clear for the second moment of normalized intensities (close to 1.5) and the *L*-test (Padilla & Yeates, 2003). These suspicions were later confirmed by a significant reduction of the crystallographic *R* factors and much better electron-density maps during refinement when the appropriate twinning law was applied (see below).

X-ray diffraction data sets were processed using *XDS* (Kabsch, 2010) and *AIMLESS* (Evans & Murshudov, 2013); the crystals diffracted to 2.5 Å resolution (Tables 3 and 4). Initial phases were obtained by molecular replacement with *Phaser* (McCoy *et al.*, 2007) using the ME from *P. horikoshii* (PDB entry 2dvm) as a search probe. *Coot* (Emsley & Cowtan, 2004) was used for manual rebuilding, and iterative refinement was carried out with *phenix.refine* (Adams *et al.*, 2010), using a least-squares function taking the hemihedral twin law into account (with a refined twinned fraction of 0.41). One monomer of AYWB-ME was found in the asymmetric unit; the biologically relevant dimer is obtained by applying the twofold crystallographic symmetry operation.

2.4. Activity assays

AYWB-ME activity was measured using the same general protocol as described previously (Saigo *et al.*, 2014). The oxidative decarboxylation of L-malate was monitored using spectrophotometric assays by adding 1.5 m*M* NAD and 30 m*M* malate in a final volume of 0.5 ml, alongside a standard reaction mixture consisting of 50 m*M* Tris–HCl pH 8.2, 5 m*M* $MgCl_2$. The reaction was started by the addition of L-malate

Table 3
Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	Rigaku MicroMax-007 HF
Wavelength (Å)	1.5418
Temperature (K)	100
Detector	MAR345 image plate
Crystal-to-detector distance (mm)	300
Rotation range per image (°)	0.5
Total rotation range (°)	178.5
Exposure time per image (s)	605
Space group	$P3_121$
a, b, c (Å)	48.24, 48.24, 291.07
α, β, γ (°)	90, 90, 120
Mosaicity (°)	0.102
Resolution range (Å)	29.470–2.498 (2.630–2.498)
Total No. of reflections	14357
No. of unique reflections	14448
Completeness (%)	98.4 (90.7)
Multiplicity	6.4 (4.1)
$\langle I/\sigma(I)\rangle$	20.2 (2.4)
$R_{\rm r.i.m.}$†	0.072 (0.806)
Overall B factor from Wilson plot (Å^2)	36.3

† Estimated $R_{\rm r.i.m.} = R_{\rm merge}[N/(N-1)]^{1/2}$, where N is the data multiplicity.

Table 4
Structure solution and refinement.

Values in parentheses are for the outer shell.

Resolution range (Å)	24.256–2.498 (2.5502–2.5022)
Completeness (%)	97.4
No. of reflections, working set	14357 (1114)
No. of reflections, test set	2511 (148)
Final $R_{\rm cryst}$	0.146 (0.3016)
Final $R_{\rm free}$	0.195 (0.3894)
No. of non-H atoms	
Protein	2876
Ion	1
Ligand	44
Water	13
Total	2934
R.m.s. deviations	
Bonds (Å)	0.008
Angles (°)	1.151
Average B factors (Å^2)	
Protein	35.1
Ion	32.6
Ligand	41.8
Water	28.9
Ramachandran plot	
Most favoured (%)	90
Allowed (%)	9

and the absorbance at 340 nm at 30°C was instantly recorded. One unit of the enzyme was defined as the amount of enzyme that catalyzed the production of 1 µmol NAD(P)H per minute. An absorption coefficient of 6.22 mM^{-1} cm^{-1} for NAD(P)H was used in the calculations.

2.5. Size-exclusion chromatography

The apparent molecular weights of the recombinant native AYWB-ME and mutant AYWB-ME_Tyr36Ala were evaluated by size-exclusion chromatography on a fast liquid-chromatography system using a Sephacryl 16/60 200 HR column (GE Healthcare Life Sciences). The column was equilibrated with 50 mM Tris–HCl pH 8.2 and was calibrated using molecular-weight standards. The sample and standards were applied separately in a final volume of 800 µl at a constant flow rate of 0.5 ml min^{-1}.

2.6. Far circular dichroism (CD) spectra

CD spectra were obtained using a Jasco J-810 spectropolarimeter using 0.5 and 0.2 cm path-length cells and averaging eight repetitive scans between 250 and 200 nm at 30°C. Approximately 50 µg of each protein in phosphate buffer (20 mM Na_2HPO_4 pH 8.0, 5 mM $MgCl_2$) was used for each assay.

2.7. Sequence and structural analysis

Protein sequences were retrieved from nonredundant databases at NCBI and three-dimensional structures were retrieved from the wwPDB, using ME as the query. The structures were analysed using *PyMOL* (v.1.8.6; Schrödinger) and *PISA* (Krissinel & Henrick, 2007).

2.8. Phylogenetic relationship studies

Phylogenetic relationships were inferred from amino-acid sequence alignments using *MEGA*7 (Felsenstein, 1985; Kumar *et al.*, 2016). The evolutionary history was inferred using the maximum-likelihood method based on the Le–Gascuel model (Le & Gascuel, 2008). The tree with the highest log likelihood is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying the neighbour-joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with the superior log-likelihood value. A discrete gamma distribution was used to model evolutionary-rate differences among sites. The rate-variation model allowed some sites to be evolutionarily invariable. The tree is drawn to scale, with branch lengths corresponding to the number of substitutions per site. The analysis involved 37 amino-acid sequences, as detailed in Supplementary Fig. S1. All positions with less than 80% site coverage were eliminated, *i.e.* less than 20% alignment gaps, missing data and ambiguous bases were allowed at any position. There were a total of 386 positions in the final data set.

3. Results and discussion

3.1. Crystal structure of AYWB-ME

The structure of AYWB-ME was solved at 2.5 Å resolution (Tables 2, 3 and 4; Fig. 1), with one monomer in the asymmetric unit. The refined atomic model comprises residues 3–389 of AYWB-ME; only Ile1, Lys2 and Arg390 are not seen in the electron density built in the model. One molecule of NAD and one Mg^{2+} cation are bound to AYWB-ME and are well defined in the electron-density map. In contrast, ADP and

pyruvate were not identified, despite being included in the crystallization solution.

AYWB-ME is organized into three domains, A, B and C (Figs. 1*a* and 1*b*), with all three displaying high identity to homologous domains of other MEs (Chang & Tong, 2003). Strikingly, AYWB-ME lacks domain D, which is present in all other reported ME structures (Xu *et al.*, 1999; Yang, Zhang *et al.*, 2002; Coleman *et al.*, 2002). Domain A (residues 3–50) is mostly α-helical, with one β-strand that forms an antiparallel two-stranded β-sheet with the equivalent strand of the other protomer in the dimer. Domain B is formed by two non-contiguous segments: residues 51–164 and 316–389. The former comprises a four-stranded parallel β-sheet with order 2–1–3–4 (β2–β5) and three α-helices (α5–α7), the first of which is on one side of the β-sheet while the latter two are on the other side. The second segment of domain B includes three α-helices (α14–α16), which complete the shielding of the β-sheet from the same side as α5. Finally, domain C (residues 165–315), the largest of the three domains, is placed as an inserted cassette within domain B and displays a Rossmann fold with a parallel five-stranded β-sheet that is sandwiched between two packs of three α-helices on both sides.

The crystal structure of AYWB-ME has one molecule in the asymmetric unit. Consistent with size-exclusion chromatography data (Saigo *et al.*, 2014), a dimeric species is readily identified in the crystal, with the two monomers related by a

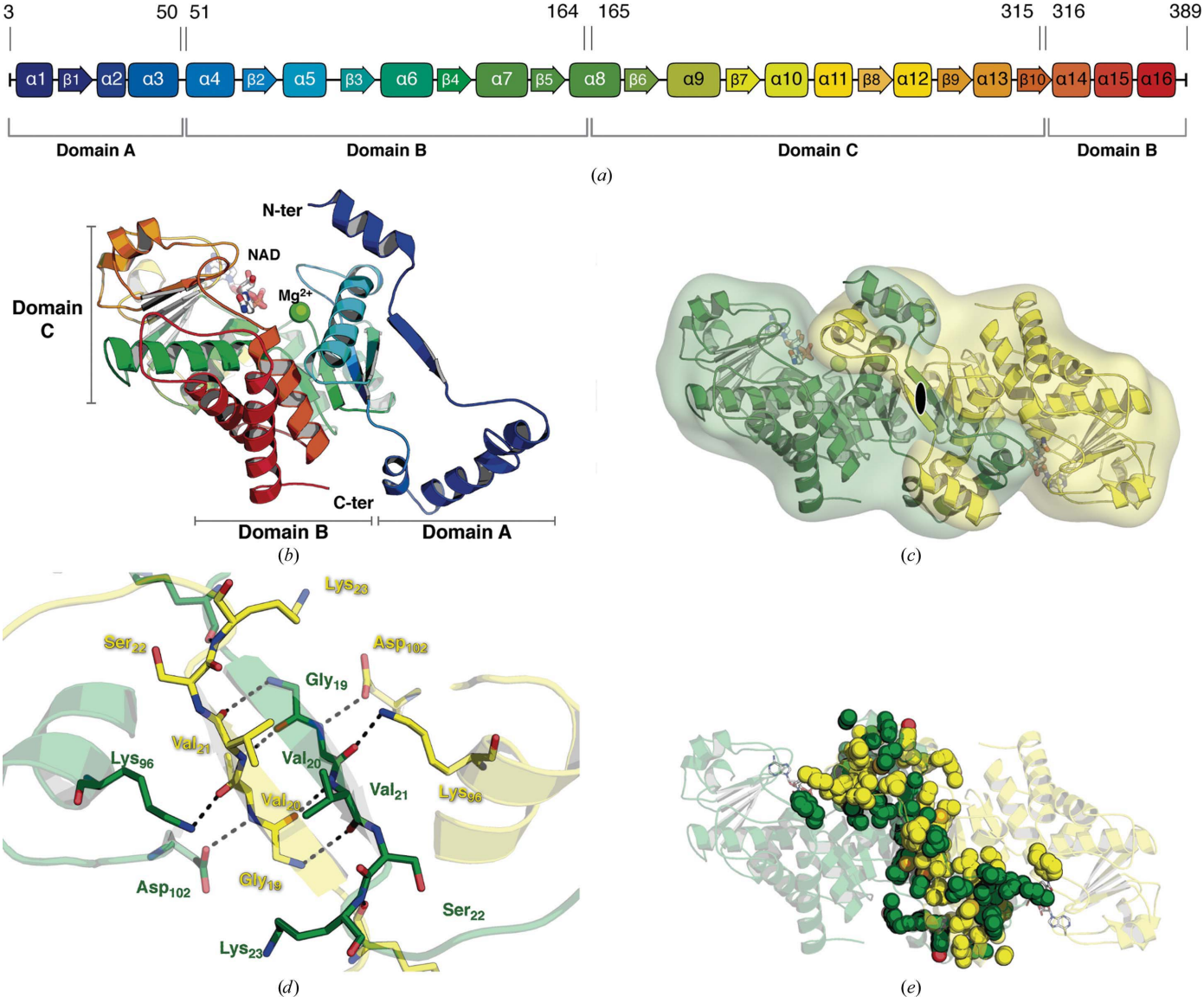


Figure 1
Crystal structure of AYWB-ME. (*a*) Schematic of secondary-structure elements arranged in domains A, B and C in the primary structure of AYWB-ME. α-Helices (α1–α16) and β-sheets (β1–β9) are labelled, and a blue-to-red colour ramp indicates the N-terminal to C-terminal direction of the polypeptide. Residue numbers mark domain boundaries. (*b*) Crystal structure of an AYWB-ME monomer in complex with NAD and Mg^{2+} in cartoon representation. NAD is shown in stick representation and Mg^{2+} in sphere representation, pinpointing the active site. The colour ramp is the same as in (*a*). The three domains A, B and C are indicated. (*c*) The twofold crystallographic axis that generates the biological dimer of AYWB-ME (one protomer is coloured yellow and the other green) is marked. (*d*) Residues forming hydrogen bonds and salt bridges at the dimer interface. (*e*) Residues located at the dimeric interface of AYWB-ME, as pinpointed by *PISA* analysis. Interface residues from each protomer are depicted in green and yellow.

twofold crystallographic axis (Fig. 1*c*). An extensive dimeric interface results from the association of domains A, which are intricately intertwined, with an additional contribution from both B domains, burying ∼4200 Å^2 of solvent-accessible surface overall. In addition to the β-sheet hydrogen bonds that bind the A domains to one another (Fig. 1*d*), numerous van der Waals contacts are additionally observed attaching domains A and B (Fig. 1*e*). Four ionic bonds further clamp this interface towards its borders, engaging the residue pairs Arg56/Glu12′ and Lys122/Glu115′. Notably, this inter-monomer association surface not only forms a stable dimer, but also contributes to building the active site, including the critical residue Tyr36′ from the other protomer of the dimer, as further elaborated below.

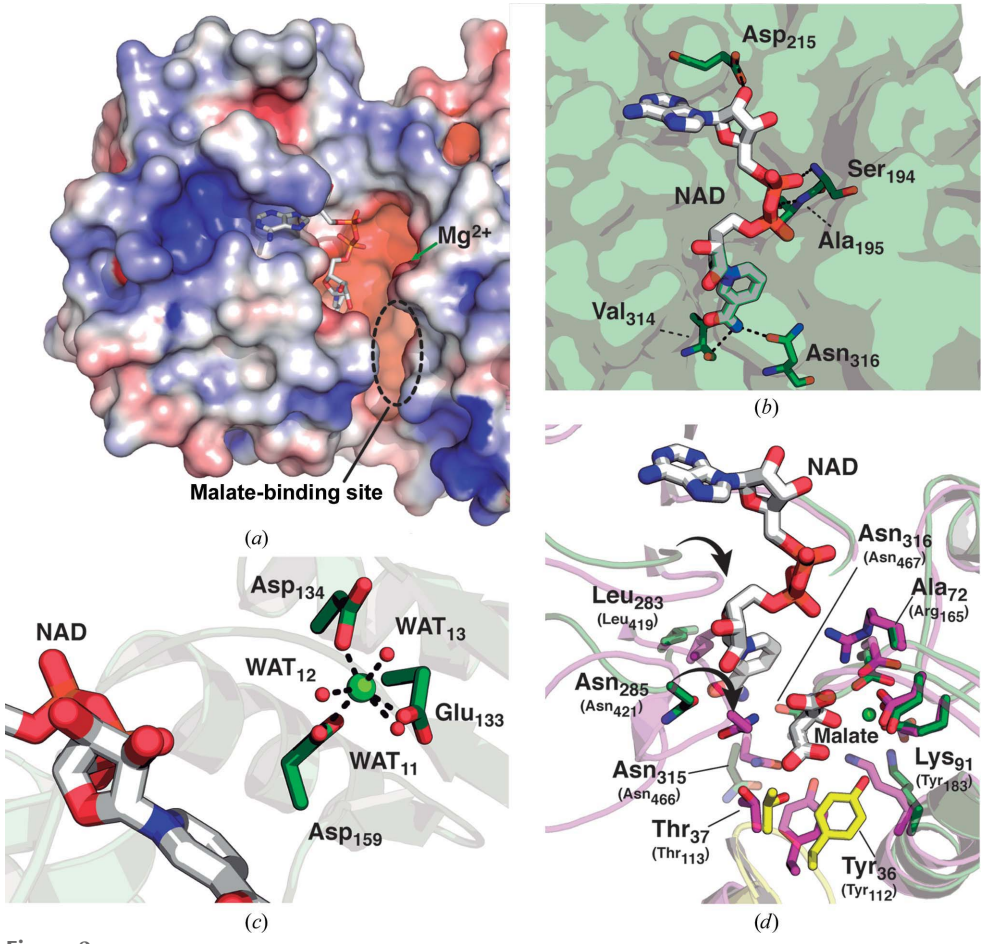


Figure 2
The binding sites for NAD, Mg^{2+} and malate in AYWB-ME. (*a*) The binding pocket for NAD (in stick representation) in AYWB-ME. The site for Mg^{2+} and the proposed malate-binding site are indicated with arrows. Positive to negative electrostatic potential is shown as a red to blue colour ramp mapped onto the solvent-exposed surface. (*b*) Residues of AYWB-ME involved in electrostatic interactions with NAD. Only residues involved in major contacts are shown. (*c*) Residues and water molecules at the metal pocket for Mg^{2+} binding in AYWB-ME. Mg^{2+} is shown as a small sphere. WAT11–WAT13 correspond to water molecules. (*d*) Predicted malate-binding site in AYWB-ME based on structural alignment with Human-ME. Residues conserved on binding oxalate (an analogue of the malate substrate) are shown in green stick representation for AYWB-ME and in purple for Human-ME. Below each residue of AYWB-ME, the number of the homologous residue in Human-ME is indicated in parentheses. Tyr36 (yellow stick model) is provided by the other monomer. NAD and oxalate are shown in stick representation and Mg^{2+} as a small sphere.

3.2. The active site of AYWB-ME

The NAD cofactor binds to domain C, within the active site of the enzyme, at the expected position for a Rossmann-fold domain (Figs. 2*a* and 2*b*). The nicotinamide moiety is snugly accommodated within its pocket, mostly through van der Waals contacts with surrounding residues and overall shape complementarity. There is a hydrogen bond between the N atom of its amide group and the side chain of Asn316, and possibly a second weaker one to the main-chain O atom of Val314 (Fig. 2*b*). Hydrogen bonds hold the NAD diphosphate in place, bound to the backbone N atoms of Ser194 and Ala195 (Fig. 2*b*). Finally, the adenosine nucleoside fragment of NAD establishes a bidentate hydrogen bond between the O2 atom of its ribose and the carboxylate group of Asp215, with its adenine base stacked between the β7–α10 and β8–α12 loops.

AYWB-ME has two putative dinucleotide-binding motifs, G*X*G*XX*G, which are highly conserved in sequence (Detarsio *et al.*, 2003; Hsieh *et al.*, 2011). One corresponds to residues 75-GLGNIG-80 in domain B; however, according to the structure this motif is not actually involved in binding NAD (Fig. 2*b*). This is consistent with site-directed mutagenesis of this conserved motif in maize NADP-ME, which showed that these residues are not essential for ME activity (Detarsio *et al.*, 2003). On the other hand, the second dinucleotide-binding motif located within loop β6–α9 in domain C (191-GAGSAG-196) is engaged in NAD binding, acting as the floor of the adenine-binding groove, in a similar way as found in other Rossmann-fold domains.

An Mg^{2+} cation was found bound within the active site, within <8 Å of the NAD nicotinamide, interacting with several residues in domain B (Fig. 2*c*). The cation is octahedrally coordinated by six O atoms: four equatorial, contributed by two water molecules and two carboxylate groups (Asp134 and Asp159), and two axial, provided by an additional water and the carboxylate of Glu133 (Fig. 2*c*).

We could not obtain the structure of AYWB-ME in complex with the substrate malate, but it can be modelled within the active site by superimposing the Human-ME structure (PDB entry 1do8; Saigo *et al.*, 2013) bound to oxalate, a malate

analogue. Malate is thus expected to bind within the reaction pocket roughly between the nicotinamide of NAD and the Mg^{2+} cation (Fig. 2*d*). Such comparative analysis reveals most of the residues involved in substrate binding (Chang & Tong, 2003) to be strictly conserved, including Leu74, Lys91, Asn285 and Asn316 (following the AYWB-ME numbering scheme). Only one position seems to be drastically modified: an arginine that is key for malate binding in Human-ME (Chang & Tong, 2003) and interacts strongly with oxalate in the crystal structure, which is substituted by Ala72 in AYWB-ME (Fig. 2*d*). Further studies of AYWB-ME in complex with malate will elucidate whether malate is bound in a subtly different way in the bacterial and human enzymes.

3.3. Structures of AYWB-ME *versus* Human-ME

A more thorough comparative analysis of the structures of AYWB-ME that we report here (PDB entry 5cee) and of Human-ME (PDB entry 1do8; Saigo *et al.*, 2013) revealed conserved motifs within domains B and C (Fig. 3; Supplementary Fig. S2). AYWB-ME is nearly 30% shorter (389 *versus* 584 amino acids), thus defining a minimal ME architecture. Domain B in AYWB-ME has a smaller, four-stranded β-sheet core instead of the five-stranded core in previously reported MEs. Domain C in Human-ME is also larger, with a fairly long β-hairpin toward its C-terminal end inserted between β10 and α14 according to AYWB-ME topology.

The size of domain A is substantially reduced in AYWB-ME with respect to the human orthologue (Fig. 3, Supplementary Fig. S2). Only α1 in the bacterial enzyme is located at an equivalent position, to α4 in the human protein, but no detectable sequence identity is conserved throughout, and the whole of domain A contains only three α-helices in AYWB-ME compared with seven in Human-ME. Furthermore, domain D of Human-ME is completely absent in AYWB-ME. Domain D has been reported to be critical for the tetrameric structure of Human-ME and other MEs (Chang & Tong, 2003). The absence of most secondary-structure elements of domain A, and the complete absence of domain D, in AYWB-ME indicates that these two domains are not essential for ME activity nor for the basic dimeric organization of these enzymes. The AYWB-ME structure now suggests that domain D may instead be likely to be involved in stabilization of the tetrameric species (Chang & Tong, 2003), an important property of previously characterized MEs. More importantly, we posit that the tight modulatory action of several effectors (Saigo *et al.*, 2014), as well as the basic dimeric architecture of MEs, are only dependent on a minimal structural organization comprising domains B and C.

Differences in the regulatory mechanisms cannot be excluded on the basis of the structural particularities observed in AYWB-ME compared with Human-ME. The latter has been captured in an open configuration, as required for substrate binding and product release, as well as in a closed form, which is likely to be the catalytically competent conformation. The closure of the active site is mediated by a rigid-body movement of domain C with respect to domain B, while domains A and D undergo a shift, ultimately reorganizing the whole tetramer (Chang & Tong, 2003). In the case of AYWB-ME the crystal structure that we report here shows an open configuration, but given the absence of the entire domain D, the smaller size of domain A and the dimerization interface of AYWB-ME, important variations in the allosteric regulation are anticipated.

Several differences can also be identified in the residues that bind NAD between AYWB-ME and Human-ME (Supplementary Fig. S3). Notably, Lys233, which is known to interact with the phosphate group of NADP in the NADP-dependent MEs from pigeon (Lys340; Kuo *et al.*, 2000) and maize (Lys435; Detarsio *et al.*, 2004), is conserved in AYWB-ME. Consistent with this finding,

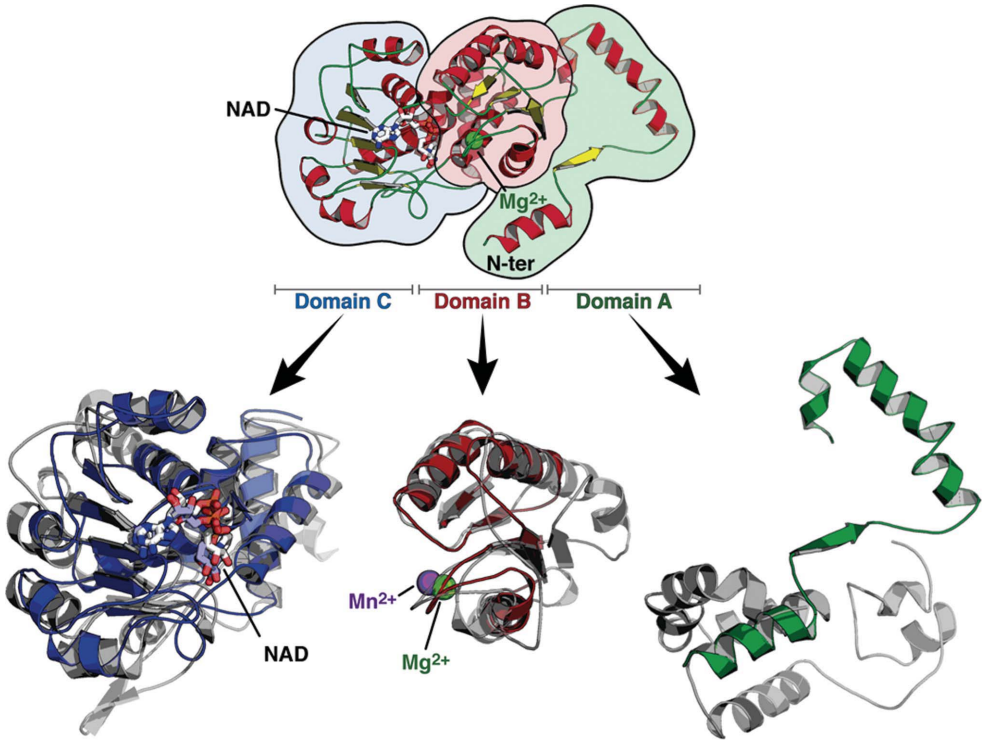


Figure 3
Structural alignment of AYWB-ME and Human-ME (PDB entry 1do8). The three-dimensional structures of domains A (green), B (red) and C (blue) of AYWB-ME are shown at the bottom of the figure, superimposed onto the corresponding domains of Human-ME (in grey). Note that domain D of Human-ME is not found in AYWB-ME. The full structure of AYWB-ME is shown at the top to reference the three-dimensional positions of the three domains.

the bacterial enzyme is indeed active with both NAD and NADP (Saigo *et al.*, 2014), making this lysine residue a plausible determinant of its dual-cofactor specificity. The structure of AYWB-ME in complex with NADP will confirm this hypothesis in future studies.

3.4. The dimeric interface of AYWB-ME: comparison with classic tetrameric MEs

The dimeric structure of AYWB-ME is novel among MEs. Human, pigeon and *A. suum* MEs are all tetrameric and are organized according to a *D*2 dimer-of-dimers architecture (Chang & Tong, 2003; Coleman *et al.*, 2002; Yang, Zhang *et al.*, 2002). Although an extensive dimeric surface interaction is found in all of these tetrameric structures, none of them follow the dimerization configuration seen in AYWB-ME (Fig. 1*e*). AYWB-ME has ∼85% more residues engaged in the dimer interface compared with Human-ME: 105 residues in the former compared with 58 residues in the latter (Supplementary Fig. S4). Hence, the stability of the AYWB-ME dimer is predicted to be approximately fourfold higher in terms of calculated energy compared with that of Human-ME. Not only is the inter-monomer interaction surface larger, but its three-dimensional configuration is absolutely novel in AYWB-ME and has not previously been reported for any other ME. Such a domain A-mediated interface makes a substantial contribution to the intimate interaction between monomers in AYWB-ME (Figs. 1*c* and 1*e*).

3.5. The active site of AYWB-ME: both monomers are involved

All of the residues previously reported to comprise the active site of Human-ME are conserved in AYWB-ME: Tyr36, Lys91, Glu133, Asp134 and Asp159 (Fig. 4). However, Tyr36 is provided by the other monomer in each of the two reaction centres of the dimer only in the bacterial enzyme (Fig. 4). Strikingly, Tyr36 and Lys91 are located at equivalent positions to Tyr112 and Lys183, respectively, in Human-ME and other MEs (Coleman *et al.*, 2002; Chang & Tong, 2003), with Tyr112 located within domain A of the same protomer. Because of the novel conformation of the active site adopted by AYWB-ME, with Tyr36 contributed by the other protomer, we measured the catalytic activity of the point mutant AYWB-ME_Tyr36Ala. As expected on structural grounds, ME activity of AYWB-ME_Tyr36Ala was not detectable at various substrate and cofactor concentrations and alternative pH conditions (data not shown), proving the critical role of Tyr36 in catalysis. The lack of activity could be owing to hindrance to dimerization, but analytic size-exclusion chromatography data proved that both wild-type AYWB-ME and the AYWB-ME_Tyr36Ala mutant behave as stable dimers in solution (with apparent molecular weights of 79.0 ± 5.0 kDa; data not shown) and do not show significant differences in their CD spectra (Supplementary Fig. S5). In conclusion, whereas Tyr36 is not critical for AYWB-ME dimerization, it is essential for catalysis and/or to stabilize a competent conformation of the active site.

3.6. Lessons from the three-dimensional structure of phytoplasma ME: a minimal scaffold for a functional ME and a plausible evolutionary pathway towards ME diversification

The crystal structure of AYWB-ME showcases the smallest of all MEs that have been characterized to date. This allows a minimal structural scaffold sufficient to retain ME activity without losing metabolite regulation to be posited. This minimal structure engages key structural motifs and catalytic residues that have previously been described. The AYWB-ME monomer is divided into three domains (A, B and C) and lacks domain D, which is known to be essential in the tetramerization of other MEs (Chang & Tong, 2003; Detarsio *et al.*, 2007). Importantly, domain A of AYWB-ME displays a novel conformation that not has not been found before in any ME. Mostly helical, domain A of AYWB-ME adopts a configuration resembling an arm that leans onto the other monomer and contributes to building the active site of the neighbouring monomer. This contribution to the reaction centre is configured through the Tyr36 residue of the arm, a particularly critical position. Indeed, helices α2 and α3 in domain A almost constitute a single continuous helix, except that Tyr36 and Thr37 interrupt this continuity, introducing a shift in the position of the main axis of α2 with respect to that of α3. The key importance of Tyr36 in AYWB-ME activity can be anticipated on the basis of its particular position, establishing a hydrogen bond to one of the waters that coordinate the Mg^{2+} cation. In close proximity to Lys91 and Glu133, the precise organization of atoms in this pocket is expected to define the malate-binding site between the cation and the dinucleotide NAD moiety.

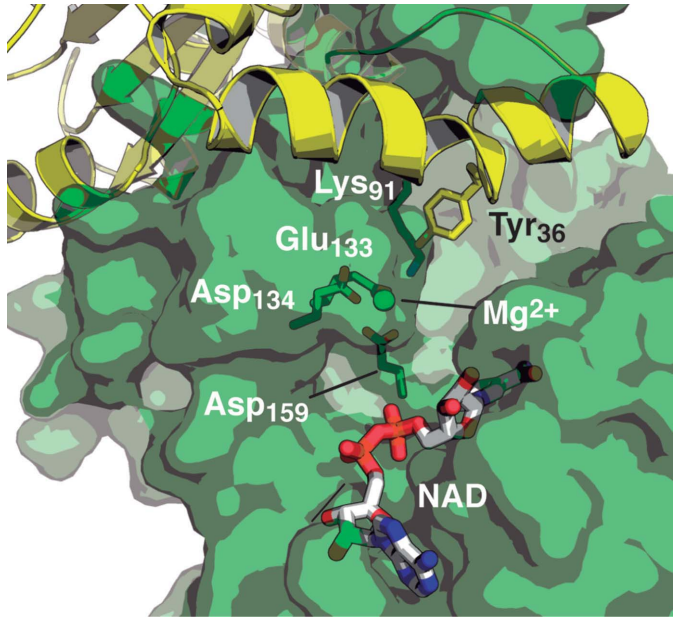


Figure 4
Residues involved in the active site of AYWB-ME. Both of the active sites of AYWB-ME (only one is depicted here) show contributions of amino acids from both monomers. The residues involved in the active site are shown in green, while Tyr36 provided by the other monomer is highlighted in yellow.

An analysis of ME sequences led to their classification into three types: (i) MEs composed of monomers of ∼60 kDa (large-subunit MEs), which are ubiquitous in eukaryotes and are present in only some prokaryotes; (ii) MEs composed of monomers of ∼40 kDa (small-subunit MEs), which are found almost exclusively in prokaryotes; and (iii) the chimeric MEs from some prokaryotes, which are fused to a phosphotransacetylase domain (Saigo *et al.*, 2014; Bologna *et al.*, 2007). Sequence and phylogenetic analyses of MEs encoded in the genomes of 36 species of archaea and eubacteria suggest that large-subunit MEs and chimeric MEs might have evolved from small-subunit MEs by gaining small sequence cassettes or an entire phosphotransacetylase domain, respectively (Supplementary Fig. S1). Among the available crystal structures of MEs, those from the eubacteria *Candidatus* Phytoplasma AYWB (AYWB-ME, analysed here), *T. maritima* (PDB entries 1vl6 and 2hae) and *S. pyogenes* (PDB entry 2a9f) and from the archaeon *P. horikoshii* (PDB entries 1ww8 and 2dvm) are all small-subunit MEs. In the case of the unicellular eukaryote *E. histolytica* (PDB entry 3nv9), the existence of a small ME has been explained by horizontal gene transfer from an anaerobic archaeon that shared the same ecological niche (Field *et al.*, 2000). This work contributes the first report of a three-dimensional structure from a small-subunit ME, AYWB-ME, unveiling the architecture of the minimal scaffold for an active ME to function properly, which is likely to be the basis from which the large diversity of MEs may have evolved.

Acknowledgements

We thank Rodolfo Rasia for assistance with circular-dichroism interpretation. AM, CEA, CSA, MFD and MA are members of the Researcher Career of CONICET and AG is a fellow of the same institution.

Funding information

This work was supported by ANPCyT and CONICET.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Alvarez, C. E., Saigo, M., Margarit, E., Andreo, C. S. & Drincovich, M. F. (2013). *Photosynth. Res.* **115**, 65–80.
Arias, C. L., Andreo, C. S., Drincovich, M. F. & Gerrard Wheeler, M. C. (2013). *Plant Mol. Biol.* **81**, 297–307.
Bertaccini, A. (2007). *Front. Biosci.* **12**, 673–689.
Bologna, F. P., Andreo, C. S. & Drincovich, M. F. (2007). *J. Bacteriol.* **189**, 5937–5946.
Casati, P., Drincovich, M. F., Edwards, G. E. & Andreo, C. S. (1999). *Photosynth. Res.* **61**, 99–105.
Chang, G.-G. & Tong, L. (2003). *Biochemistry*, **42**, 12721–12733.
Coleman, D. E., Rao, G. S. J., Goldsmith, E. J., Cook, P. F. & Harris, B. G. (2002). *Biochemistry*, **41**, 6928–6938.
Daily, M. D. & Gray, J. J. (2009). *PLoS Comput. Biol.* **5**, e1000293.
Detarsio, E., Alvarez, C. E., Saigo, M., Andreo, C. S. & Drincovich, M. F. (2007). *J. Biol. Chem.* **282**, 6053–6060.
Detarsio, E., Andreo, C. S. & Drincovich, M. F. (2004). *Biochem. J.* **382**, 1025–1030.
Detarsio, E., Maurino, V. G., Alvarez, C., Müller, G., Andreo, C. S. & Drincovich, M. F. (2008). *Plant Mol. Biol.* **68**, 355–367.
Detarsio, E., Gerrard Wheeler, M. C., Campos Bermúdez, V. A., Andreo, C. S. & Drincovich, M. F. (2003). *J. Biol. Chem.* **278**, 13757–13764.
Drincovich, M. F., Casati, P. & Andreo, C. S. (2001). *FEBS Lett.* **490**, 1–6.
Edwards, G. E. & Andreo, C. S. (1992). *Phytochemistry*, **31**, 1845–1857.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, P. R. (2011). *Acta Cryst.* D**67**, 282–292.
Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.
Felsenstein, J. (1985). *Evolution*, **39**, 783–791.
Field, J., Rosenthal, B. & Samuelson, J. (2000). *Mol. Microbiol.* **38**, 446–455.
Gerrard Wheeler, M. C., Arias, C. L., Maurino, V. G., Andreo, C. S. & Drincovich, M. F. (2009). *FEBS J.* **276**, 5665–5677.
Gerrard Wheeler, M. C., Arias, C. L., Tronconi, M. A., Maurino, V. G., Andreo, C. S. & Drincovich, M. F. (2008). *Plant Mol. Biol.* **67**, 231–242.
Hogenhout, S. A., Oshima, K., Ammar, E.-D., Kakizawa, S., Kingdom, H. N. & Namba, S. (2008). *Mol. Plant Pathol.* **9**, 403–423.
Hsieh, J.-Y., Chen, M.-C. & Hung, H.-C. (2011). *PLoS One*, **6**, e25312.
Hsu, R. Y. (1982). *Mol. Cell. Biochem.* **43**, 3–26.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 133–144.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Kube, M., Mitrovic, J., Duduk, B., Rabus, R. & Seemüller, E. (2012). *ScientificWorldJournal*, **2012**, 185942.
Kumar, S., Stecher, G. & Tamura, K. (2016). *Mol. Biol. Evol.* **33**, 1870–1874.
Kuo, C.-C., Tsai, L.-C., Chin, T.-Y., Chang, G.-G. & Chou, W.-Y. (2000). *Biochem. Biophys. Res. Commun.* **270**, 821–825.
Laporte, M., Shen, B. & Tarczynski, M. C. (2002). *J. Exp. Bot.* **53**, 699–705.
Le, S. Q. & Gascuel, O. (2008). *Mol. Biol. Evol.* **25**, 1307–1320.
Liu, Y., Ren, D., Pike, S., Pallardy, S., Gassmann, W. & Zhang, S. (2007). *Plant J.* **51**, 941–954.
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.
Padilla, J. E. & Yeates, T. O. (2003). *Acta Cryst.* D**59**, 1124–1130.
Saigo, M., Bologna, F. P., Maurino, V. G., Detarsio, E., Andreo, C. S. & Drincovich, M. F. (2004). *Plant Mol. Biol.* **55**, 97–107.
Saigo, M., Golic, A., Alvarez, C. E., Andreo, C. S., Hogenhout, S. A., Mussi, M. A. & Drincovich, M. F. (2014). *Microbiology*, **160**, 2794–2806.
Saigo, M., Tronconi, M. A., Gerrard Wheeler, M. C., Alvarez, C. E., Drincovich, M. F. & Andreo, C. S. (2013). *Photosynth. Res.* **117**, 177–187.
Strauss, E. (2009). *Science*, **325**, 388–390.
Tronconi, M. A., Fahnenstich, H., Gerrard Wheeler, M. C., Andreo, C. S., Flügge, U.-I., Drincovich, M. F. & Maurino, V. G. (2008). *Plant Physiol.* **146**, 1540–1552.
Tronconi, M. A., Maurino, V. G., Andreo, C. S. & Drincovich, M. F. (2010). *J. Biol. Chem.* **285**, 11870–11879.
Xu, Y., Bhargava, G., Wu, H., Loeber, G. & Tong, L. (1999). *Structure*, **7**, 877–889.
Yang, Z., Lanks, C. W. & Tong, L. (2002). *Structure*, **10**, 951–960.
Yang, Z., Zhang, H., Hung, H.-H., Kuo, C.-C., Tsai, L.-C., Yuan, H. S., Chou, W.-Y., Chang, G.-G. & Tong, L. (2002). *Protein Sci.* **11**, 332–341.